Journal of Engineering Issues and Solutions

Geospatial Integration of Landslide Susceptibility and Road Network Vulnerability in Mountainous Terrain

Buddhi Raj Joshi[1], Netra Prakash Bhandary[2*], Indra Prasad Acharya[3], Niraj K.C.[4], Chakra Bhandari[5]

[1]*School of Engineering, Faculty of Science and Technology, Pokhara University, Kaski 33700, Nepal*
Email: buddhi.rajjoshi@pu.edu.np
[2]*Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime, 790-8577, Japan*
Email: netra@ehime-u.ac.jp
[3]*Department of Civil Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Lalitpur 44700, Nepal*
Email: indrapd@ioe.edu.np
[4]*Department of Geomatics Engineering, Pashchimanchal Campus, Institute of Engineering, Tribhuvan University, Kaski 33700, Nepal*
Email: niraj@wrc.edu.np
[5]*Madan Bhandari College of Engineering, Madan Bhandari Memorial Academy Nepal, Urlabari 56604, Nepal*
Email: chakra.bs@mbman.edu.np
** Correspondence: netra@ehime-u.ac.jp*

Received: May 6, 2026; Revised: June 3, 2026; Accepted: June 25, 2026

Abstract

Transportation networks in hilly regions are often subject to landslides, which induce large socio-economic losses and impede emergency response and regional communication. Hence, it is important to identify landslide prone locations and evaluate the vulnerability of road infrastructure for sustainable transportation planning and catastrophe high risk reduction. This research develops a geospatial machine learning framework for landslide susceptibility mapping and road network vulnerability assessment in Kaski, Nepal, a mountainous region characterized by complicated topography, intense monsoonal rainfall, and fast land use change. A landslide susceptibility model was formulated by employing the Extreme Gradient Boosting (XG-Boost) method with 12 causative elements, including geomorphological, topographic, hydrological, geological, and anthropogenic variables. The Variance Inflation Factor (VIF) was used to check multicollinearity; all variables were under the permitted limit (VIF < 5). The model showed good predictive performance with 93% accuracy, recall, and F1-score, 98% AUC, and a Kappa value of 0.861. The resulting landslide susceptibility map generated shows that most of the study region (59.69%) falls under very low susceptibility; 12.18% is under high to very high susceptibility with localized hotspots. Furthermore, the feature significance analysis showed the major factors influencing the incidence of landslides were land cover (13.9%), rainfall (13.5%), elevation (12.5%), NDVI (12.2%), and slope (11.7%), whereas curvature was the least relevant component. Spatial overlay of the susceptibility map with the road network suggests that about 693 km (23.3%) of the entire road length falls under high and very high landslide-vulnerable zones. The most vulnerable roads were unclassified roads, with 34.1% of their total length in high and very high susceptibility groups, followed by primary routes (21.7%) and residential roads (8.3%). The results emphasize the significant vulnerability of major road infrastructure and the urgent need for mitigation measures at a site-specific level. The developed geospatial framework is an effective and flexible method for landslide vulnerability assessment, which facilitates informed decision-making for sustainable infrastructure development in landslide-prone mountainous regions.

Keywords: *Landslide susceptibility mapping, XG-Boost model, road network vulnerability, infrastructure vulnerability assessment, disaster risk reduction*

1. Introduction

Landslides are masses of rocks, debris, and earth that roll down a slope owing to gravity. Landslides are recognized as one of the most dangerous phenomena globally, resulting in considerable loss of life and property (Bhattarai & Bhandary, 2026a; Froude & Petley, 2018; Petley, 2012). They are generally driven by a combination of intrinsic causes, including lithology, structure, weathering, and extrinsic factors, such as severe precipitation, seismic activity, or human involvement (Wang et al., 2025). Nepal is also considered a hotspot for landslides due to its steep topography, active tectonics, and heavy rains, apart from being the eleventh most seismically active nation, making it prone to recurring landslides (Bhattarai & Bhandary, 2025). Landslides in the Nepal Himalaya are governed by geography and active thrust faults and are further exacerbated by uncontrolled manmade activities, such as hasty road development and mismanaged urban expansion.

