Geographical Studies and Development
Volume 1 Issue 1 (July-Dec 2026) | **ISSN**: Applied | **Page No**: 01-22

Machine-learning-based two-dimensional building exposure assessment to rainfall-triggered landslides in the South Sikkim Himalaya

Shivam Priyadarshi[1]

[1] Department of Geography, School of Earth, Biological and Environmental Sciences,
Central University of South Bihar, Gaya, India

** Corresponding author: shivampriyadarshi@cusb.ac.in*

Abstract

Landslide risk in the Sikkim Himalaya is increasingly shaped by the building stock accumulating on steep, weathered slopes, yet quantitative information on the exposure of buildings to rainfall-triggered slope failure remains scarce at the settlement scale. This study couples a Random Forest (RF) landslide susceptibility model with a settlement-scale building inventory to quantify building exposure in South Sikkim, Indian Himalaya. Ten conditioning factors—slope, aspect, curvature, distance to rivers, distance to roads, distance to lineaments, Topographic Wetness Index (TWI), Land Use/Land Cover (LULC), Normalized Difference Vegetation Index (NDVI) and Stream Power Index (SPI)—were combined with a multi-source landslide inventory to train the classifier and produce a five-class susceptibility surface, which was then intersected with the building layer. The RF model attained an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.802, indicating good discriminative capability. High-susceptibility zones cover approximately 24% of the study area but contain 38% of the buildings, revealing a systematic over-representation of the built environment in the most hazardous terrain and a strong potential for cascading disruption of settlement services. The framework is data-parsimonious, reproducible, and directly transferable to comparable Himalayan settlements, providing an operational basis for hazard-informed regulation of new construction and for integrating exposure into landslide risk reduction in mountain regions.

Keywords: Landslide susceptibility; Random Forest; building exposure; rainfall-triggered landslides; South Sikkim; Himalaya; disaster risk reduction

1. Introduction

Landslides are among the most destructive geohazards of mountain regions, causing recurrent loss of life, infrastructure, and economic assets (Das et al., 2026). The Himalayan arc is a global hotspot: the Indian Himalayan Region (IHR) accounts for roughly 15 % of the world's rainfall-induced landslides (Dikshit et al., 2020; Saha et al., 2021), for around 80 % of all landslides recorded in India (J. Das et al., 2023), and for landslide damage exceeding US $1 billion and approximately 200 deaths per year — some 30 % of global landslide fatalities (Koley et al., 2019). Within India, nearly 15 % of the national territory is landslide-prone (Das et al., 2026), and the Darjeeling–Sikkim segment alone contributes 40–43 % of the country's landslide-susceptible area (J. Das et al., 2023; Sarkar & Saha, 2026). Slope failure across the range is driven by the interaction of steep relief, active tectonics, fragile lithology, and monsoonal rainfall, with high-intensity storms acting as the dominant trigger (Dikshit et al., 2020; Koley et al., 2019; Rafique & Joshi, 2024).

The Sikkim Himalaya, on the eastern flank of the IHR, concentrates most of the conditions that predispose slopes to failure: fragile Lesser and Higher Himalayan lithologies, steep relief, active seismicity, monsoonal precipitation, and rapidly intensifying anthropogenic pressures (M. R. Das et al., 2026; J. Das et al., 2023; Nawazuzzoha et al., 2025). Under the National Landslide Susceptibility Mapping (NLSM) programme, some 3,371 active landslides have been mapped in the state, and the 1:50,000 susceptibility map places 18 % of the area in the "High" class while capturing 65 % of historical failures (Bansal et al., 2022). At the settlement scale, the consequences are already visible: parts of Gangtok have been assessed as very high (7.51 %) or high (18.18 %) landslide risk zones (Zimik et al., 2022), several urban neighbourhoods have been abandoned after damage to housing, schools, and offices from persistent ground displacement (Bhasin et al., 2023), and rain-triggered slope failure recurrently disrupts settlements and roads across the state (Koley et al., 2019).

Damaging events across the Himalaya routinely translate slope instability into building loss: extreme-rainfall episodes have collapsed multi-storey structures and buried villages under rockslides (Kumar et al., 2017; Rajamani & Iyer, 2024), and in Sikkim the 2023 glacial lake outburst flood triggered secondary landslides that damaged approximately 200 buildings, 90 % of them from just two adjacent failures (Sattar et al., 2025). Anthropogenic modification of the hillslope system — deforestation, road cutting, and unplanned construction — has compounded these pressures across the Darjeeling–Sikkim Himalaya (M. R. Das et al., 2026;

Sarkar & Saha, 2026), with cultivated soils on cleared slopes exhibiting reduced strength and elevated pore-water pressure that further predisposes them to failure during prolonged or intense rainfall (Sarkar & Saha, 2026).

Landslide susceptibility mapping (LSM) is the standard entry point for translating hazard understanding into land-use and mitigation decisions (M. R. Das et al., 2026; Rafique & Joshi, 2024). In Sikkim, LSM has progressed rapidly from heuristic and bivariate statistical models — notably AHP, frequency ratio, information value, and logistic regression, and their hybrids such as AHP–FR (AUC 0.81–0.85) (M. R. Das et al., 2026) — to a family of data-driven machine learning classifiers (Dikshit et al., 2020; J. Das et al., 2023; Saha et al., 2021). Recent Sikkim studies span Random Forest and stacked ensembles of ANN, SVM, and GBDT in South Sikkim (Nawazuzzoha et al., 2025), Gradient Boosting Machine reaching AUC values as high as 0.99 (Roy et al., 2025), and CNN-based deep learning in East Sikkim with AUC of 0.918–0.933 (Saha et al., 2021). Across these studies, distance to roads, rainfall, elevation, and slope emerge consistently as leading predictors (Nawazuzzoha et al., 2025; Roy et al., 2025), and RF in particular has proven robust to correlated Himalayan predictor sets and to the presence–absence framing typical of inventory-based training data.

Despite these advances, the use of deep learning and artificial intelligence, which shows great promise in geohazard studies, remains limited for the Indian Himalayan region (Dikshit et al., 2020). Furthermore, despite the entire region being highly landslide-prone, most studies have focused on few regions and large areas have been neglected (Dikshit et al., 2020). There is a notable lack of research specifically focused on the South Sikkim district and an evident lack of application of machine learning techniques in that area (Nawazuzzoha et al., 2025).