The Kaski District of the Gandaki Province is experiencing frequent landslide events because of the variable topography of mid-hill mountains and high Himalayas as well as complex geological structures and abundant rainfall during the monsoon season (Pathak et al., 2025). The exposure distribution is again broadly distributed in the Middle Hills, with most of the exposure around Pokhara Valley in Kaski District being high (Kincey et al., 2024). The modified frequency ratio models identified elevation, geology, and land cover as significant factors for landslides in the Kaski district but with a moderate predictive capability (AUC: 68.87%) (Niraj Baral et al., 2021).

The resilience of critical transportation infrastructure, particularly road networks, is increasingly challenged by the heightened frequency and intensity of landslide events, a trend intensified by climate change and anthropogenic landscape modifications (Bhandari et al., 2026). Roads serve as the basic social connectivity systems, lifelines of regional economies, and emergency response systems; their disruption due to slope failures can lead to significant economic losses, prolonged detours, and compromised public safety. In Gandaki Province, the Beni Jomsom Road clearly demonstrates the relationship between road construction and recurrent slope instabilities such as translational slides, debris flows, and rockfalls (Roubal, 2014). Similarly, in the Narayanghat-Kathmandu Road corridor, infrastructure design has revealed that critical elements such as bridges, culverts, and retaining walls are often situated within hazardous zones (K. K. Sharma et al., 2025) and require immediate action for their management. While roads are essential for last-mile connectivity, they are among the most vulnerable infrastructure elements because they are directly exposed to landslides and other slope failures (Adhikari & Dhakal, 2026).

Consequently, developing robust methodologies for assessing road vulnerability to landslides has become a paramount concern for geoscientists, transportation planners, and disaster risk reduction agencies. Traditional approaches to landslide risk assessment have often treated the hazard (the landslide) and the element at risk (the road network) as separate analytical domains (Chowdhuri et al., 2022; Yalcin, 2008). While these maps are effective for land-use zoning and regional hazard planning, they are rarely constructed with the explicit geometric and topological characteristics of road networks. Conversely, road vulnerability frequently relies on empirical damage data or localized slope stability analyses, which lack the spatial continuity and predictive power required for road network-wide planning. Bridging this gap requires a methodological shift from viewing the road proximity as a landslide causative factor within a susceptible zone to understanding the overlapping relationship between the linear infrastructure and the continuous susceptible zone.

While existing research has focused either on Landslide Susceptibility Mapping (LSM) in the Nepal Himalaya (Joshi et al., 2026) along with landslide risk assessment (Joshi et al., 2025) or on slope instabilities along road corridors (Robson et al., 2025), limited attention has been given to explicitly integrating landslide susceptibility with transportation network vulnerability assessment. Continuous advancements in landslide susceptibility mapping have demonstrated that machine learning approaches outperform conventional statistical methods (Lee & Pradhan, 2007; Pham et al., 2016; Chen et al., 2017; Pourghasemi & Rossi, 2017; Ullah et al., 2025). Extreme Gradient Boosting (XGBoost) has been demonstrated to be particularly effective (Chowdhuri et al., 2022) in terms of accuracy, lack of overfitting, and its ability to handle complex interactions among the predictor variables (Pawłuszek-Filipiak & Lewandowski, 2025).

This study develops a novel framework for assessing road vulnerability by explicitly integrating a high-resolution road network with a multi-class landslide susceptibility map. By explicitly linking susceptibility zones to road segments, the study addresses a key gap in literature, offering a more operationally relevant framework for infrastructure planning and risk mitigation. The Sendai Framework for Disaster Risk Reduction emphasizes the necessity of evidence-based approaches to reduce infrastructure vulnerability in disaster-prone regions (UNDRR, 2015). This study addresses these gaps in three ways: (1) by identifying the main factors controlling landslide occurrence and distribution; (2) by developing a reliable landslide susceptibility model using the XGBoost machine learning approach, supported by field-validated landslide data; and (3) by assessing the vulnerability of the road network to landslide hazards. The findings will support road maintenance and investment planning for mountain roads in Nepal.

2. Study Area

Kaski District lies in the Gandaki Province of Nepal with latitudes between 28° 15' 00" and 28° 30' 00" and longitudes between 83° 45' 00" and 84° 15' 00" (Fig. 1). The district is characterized by extremely rugged topography, ranging from the low-lying Pokhara Valley (≈450 masl) to the high peaks of the Annapurna range (>8,000 m). Elevations above 4,599 m produce pronounced climatic and ecological variations that significantly influence landslide occurrence.