Landslide risk is conventionally decomposed into hazard, exposure, and vulnerability (Kato et al., 2021). The Sikkim literature reviewed above is overwhelmingly weighted towards the hazard component: susceptibility maps of increasing statistical sophistication have proliferated (J. Das et al., 2023; M. R. Das et al., 2026; Nawazuzzoha et al., 2025; Roy et al., 2025; Saha et al., 2021), yet the exposure of the built environment, the identification and quantification of buildings actually located in hazardous terrain — remains addressed at only a handful of sites. Sattar et al. (2025) tracked change in building exposure along the Teesta corridor between 2013 and 2023 and showed that exposure is growing faster than the underlying hazard footprint, an observation with direct implications for the future loss burden. Yet, to our knowledge, no published study has intersected a Sikkim-specific machine-learning susceptibility surface with a settlement-scale building layer for South Sikkim, the district identified by Nawazuzzoha et al. (2025) as both highly susceptible and understudied, to produce a quantitative, planning-ready exposure assessment.

Settlements in South Sikkim are typically strung along ridges and road corridors, and multi-storey reinforced-concrete buildings founded on cut-and-fill benches have become increasingly common in the last two decades. This expansion places new construction on precisely the mid-slope and ridge-flank positions that are most susceptible to failure — a pattern documented across the Darjeeling–Sikkim Himalaya (Laha et al., 2024; Shahi et al., 2022) and, in the wider eastern Himalaya, associated with rising building and roadway exposure to mass movement (Chen et al., 2024; Dubey et al., 2023). Recurrent slope failures at single sites are characteristic of the region: the settlement of Pathing, in South Sikkim, experienced damaging landslides in December 2020, 2021, and 2022, consistent with the 30-year Himalaya-wide pattern of landslide persistence documented by Chen et al. (2024), who report average recovery times of about three years in Bhutan and eastern India. This combination of steep, structurally weakened terrain, monsoon-dominated triggering, growing settlement footprint, and recurrent failure makes South Sikkim a representative and time-critical test bed for building-scale exposure assessment.

2. Study Area

The study area is situated in the South Sikkim region of the Indian state of Sikkim, located within the Eastern Himalaya. Sikkim is bordered by Nepal to the west, the Tibet Autonomous Region of China to the north and northeast, Bhutan to the southeast, and the Indian state of West Bengal to the south. The Eastern Himalaya, including Sikkim, represents one of the most geologically active and geomorphologically dynamic mountain regions globally (Chen et al., 2024; Dubey et al., 2023). The South Sikkim region is characterized by steep-to-moderate hillslopes, deeply incised tributaries of the Teesta and Rangit rivers, and a dispersed pattern of ridge-top and mid-slope settlements. Elevation across the study area ranges from a few hundred metres above mean sea level along the valley floors to more than 2,000 m on the higher ridges, and slope gradients in excess of 25° are common in and around the settled areas. Such steep terrain and rugged topography are characteristic of the broader Himalayan system, where landslides constitute a prominent and recurring geohazard (KC et al., 2022; Rafique & Joshi, 2024).

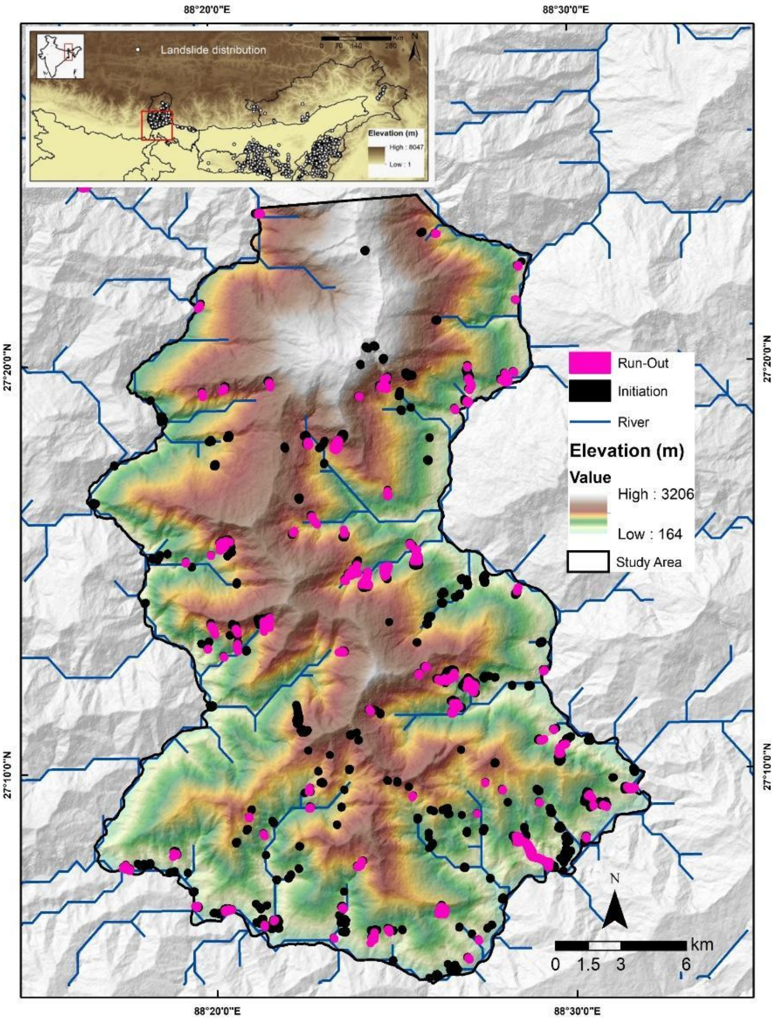

Figure 1. Location of the study area in South Sikkim, India showing landslide initiation and runout in the region.

Geologically, the region is underlain by rocks of the Lesser Himalayan and Higher Himalayan sequences, cut by regional lineaments and thrust structures that condition the distribution of weathered and fractured material. The Lesser Himalaya, also referred to as the thrust-and-fold belt, exhibits complex geological settings due to the folding and thrusting of crystalline sequences (Shahi et al., 2022). Across the broader Himalayan arc, lithology, structural discontinuities, and proximity to faults are recognized as critical conditioning factors for slope instability (Beigh & Bukhari, 2024; Naik & Palakuzhiyil, 2024; Rafique & Joshi, 2024). In Sikkim specifically, structure and lithology play important roles in landslide triggering, with the landslide hazard index being highest for rainfall followed by structures, lithology, slope, land use/land cover, and drainage density (Rafique & Joshi, 2024). The Indian Himalayan Region, owing to its active tectonics and fragile geological conditions, is among the most vulnerable zones to landslide hazards globally (Laha et al., 2024; Rafique & Joshi, 2024).

Large-scale slope failures have been associated with major seismic events in the region, including the 2011 Sikkim earthquake (Rafique & Joshi, 2024).