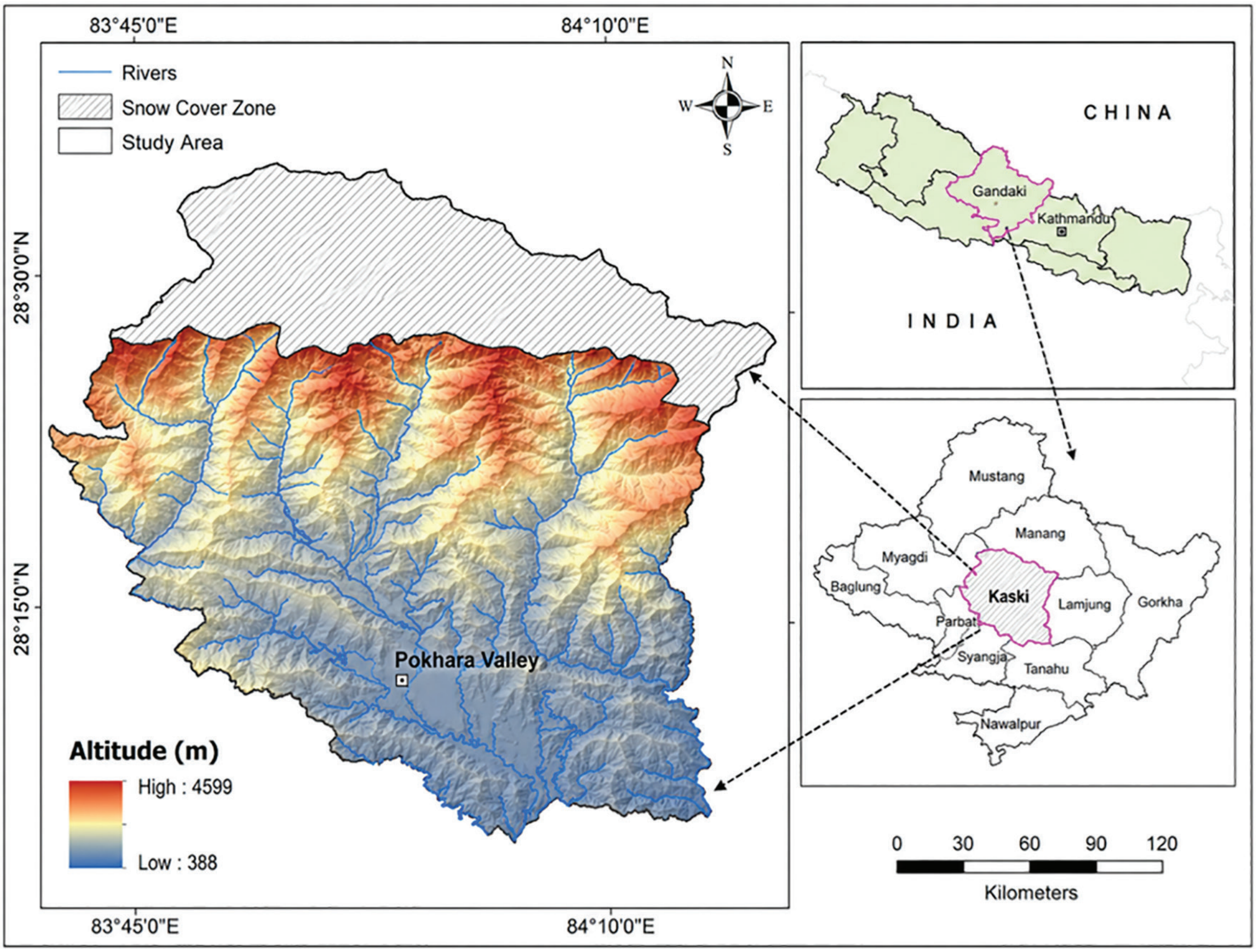

Figure 1: Geographical location map of the study area along with the digital elevation model and topographic features (rivers and streams).