Climatically, the area lies within the Indian summer monsoon regime, with more than 80% of the annual precipitation delivered between June and September. This monsoonal precipitation pattern is a defining characteristic of the Himalayan hazard environment, where high-intensity rainfall spells during the monsoon period serve as the principal trigger of shallow translational slides, debris flows, and slumps that damage buildings and disrupt roads (Kincey et al., 2024; Laha et al., 2024; Rafique & Joshi, 2024). The eastern Himalaya receives substantial monsoon rainfall, with spatial gradients strongly influenced by orographic effects as moist monsoon air ascends the mountain slopes (Krishna, 2025). Froude and Petley (2018), as cited in the literature, highlighted that the Indian Himalaya accounts for approximately 16% of the world's rainfall-induced landslides (Laha et al., 2024). Almost all landslides in the Sikkim region have been reported during the rainy season (Rafique & Joshi, 2024), consistent with the broader pattern across the Himalaya where monsoon-triggered mass movements represent the dominant hazard mechanism (Chen et al., 2024; Kincey et al., 2024).

The relationship between intense precipitation and slope failure is well established across the Himalayan system. In the Darjeeling-Sikkim Himalaya, landslides are annually triggered by tropical rainfall, leading to sudden loss of wealth, houses, and even casualties (Laha et al., 2024). Similarly, across Nepal and the broader central and eastern Himalaya, rainfall-triggered landslides impose a chronic impediment to sustainable livelihoods, with millions of people residing in critical zones potentially exposed to such catastrophes (Dubey et al., 2023; Kincey et al., 2024).

Settlements in South Sikkim are typically strung along ridges and road corridors, and multi-storey reinforced-concrete buildings founded on cut-and-fill benches have become increasingly common in the last two decades. This expansion places new construction on precisely the mid-slope and ridge-flank positions that are most susceptible to failure — a pattern documented across the Darjeeling–Sikkim Himalaya (Laha et al., 2024; Shahi et al., 2022) and, in the wider eastern Himalaya, associated with rising building and roadway exposure to mass movement (Chen et al., 2024; Dubey et al., 2023). Recurrent slope failures at single sites are characteristic of the region: the settlement of Pathing, in South Sikkim, experienced damaging landslides in December 2020, 2021, and 2022, consistent with the 30-year Himalaya-wide pattern of landslide persistence documented by Chen et al. (2024), who report average recovery times of about three years in Bhutan and eastern India. This combination of steep, structurally weakened terrain, monsoon-dominated triggering, growing settlement footprint, and recurrent failure makes South Sikkim a representative and time-critical test bed for building-scale exposure assessment.

3. Materials and Methods

3.1 Overall workflow

The workflow (Figure 2) comprises four stages: (i) preparation of a landslide inventory and a set of ten conditioning factors; (ii) training of a Random Forest classifier to produce a landslide susceptibility surface, with independent accuracy assessment; (iii) preparation of a building infrastructure layer for the study area; and (iv) exposure analysis, in which the susceptibility surface is intersected with the building layer to identify buildings located within high and very high susceptibility zones.

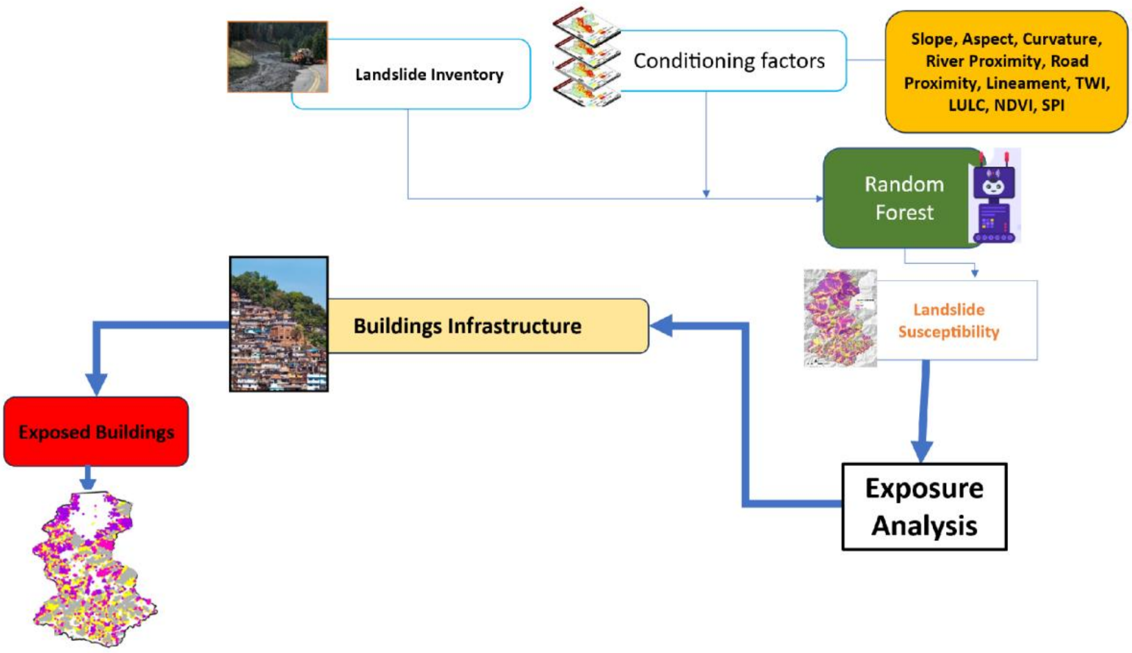

Figure 2. Methodological workflow for landslide susceptibility mapping and building exposure analysis in South Sikkim.

3.2 Landslide inventory

A landslide inventory for the study area was compiled from available records of past events, complemented by mapping of recent landslides visible in satellite imagery and reported by local sources, including the citizen-monitoring platform Save The Hills, which documents recurrent failures such as those at Pathing in December 2020, 2021, and 2022. Each event was represented as a point or polygon and attributed with, where possible, the type of movement and the approximate date of occurrence. Non-landslide samples were drawn from areas of gentle terrain and known stable ground, at a minimum distance from the mapped landslides to reduce spatial contamination. The combined presence–absence dataset was split into training and testing subsets for model calibration and validation.

3.3 Conditioning factors

Ten conditioning factors were selected based on their physical relevance to rainfall-triggered slope failure in Himalayan terrain and their established use in the landslide susceptibility literature (Beigh & Bukhari, 2024; KC et al., 2022; Naik & Palakuzhiyil, 2024; Rafique & Joshi, 2024). These factors are widely recognized as key contributors to slope instability across the Himalaya and have been employed in numerous susceptibility assessments using both statistical and multicriteria approaches (Kincey et al., 2024; Laha et al., 2024; Shahi et al., 2022).

The selected factors comprise five terrain descriptors — slope, aspect, curvature, Topographic Wetness Index (TWI), and Stream Power Index (SPI); two proximity descriptors — distance to rivers and distance to roads; one structural descriptor — distance to lineaments; and two land-surface descriptors — Land Use/Land Cover (LULC) and Normalized Difference Vegetation Index (NDVI). Slope gradient, aspect, and curvature are consistently identified among the most influential topographic variables controlling landslide occurrence (Beigh & Bukhari, 2024; KC et al., 2022; Rafique & Joshi, 2024). TWI and SPI capture the hydrological behavior of hillslopes, reflecting soil saturation potential and erosive power of surface runoff, respectively (Naik & Palakuzhiyil, 2024; Rafique & Joshi, 2024). Proximity to drainage networks and roads are critical factors, as stream undercutting and road-cut excavation destabilize slopes (Beigh & Bukhari, 2024; Naik & Palakuzhiyil, 2024; Uniyal, 2010). Distance to lineaments captures the influence of structural discontinuities on rock mass strength and weathering depth (Beigh & Bukhari, 2024; Rafique & Joshi, 2024). LULC and NDVI characterize vegetation cover and land use patterns that modulate infiltration, root cohesion, and surface erosion (Laha et al., 2024; Rafique & Joshi, 2024; Shahi et al., 2022).

Terrain factors were derived from a digital elevation model (DEM), hydrological factors from the DEM-extracted drainage network, road and lineament proximities from vector layers, LULC from supervised classification of multispectral imagery, and NDVI from the same imagery (Beigh & Bukhari, 2024; Naik & Palakuzhiyil, 2024; Rafique & Joshi, 2024). All factors were resampled to a common raster grid and clipped to the study area, following standard GIS-based susceptibility mapping procedures (KC et al., 2022; Kincey et al., 2024; Shahi et al., 2022).

Table 1. Conditioning factors used for Random Forest based landslide susceptibility modelling.

No.	Factor	Type	Source / Derivation
1	Slope	Terrain (continuous)	Derived from DEM
2	Aspect	Terrain (categorical)	Derived from DEM
3	Curvature	Terrain (continuous)	Derived from DEM
4	River Proximity	Hydrological (continuous)	Euclidean distance to drainage network
5	Road Proximity	Anthropogenic (continuous)	Euclidean distance to road network
6	Lineament	Structural (continuous)	Euclidean distance to lineaments
7	Topographic Wetness Index (TWI)	Hydrological (continuous)	Derived from DEM (upslope area and slope)
8	Land Use / Land Cover (LULC)	Land surface (categorical)	Supervised classification of satellite imagery
9	NDVI	Land surface (continuous)	Red and NIR bands of satellite imagery
10	Stream Power Index (SPI)	Hydrological (continuous)	Derived from DEM (slope and upslope area)

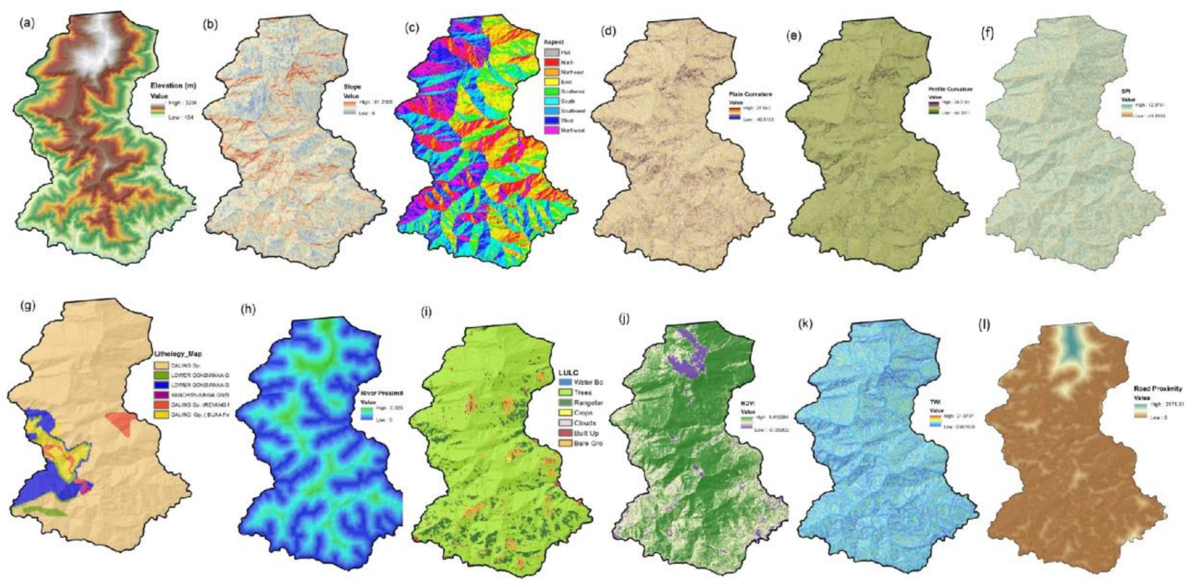

Figure 3. Spatial distribution of the ten conditioning factors used in the Random Forest model

3.4 Random Forest susceptibility modelling

Landslide susceptibility is modelled as a supervised binary classification problem using the Random Forest (RF) algorithm. RF is an ensemble learning method based on bagging that constructs multiple decision trees, each grown on a bootstrap sample of the training data, with each split selected from a random subset of predictors (Ageenko et al., 2022; Nhu et al., 2020; Rabby et al., 2023). The final class assignment is determined by aggregating the votes of all trees (Ageenko et al., 2022; Chen et al., 2023), and the resulting probability can be interpreted as a continuous susceptibility index.

RF is particularly well suited to landslide susceptibility mapping for several reasons. It handles both continuous and categorical variables without requiring discretisation (Dwivedi & Jain, 2025; Nhu et al., 2020). It is robust to correlated and redundant inputs (Kalantar et al., 2020) and outliers (Pourghasemi et al., 2020). It provides an internal measure of variable importance (Meena et al., 2022; Nhu et al., 2020). Furthermore, RF has consistently demonstrated strong performance in comparative studies of machine learning models for landslide susceptibility (Ageenko et al., 2022; Kalantar et al., 2020; Ullah et al., 2024), achieving high prediction accuracies across diverse study areas (Adnan et al., 2020; Nhu et al., 2020; Nnanwuba et al., 2022).