Geologically, the district spans both the Lesser and Higher Himalayan zones, with the Main Central Thrust (MCT), a major tectonic boundary, passing through the area and juxtaposing low-grade metamorphic rocks of the Lesser Himalaya against high-grade metamorphic rocks and gneisses of the Higher Himalaya (Searle et al., 2008). Dominant lithologies include phyllites, schists, quartzites, gneisses, and migmatites, which vary in their degree of weathering and fracturing, thereby influencing slope stability. The region experiences an extreme monsoonal climate, with annual rainfall ranging from 3,000 to 4,500 mm, concentrated during the monsoon season (June to September), which frequently triggers shallow landslides and debris flows (Basnet et al., 2020). The combination of steep slopes, complex geology, intense rainfall, and ongoing road expansion makes the area highly susceptible to landslides.

3. Materials and Methods

This study compiled and integrated diverse geospatial datasets sourced from national and international repositories, selected based on an extensive literature review (Table 1).

Table 1: Description of the data types used in this study along with their format and sources

S.N.	Data	Source	Format	Use
1	DEM	ASTER Dem	Raster, 30 m	Geo-morphological & Topographic features
2	Rainfall (from 2000 to 2025)	Department of Hydrology & Meteorology (DHM)	Time-series	Precipitation & soil erosivity Analysis
3	Geology	Department of Mines & Geology (DMG), 1994	Vector	Lithological classification
4	Soil Types	SOTER Nepal, FAO (2004)	Vector	Soil classification
5	Landsat - 8 images	USGS / NASA / GEE	Raster	NDVI Analysis
6	LULC	ESRIC, ICIMOD	Raster	
7	Road datasets	Open Street Map, Google earth digitization	Vector	Infrastructure overlay
8	Landslide Inventory data (total 646)	DMG records (371), Google earth digitization, and Field investigation	Vector: polygon	LSM development, and Road vulnerability assessment

3.1 Landslide Inventory

Landslide inventories are fundamental datasets in landslide susceptibility studies, as they provide essential information on the spatial and temporal distribution of past landslide events. A landslide inventory typically consists of mapped locations, types, dates, and characteristics of landslides (Joshi et al., 2026). The quality, completeness, and accuracy of inventories significantly influence the reliability of susceptibility and hazard assessments. In this study, a total of 646 landslide polygons were compiled, including 371 obtained from the Department of Mines and Geology (DMG) and the remainder digitized from Google Earth imagery. Among these, 405 landslides located within a 200 m buffer of the road network were found to affect road infrastructure (Fig. 2).

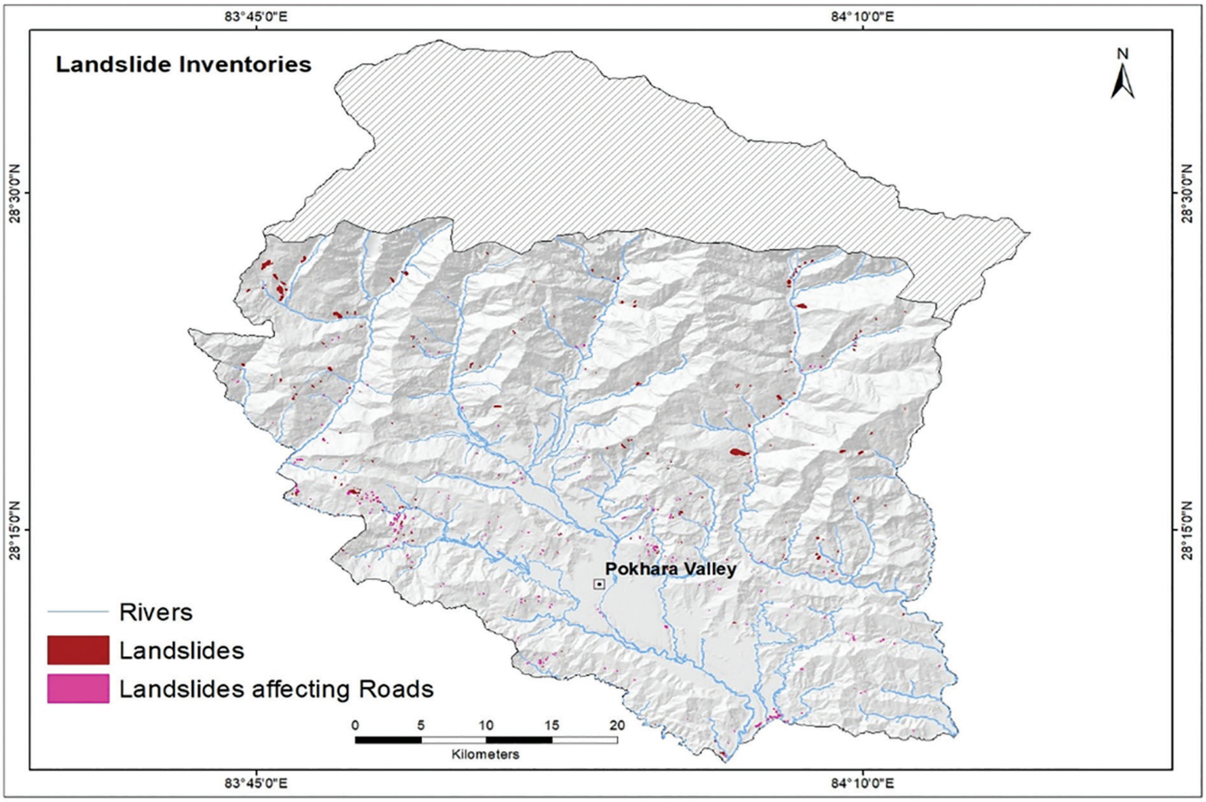

Figure 2: Integrated landslide inventory compiled from historical records, Google Earth digitization, and field surveys, including landslides affecting road networks (shown in pink).

3.2 Landslide Causative Factors

Reliable landslide susceptibility fundamentally relies on the correct identification and analysis of causative factors (Reichenbach et al., 2018). This study considers twelve causative factors based on specific site conditions (Chowdhury, 2023; Singh et al., 2024) along with their documented influence on landslide occurrence in the Nepal Himalaya (Gautam et al., 2021; Huang & Chen, 2024), which includes annual rainfall, elevation, slope, curvature, river proximity, aspects, topographic wetness index (TWI), normalized difference vegetation index (NDVI), road proximity, geology, soil types, and land use land cover (LULC).

The Department of Hydrology and Metrology (DHMG) of Nepal provides the rainfall data to analyze the precipitation frequency and intensity of landslide occurrences. Long-term annual rainfall defines the baseline climate, but intensity and duration of individual storms are even more critical for event-based landslide susceptibility mapping (Dahal & Hasegawa, 2008). The 30 m resolution ASTER satellite digital elevation model (DEM) of the study area was extracted to create various topographical features, such as elevation, slope, aspect, and TWI. Elevation itself doesn't cause landslides, but it acts as a powerful driver for other factors, including slopes, rainfall, and geology. The slope data help examine the steep gradients and unstable slope characteristics in the area. Aspect controls exposure to sunlight, wind, and prevailing rainfall. In many regions, south-facing slopes (Northern Hemisphere) are drier and more weathered, while north-facing slopes retain more moisture, increasing pore water pressure and influencing vegetation type and root cohesion. The topographic wetness index (TWI) examines drainage conditions, saturation levels, and erosion factors. TWI identifies areas where water is likely to accumulate based on upstream contributing areas and local slopes.

Proximity to rivers influences slope stability due to undercutting and saturation levels and shows that being nearer to a river leads towards frequent slope failure. NDVI quantifies vegetation health and density. From NDVI indices, we can determine how vegetation affects slope stability, resilience, and vulnerability when considering landslide risk by analyzing the role of vegetation. High NDVI (dense forest) typically correlates with lower susceptibility due to root cohesion. Low NDVI (sparse/bare soil) correlates with higher susceptibility. Anthropogenic factors focused on proximity to roads, calculated as Euclidean distance from the district's road network and classified into buffer zones reflecting the zone of influence of road construction on slope stability.

Likewise, information on soil composition combined with moisture content and stability assessment was derived from the SOTER database of Nepal, while rock types, permeability, and lineament were extracted from the archives of Nepal's Department of Mines and Geology (DMG) (DMG., 2022). Geology is a primary controlling factor for landslides, in which different rock types have different strengths, weathering rates, and permeability. Weak rock seems highly susceptible, while strong rocks are generally stable. The orientation of bedding planes, faults, and foliation highly influences instability, in which dips parallel to the slope are critically dangerous. Each soil type has unique geotechnical properties that lead to shear failure, weathering, deep crack production, rapid strength loss, perched water table, and liquefaction. LULC represents the combined effect of human activity and vegetation. Forests generally increase stability through root reinforcement and evapotranspiration, while deforestation drastically increases landslide susceptibility. The barren/agricultural land has a much higher susceptibility due to a lack of deep roots and soil disturbance, while urban areas shows complex effects. Excavating foundations can create instability, but retaining walls and drainage can help stabilize slopes locally. Fig. 3 presents the landslide causative factors used in this study.