The model was trained on a presence–absence dataset using conditioning factors as predictors, treating landslide susceptibility as a binary classification between landslide and non-landslide locations (Ageenko et al., 2022; Kalantar et al., 2020; Rabby et al., 2023). Key hyperparameters—the number of trees and the number of predictors sampled at each split—were tuned on the training subset (Dwivedi & Jain, 2025; Nnanwuba et al., 2022; SP. et al., 2026). The trained model was then applied across the entire study area to produce a continuous susceptibility surface. For reporting and exposure analysis, this surface was reclassified into five susceptibility classes (Very Low, Low, Moderate, High, Very High) using the natural breaks classification scheme (Nhu et al., 2020; Rabby et al., 2023; SP. et al., 2026).

3.5 Accuracy assessment

Model performance was evaluated on the independent test subset using the Receiver Operating Characteristic (ROC) curve and the associated Area Under the Curve (AUC). The AUC is a threshold-independent measure that summarises the trade-off between the true-positive and false-positive rates across all decision thresholds, taking values from 0.5 (no skill) to 1.0 (perfect discrimination); values above 0.8 are conventionally interpreted as indicating good discriminative performance in landslide susceptibility studies (Ageenko et al., 2022; Reichenbach et al., 2018 [add to references if not present]; Rabby et al., 2023).

3.6 Building infrastructure and exposure analysis

Building exposure was quantified by spatially intersecting individual building footprints with the reclassified susceptibility surface. Each building was assigned the susceptibility class that covered the majority of its footprint, and the resulting attribute was used to compute the number and percentage of buildings in each class. Following the study objective, the analysis emphasised buildings located within the high and very high susceptibility zones, where damaging landslide impacts are most probable (Kincey et al., 2024; Sattar et al., 2025). The proportions of the study area and of the building stock falling within each susceptibility class are reported at the study-area level.

4. Results

4.1 Landslide susceptibility map

The Random Forest model produced a continuous susceptibility surface across the study area, which was reclassified into five classes for reporting (Figure 4). Very Low and Low susceptibility classes are dominant on gentle valley-floor and lower-slope terrain, whereas High and Very High susceptibility classes are concentrated on steep hillslopes, along road corridors, and in the vicinity of lineaments and incised drainage lines — a pattern that is consistent with field understanding of rainfall-triggered slope failure in the Eastern Himalaya. The High susceptibility class alone occupies approximately 24% of the study area, indicating that a substantial fraction of the region is intrinsically prone to landslide initiation.

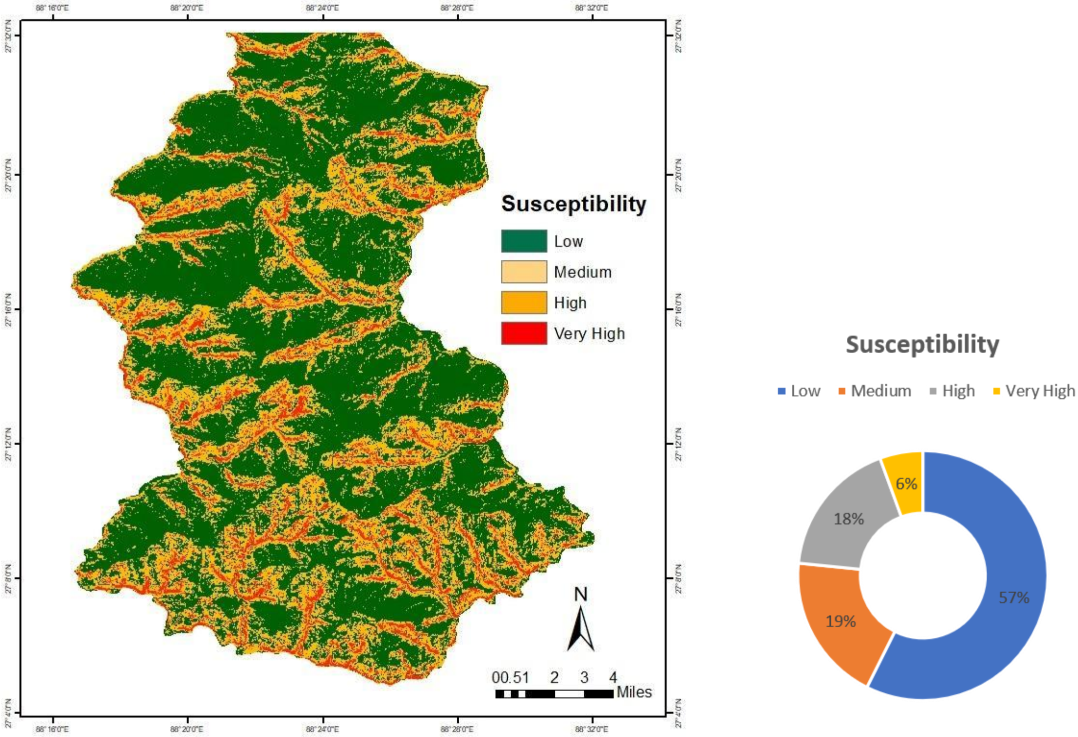

Figure 4. Landslide susceptibility map of the South Sikkim produced by the Random Forest model, classified into five susceptibility classes (Low to Very High).

4.2 Model accuracy

The Random Forest model achieved an AUC of 0.802 on the independent test subset (Figure 5). This value lies within the range generally interpreted as "good" discriminative performance and is comparable to AUC values reported for Random Forest based landslide susceptibility studies elsewhere in the Himalaya. The ROC curve indicates that the model attains a high true positive rate at moderate false positive rates, supporting the reliability of the derived susceptibility surface for subsequent exposure analysis.

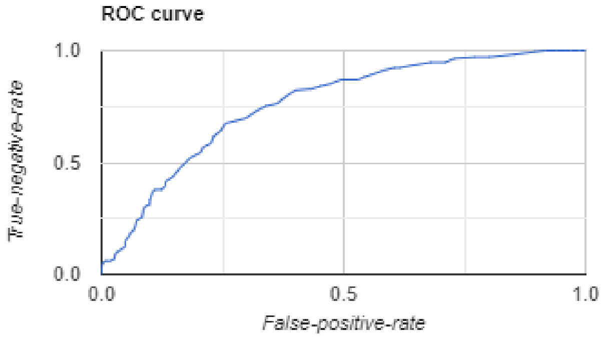

Figure 5. ROC curve for the Random Forest based landslide susceptibility model. Area Under the Curve (AUC) = 0.802.

4.3 Exposure of building infrastructure

Intersecting the building layer with the reclassified susceptibility surface reveals a pronounced concentration of the built environment on hazardous terrain (Table 2). While the high-susceptibility class covers 24 % of the study area, it contains 38 % of the buildings — a 14-percentage-point gap that indicates systematic over-representation of the building stock in the most landslide-prone parts of the landscape. This concentration reflects the geography of habitable ground in South Sikkim: the sites accessible from roads, gently enough graded for cut-and-fill construction, and close to water are the same sites at which the physical drivers identified by the Random Forest model , steep gradient, high wetness index, and road proximity, also converge.

Table 2. Areal share of susceptibility classes and share of buildings

Susceptibility class	Share of area (%)	Share of buildings (%)
High susceptibility	24	38

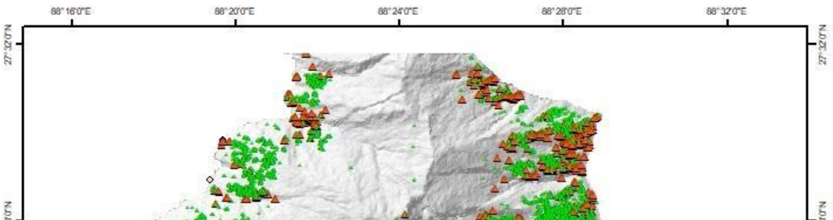

Figure 6. Distribution of buildings across landslide susceptibility classes in the South Sikkim, highlighting the buildings highly exposed.

5. Discussion

5.1 Model performance and factor relevance

The AUC of 0.802 places the present model in the "good" performance band and squarely within the range reported for machine-learning susceptibility studies elsewhere in Sikkim: below the CNN AUC of 0.918–0.933 obtained by Saha et al. (2021) in East Sikkim on a larger training set, close to the hybrid AHP–FR AUC of 0.81 reported for the state by M. R. Das et al. (2026), and below the Gradient Boosting Machine AUC of 0.99 reported by Roy et al. (2025) — noting that AUC comparisons across studies must be qualified by differences in inventory density, sample balance, and spatial-block validation. The value obtained here is therefore consistent with the wider Sikkim literature without approaching the near-perfect scores that typically indicate spatial-autocorrelation optimism (Reichenbach et al., 2018 [if in library]). Spatially, the concentration of the High class along steep hillslopes, road corridors, and incised channels reproduces the pattern identified by Nawazuzzoha et al. (2025) for the same district, in which distance to roads and rainfall emerged as the leading conditioning factors. In our RF model the same slope-and-road signature is reconstructed from a smaller, physically motivated predictor set of ten variables — evidence that a compact predictor set is sufficient to reproduce the dominant susceptibility geography of South Sikkim, and that additional variables in earlier studies were probably serving redundant physical roles.

5.2 Over-exposure of the building stock

The central finding — that the high-susceptibility class holds 38 % of the building stock while occupying only 24 % of the area — is best understood as a signature of Himalayan settlement geography rather than as a statistical artefact. In steep terrain, habitable ground is scarce, and the criteria that make a site habitable (road access, moderate rather than extreme gradient, proximity to water) coincide with several of the physical drivers that the RF model identifies as diagnostic of susceptibility. Exposure is therefore not distributed in proportion to land area but is amplified in the very zones where hazard is highest. This mechanism is consistent with the exposure trajectory reported by Sattar et al. (2025) along the Teesta corridor, where building exposure grew faster than the underlying hazard footprint between 2013 and 2023, and with the risk-vulnerability pattern documented for Gangtok by Zimik et al. (2022), in which the very high and high-risk classes together occupy 25.7 % of area — a value close to the 24 % reported here for South Sikkim, suggesting a broadly consistent hazard geography across the state. Because the building stock in high-susceptibility zones concentrates essential services

— health facilities, schools, administrative offices, and connective infrastructure — even localised failures can propagate rapidly into settlement-scale disruption, consistent with the cascading-impact framing set out by Kato et al. (2021).

5.3 Implications for construction regulation and risk reduction

Because 38 % of the building stock is already located within high-susceptibility zones, risk reduction in South Sikkim must proceed on two parallel tracks. Retrospective mitigation — slope stabilisation, drainage improvement, structural retrofitting of the most exposed clusters, and, where the risk-benefit balance justifies it, planned relocation — is unavoidable for the existing stock. In parallel, hazard-informed regulation of new construction is required to prevent further concentration of assets and population on the most hazardous slopes, echoing the recommendations of Bansal et al. (2022) and Yadav and Kumar (2023) that susceptibility and microzonation maps should be embedded in land-use control at the block level. The map produced here is at a scale appropriate for such screening: it can be adopted directly by the district planning authority and by the State Disaster Management Authority to filter proposed construction sites, and the 38 % exposure share can serve as a monitored indicator of whether new construction is worsening or improving the state of exposure over time. Extension of the framework to the NH-10 corridor and to inventories of critical facilities (schools, hospitals, government offices) would translate the exposure metric into an operational input for emergency-response planning.

5.4 Limitations and future work

Several limitations qualify these results. First, the landslide inventory is drawn from heterogeneous sources and inevitably under-represents older, vegetated, or small failures; the susceptibility model is therefore conditioned by the observable rather than the complete population of landslides. Second, the ten conditioning factors, though widely used, do not explicitly represent event-specific rainfall triggers; introducing rainfall thresholds and antecedent moisture would move the model from static susceptibility towards event-based hazard. Third, the exposure analysis reported here is two-dimensional and treats all buildings as equivalent. Future work will extend the framework in three directions: (i) attributing each building with height and use type to produce a three-dimensional exposure surface expressed in floor space and population; (ii) incorporating runout modelling so that buildings downslope of unstable ground, and not only those on unstable ground, are counted as exposed; and (iii) assessing the risk associated with new building construction as an operational planning tool.

6. Conclusions

This study moves the analysis of landslide risk in South Sikkim from a hazard-only to a hazard-plus-exposure framing. A Random Forest susceptibility model built on ten physically motivated conditioning factors (AUC = 0.802) was intersected with a settlement-scale building layer to produce, to our knowledge, the first quantitative building-exposure assessment for the region. The key result — that the high-susceptibility class occupies 24 % of the area but contains 38 % of the buildings — is not merely descriptive: it identifies a structural feature of Himalayan settlement geography, in which the scarcity of habitable ground concentrates buildings on the same sites that concentrate physical hazard. Three implications follow. Scientifically, exposure cannot be inferred from land area and must be measured directly on the building layer, or it will be systematically underestimated. Practically, existing settlements require dual-track mitigation combining retrospective stabilisation with hazard-informed regulation of new construction, and the map produced here can be inserted directly into district-level planning workflows. Methodologically, the compact ten-factor RF pipeline is deliberately transferable to other data-scarce Himalayan settlements, in Sikkim and beyond. Future work will attribute each building with height, use-type and occupancy to produce a three-dimensional exposure surface, incorporate runout modelling so that downslope buildings are also counted as exposed, and formalise the risk of new building construction as an operational planning indicator.