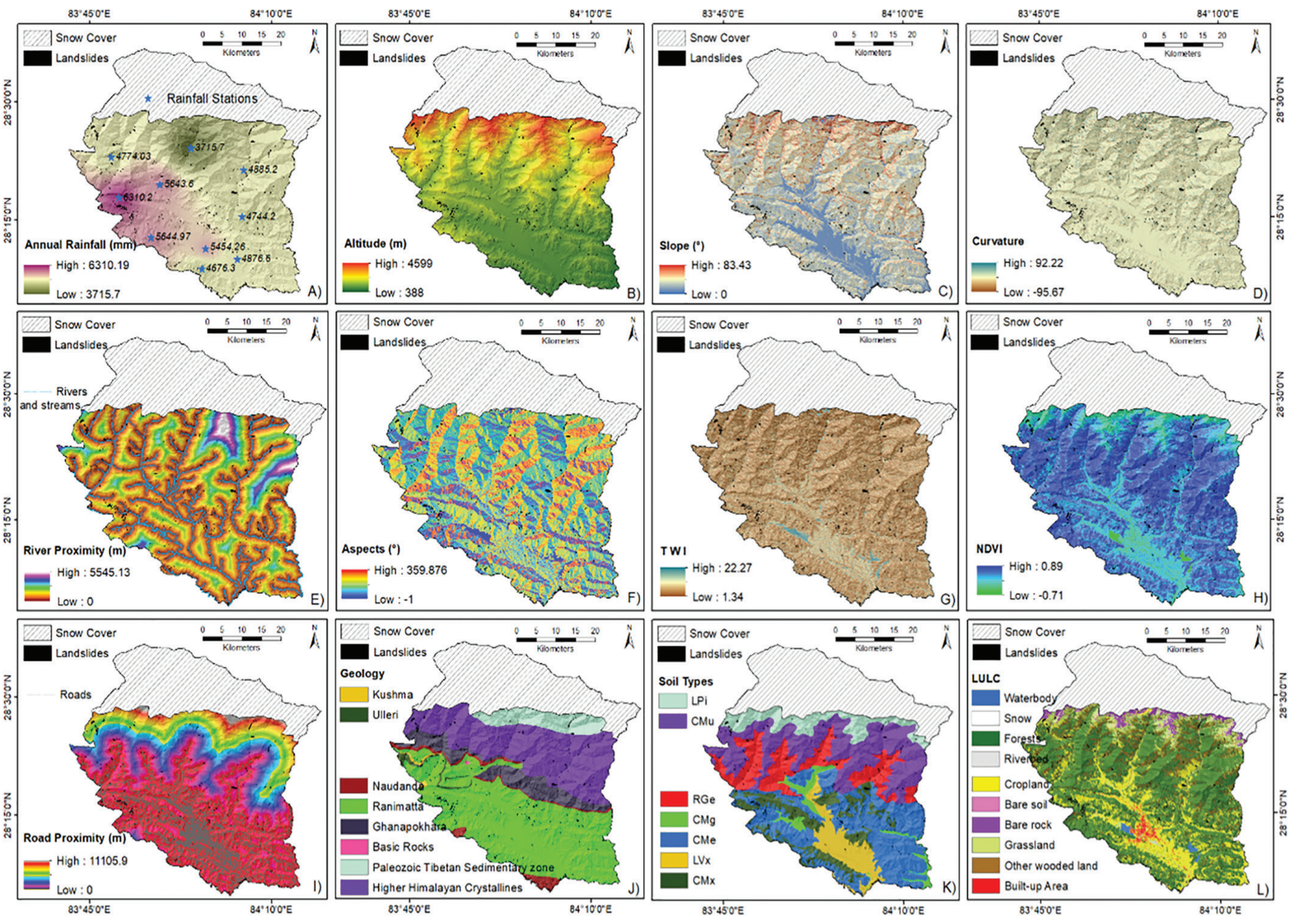

Figure 3: Spatial distribution of the landslide conditioning factors considered in the atudy: A) Annual Rainfall, B) Elevation, C) Slope, D) Curvature, E) River Proximity, F) Aspects, G) TWI, H) NDVI, I) Road Proximity, J) Geology, K) Soil Types (LPi: Gelic Leptosols, CMu: Humic Cambisols, RGe: Eutric Regosols, CMg: Gleyic Cambisols, CMe: Eutric Cambisols, LVx: Haplic Luvisols, & CMx: Chromic Cambisols), and L) LULC