Acknowledgements

The author gratefully acknowledges the Department of Geography, School of Earth, Biological and Environmental Sciences, Central University of South Bihar, Gaya, for institutional support. The citizen-monitoring platform Save The Hills is acknowledged for documentation of recurrent landslides in South Sikkim.

Author Contributions

Shivam Priyadarshi: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, writing — original draft.

Data Availability Statement

The landslide inventory, conditioning factor rasters, Random Forest susceptibility outputs, and building layer used in this study are available from the corresponding author on reasonable request.

Declaration of Competing Interests

The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adnan, M. S. G., Rahman, S., Ahmed, N., Ahmed, B., Rabbi, Md. F., & Rahman, R. M. (2020). Improving Spatial Agreement in Machine Learning-Based Landslide Susceptibility Mapping. Remote Sensing, 12(20), 3347. https://doi.org/10.3390/rs12203347

Ageenko, A., Hansen, L. C., Lyng, K. L., Bodum, L., & Arsanjani, J. J. (2022). Landslide Susceptibility Mapping Using Machine Learning: A Danish Case Study. Isprs International Journal of Geo-Information, 11(6), 324. https://doi.org/10.3390/ijgi11060324

Bansal, B. K., Verma, M., Gupta, A. K., & Prasath, R. A. (2022). On mitigation of earthquake and landslide hazards in the eastern Himalayan region. Natural Hazards, 114(2), 1079–1102. https://doi.org/10.1007/s11069-022-05448-y

Beigh, I. H., & Bukhari, S. K. (2024). Landslide susceptibility assessment using GIS-based multicriteria decision analysis (MCDA) along a part of National Highway-1, Kashmir- Himalayas, India. Applied Geomatics, 16(2), 451–469. https://doi.org/10.1007/s12518-024-00559-6

Bhasin, R., Aslan, G., & Dehls, J. (2023). Ground Investigations and Detection and Monitoring of Landslides Using SAR Interferometry in Gangtok, Sikkim Himalaya. Geohazards, 4(1), 25–39. https://doi.org/10.3390/geohazards4010003

Chen, T. K., Kincey, M., Rosser, N., & Seto, K. C. (2024). Identifying recurrent and persistent landslides using satellite imagery and deep learning: A 30-year analysis of the Himalaya. The Science of the Total Environment, 922, 171161. https://doi.org/10.1016/j.scitotenv.2024.171161

Chen, C., Shen, Z., Weng, Y., You, S., Lin, J., Li, S., & Wang, K. (2023). Modeling Landslide Susceptibility in Forest-Covered Areas in Lin'an, China, Using Logistical Regression, a Decision Tree, and Random Forests. Remote Sensing, 15(18), 4378. https://doi.org/10.3390/rs15184378

Das, J., Saha, P., Mitra, R., Alam, A., & Kamruzzaman, Md. (2023). GIS-based data-driven bivariate statistical models for landslide susceptibility prediction in Upper Tista Basin, India. Heliyon, 9(5), e16186. https://doi.org/10.1016/j.heliyon.2023.e16186

Das, M. R., Gautam, G. K., Jain, S., Bhat, M. F., Mankar, A. K., & Koner, R. (2026). A comparative analysis of AHP, FR, AHP-FR and LR models for landslide susceptibility mapping in Sikkim Himalaya, India. Earth Surface Processes and Landforms, 51(2). https://doi.org/10.1002/esp.70257

Dikshit, A., Sarkar, R., Pradhan, B., Segoni, S., & Alamri, A. (2020). Rainfall Induced Landslide Studies in Indian Himalayan Region: A Critical Review. Applied Sciences, 10(7), 2466. https://doi.org/10.3390/app10072466

Dubey, S., Sattar, A., Goyal, M. K., Allen, S., Frey, H., Haritashya, U. K., & Huggel, C. (2023). Mass Movement Hazard and Exposure in the Himalaya. Earth S Future, 11(9). https://doi.org/10.1029/2022ef003253

Dwivedi, M. K., & Jain, K. (2025). Unveiling the layers of decision-making in landslide susceptibility: a critical exploration of multi-criteria analysis, bivariate statistics, and knowledge-based approaches for complex relationships of factors and their weightage. https://doi.org/10.21203/rs.3.rs-6750285/v1

Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. A., & Shabani, F. (2020). Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. Remote Sensing, 12(11), 1737. https://doi.org/10.3390/rs12111737

Kato, T., Rambali, M., & Blanco, V. M. F. (2021). Strengthening climate resilience in mountainous areas. https://doi.org/10.1787/1af319f0-en

KC, D., Dangi, H., & Hu, L. (2022). Assessing Landslide Susceptibility in the Northern Stretch of Arun Tectonic Window, Nepal. Civileng, 3(2), 525–540. https://doi.org/10.3390/civileng3020031

Kincey, M., Rosser, N., Swirad, Z., Robinson, T., Shrestha, R., Pujara, D. S., Basyal, G. K., Densmore, A. L., Arrell, K., Oven, K., & Dunant, A. (2024). National-Scale Rainfall-Triggered Landslide Susceptibility and Exposure in Nepal. Earth S Future, 12(2). https://doi.org/10.1029/2023ef004102

Koley, B., Nath, A., Saraswati, S., Bandyopadhyay, K., & Ray, J. K. (2019). Assessment of Rainfall Thresholds for Rain-Induced Landslide Activity in North Sikkim Road Corridor in Sikkim Himalaya, India. Journal of Geography Environment and Earth Science International, 1–14. https://doi.org/10.9734/jgeesi/2019/v19i330086

Krishna, A. P. (2025). Sikkim Glacial Lake Outburst Flood (GLOF) Disaster of 2023: Geoinformatics- Based Assessment of Causes and Impact. https://doi.org/10.21203/rs.3.rs-7405556/v1

Kumar, A., Asthana, A. K. L., Priyanka, R. S., Jayangondaperumal, R., Gupta, A. K., & Bhakuni, S. S. (2017). Assessment of landslide hazards induced by extreme rainfall event in Jammu and Kashmir Himalaya, northwest India. Geomorphology, 284, 72–87. https://doi.org/10.1016/j.geomorph.2017.01.003

Laha, C., Chattoraj, S., Prasain, G. P., & Sharma, P. (2024). Risk Resilience of Growing Settlements in Landslide Prone Hilly Areas: Case Study on Kalimpong-I Block, Darjeeling District, West Bengal. https://doi.org/10.21203/rs.3.rs-3676394/v1

Meena, S. R., Puliero, S., Bhuyan, K., Floris, M., & Catani, F. (2022). Assessing the importance of conditioning factor selection in landslide susceptibility for the province of Belluno (region of Veneto, northeastern Italy). Natural Hazards and Earth System Science, 22(4), 1395–1417. https://doi.org/10.5194/nhess-22-1395-2022

Naik, D. D., & Palakuzhiyil, Y. (2024). Landslide Susceptibility Mapping using Geospatial framework -A Study of Pinder Basin, Chamoli District, Uttarakhand, India. https://doi.org/10.21203/rs.3.rs-4006670/v1

Nawazuzzoha, M., Rashid, Md. M., Iqbal, F., Suheb, Shakeel, A., Arabameri, A., & Naqvi, H. R. (2025). Advanced Machine Learning Models for Landslide Susceptibility Mapping in South Sikkim Himalayas, India. Geological Journal, 61(2), 371–387. https://doi.org/10.1002/gj.5234

Nhu, V., Shirzadi, A., Shahabi, H., Chen, W., Clague, J. J., Geertsema, M., Jaafari, A., Avand, M., Miraki, S., Asl, D. T., Pham, B. T., Ahmad, B. B., & Lee, S. (2020). Shallow Landslide Susceptibility Mapping by Random Forest Base Classifier and Its Ensembles in a Semi-Arid Region of Iran. Forests, 11(4), 421. https://doi.org/10.3390/f11040421

Nnanwuba, U. E., Qin, S., Adeyeye, O. A., Cosmas, N. C., Yao, J., Qiao, S., & Egwuonwu, E. M. (2022). Prediction of Spatial Likelihood of Shallow Landslide Using GIS-Based Machine Learning in Awgu, Southeast/Nigeria. Sustainability, 14(19), 12000. https://doi.org/10.3390/su141912000

Pandey, B. W., & Prasad, A. S. (2018). Slope vulnerability, mass wasting and hydrological hazards in himalaya: a case study of Alaknanda Basin, Uttarakhand. Terrae Didatica, 14(4), 395–404. https://doi.org/10.20396/td.v14i4.8654111

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerdà, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-60191-3

Rabby, Y. W., Li, Y., & Hilafu, H. (2023). An objective absence data sampling method for landslide susceptibility mapping. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-28991-5

Rafique, M. F., & Joshi, V. (2024). Information Value Model Based Mapping of Updated Spatial and Temporal Landslide Susceptibility: A Case Study from East Sikkim District, India's Northeastern Himalayas. https://doi.org/10.20944/preprints202404.0066.v1

Rajamani, S. K., & Iyer, R. S. (2024). The Phenomenon of Eco-Anxiety and Distress Related to Climate Change and Environmental Degradation. 156–177. https://doi.org/10.4018/979-8-3693-2177-5.ch011

Roy, S. K., Dey, S., Das, J., Hossen, B., Das, S., Hasan, Md. M., & Mojumder, P. (2025). Utilising Machine Learning Approaches for Enhanced Landslide Susceptibility Mapping in Sikkim, India. Geological Journal, 60(5), 1150–1169. https://doi.org/10.1002/gj.5198

Saha, S., Roy, J., Hembram, T. K., Pradhan, B., Dikshit, A., Maulud, K. N. A., & Alamri, A. (2021). Comparison between Deep Learning and Tree-Based Machine Learning Approaches for Landslide Susceptibility Mapping. Water, 13(19), 2664. https://doi.org/10.3390/w13192664

Sarkar, S., & Saha, K. (2026). Evaluating the influence of anthropogenic hillslope modulation in relation to shallow landslides in the Rishikhola region, Darjeeling–Sikkim Himalaya, India. Scientific Reports, 16(1). https://doi.org/10.1038/s41598-025-30543-y

Sattar, A., Cook, K., Kant, S., Berthier, É., Allen, S., Rinzin, S., Vries, M. V. W. de, Haeberli, W., Kushwaha, P., Shugar, D. H., Emmer, A., Haritashya, U. K., Frey, H., Gurudin, K. S. K., Rajak, R., Hossain, F., Huggel, C., Mergili, M., Azam, M. F., … Bhat, S. Y. (2025). The Sikkim flood of October 2023: Drivers, causes, and impacts of a multihazard cascade. Science, 387(6740). https://doi.org/10.1126/science.ads2659

Shahi, Y. B., Kadel, S., Dangi, H., Adhikari, G. P., KC, D., & Paudyal, K. R. (2022). Geological Exploration, Landslide Characterization and Susceptibility Mapping at the Boundary between Two Crystalline Bodies in Jajarkot, Nepal. Geotechnics, 2(4), 1059–1083. https://doi.org/10.3390/geotechnics2040050

Sharma, A., Schwartz, S. M., & Méndez, E. (2013). Hospital volume is associated with survival but not multimodality therapy in Medicare patients with advanced head and neck cancer. Cancer, 119(10), 1845–1852. https://doi.org/10.1002/cncr.27976

SP., D., Das, J., R, J., TM, S., V, S., & RS, L. (2026). Integrating Geo-Environmental Factors and Ensemble Machine Learning for Landslide Susceptibility Assessment in Kodaikanal and its Environs, India. https://doi.org/10.21203/rs.3.rs-8113160/v1

Ullah, M., Tang, B., Huangfu, W., Yang, D., Wei, Y., & Qiu, H. (2024). Machine Learning-Driven Landslide Susceptibility Mapping in the Himalayan China–Pakistan Economic Corridor Region. Land, 13(7), 1011. https://doi.org/10.3390/land13071011

Uniyal, A. (2010). Disaster management strategy for potential slide zones of Kumarkhera in Narendra Nagar township of Tehri Garhwal district, Uttarakhand, India. Disaster Prevention and Management an International Journal, 19(3), 358–369. https://doi.org/10.1108/09653561011052529

Yadav, P., & Kumar, S. (2023). CAUSES, IMPACTS & MITIGATION MEASURES FOR FLASH FLOODS IN INDIAN HIMALAYAN REGION. Interantional Journal of Scientific Research in Engineering and Management, 07(04). https://doi.org/10.55041/ijsrem19803

Zimik, H. V., Angchuk, T., Misra, A. K., Ranjan, R. K., Wanjari, N., & Basnett, S. (2022). GIS-based identification of potential watershed recharge zones using analytic hierarchy process in Sikkim Himalayan region. Applied Water Science, 12(11). https://doi.org/10.1007/s13201-022-01758-5