3.3 Data Sampling and Preprocessing

In this study, a pixel-based sampling approach was adopted to prepare landslide and non-landslide datasets for susceptibility modeling. A total of 646 landslide polygons collected from different sources (Table 1) were transformed into 3,442 landslide pixels (30 m resolution). An equal number of non-landslide pixels were randomly generated from stable terrain, specifically within buffer zones located 300 m away from existing landslides, to minimize spatial overlap and ensure reliable representation of stable conditions (Joshi et al., 2026). The final dataset was assigned to binary classes, where landslide pixels were labeled as 1 and non-landslide pixels as 0. The combined dataset was then randomly divided into training (70%) and testing (30%) subsets using simple random sampling to reduce spatial bias and support robust model validation. Before model development, multicollinearity among the 12 selected landslide conditioning factors was evaluated using the Variance Inflation Factor (VIF) test (Eq. 1). A sequential backward elimination approach was applied, in which variables exceeding the VIF threshold of 5 were iteratively removed, and VIF values were recalculated until all remaining predictors satisfied the acceptable level of independence. This ensured reduced redundancy among explanatory variables and improved model stability and predictive performance (Mejia-Manrique et al., 2025).

$$VIF_i = \frac{1}{1-R_i^2} \quad (1)$$

Where R_i^2 represents the coefficient of determinants that regress all the causative factors on the others.

3.4 Methodology

The methodology developed in this study, as shown in Fig. 4, represents an effective approach for assessing landslide susceptibility and road vulnerability by integrating geospatial data, landslide inventory information, and the XG-Boost machine learning model. The framework facilitates the identification of landslide-prone zones and vulnerable road networks, thereby supporting hazard mitigation and sustainable infrastructure planning.

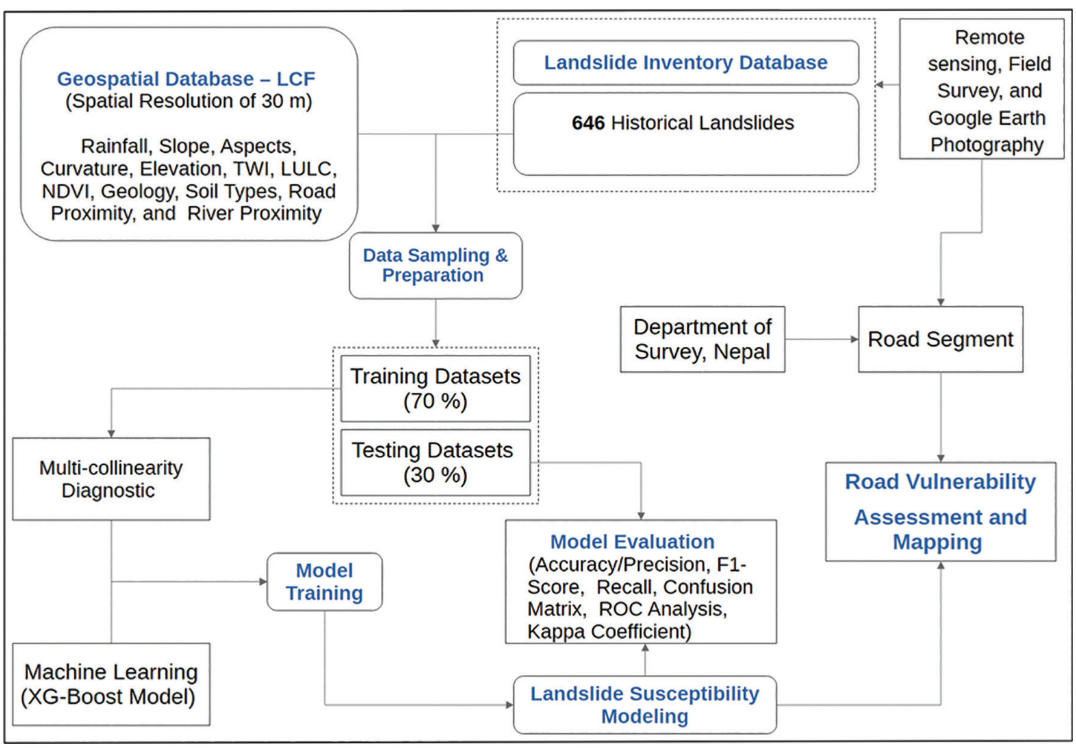

Figure 4: Flowchart showing the research approach and methodological framework for the landslide susceptibility mapping and road network vulnerability assessment

A geospatial database was developed using landslide conditioning factors with a spatial resolution of 30 m, including rainfall, slope, aspect, curvature, elevation, topographic wetness index (TWI), land use/land cover (LULC), normalized difference vegetation index (NDVI), geology, soil type, road proximity, and river proximity. A landslide inventory database comprising 646 landslide locations was prepared through the integration of remote sensing data, field surveys, and Google Earth imagery. The inventory data were combined with the conditioning factors, and the resulting dataset was randomly divided into training (70%) and testing (30%) subsets. Before model development, multicollinearity analysis was conducted to identify and eliminate highly correlated variables, thereby improving model reliability and reducing redundancy among predictors. Subsequently, the XGBoost machine learning algorithm (Section 3.4.1) was trained using the selected conditioning factors and landslide occurrences to establish the relationship between landslide events and their controlling factors. The trained model was then applied to generate a landslide susceptibility map for the entire study area. Model performance was evaluated using the testing dataset through several statistical indicators, including accuracy, precision, recall, F1-score, confusion matrix, receiver operating characteristic (ROC) curve, and Kappa coefficient (Section 3.4.2). Finally, road network data obtained from the Department of Survey, Nepal, and OpenStreetMap (OSM) were segmented and integrated with the generated landslide susceptibility map to assess the exposure and vulnerability of road infrastructure (Section 3.4.3).

3.4.1 XG-Boost Algorithm

XG-Boost is an advanced gradient boosting algorithm that builds an ensemble of decision trees sequentially, minimizing prediction errors and providing high accuracy for classification and regression tasks, making it widely used in landslide susceptibility mapping (Tien Bui et al., 2016; Khan et al., 2025). The algorithm iteratively builds decision trees by initializing predictions, computing residuals, constructing new trees, and updating predictions (Joshi et al., 2026). The model was implemented using the XG-Boost library in Python, with hyperparameters optimized through grid search and cross-validation. The model was trained on the balanced dataset of 646 landslide and 646 non-landslide points, with the 12 conditioning factors as predictors.

The XGBoost algorithm predicts the mean of the target variable (Eq. 2) for all training instances, which serves as the initial prediction before any iterative corrections. Each iteration calculates residuals by subtracting the current predictions from the actual target values (Eq. 3).

$$\hat{y}_i = mean\ (y) \qquad (2)$$

$$r_i = y_1 - \hat{y}_i \qquad (3)$$

where y_1 is the residual for the i^{th} instance, y_1 is the base value, and $\hat{y}_i$ represents the initial prediction. A new decision tree is fitted to the residuals by optimizing splits, and predictions are updated by adding a fraction (learning rate, η) of the tree's output to improve accuracy gradually (**Eq.4**).

$$\hat{y}_i^t = \hat{y}_i^{t-1} + \eta Tree_m(x_i) \qquad (4)$$

In this equation, $\hat{y}_i^t$ is the prediction is for instance i at the current iteration t, $\hat{y}_i^{t-1}$ is the prediction from the previous iteration t-1. Similarly, $Tree_m\ (x_i)$ is the new prediction of the new decision being added from m^{th} tree for instance x_i. For a fixed number of boosting rounds, the model iteratively reduces residuals while minimizing a regularized objective function (**Eq.5**) that balances prediction error and model complexity.

$$Obj(\theta) = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \Omega(f_m) \qquad (5)$$

where, θ represents model parameter, $L(y_i, \hat{y}_i)$ is the loss function that represents the error, and $\Omega(f_m)$ is the regularization term that penalizes model complexity to improve generalization. To improve generalization and prevent overfitting, it incorporates both L1 (Lasso) (**Eq. 6**) and L2 (Ridge) regularization (**Eq. 7**)

$$\Omega(f_m)=\lambda_1 \sum_{j=1}^{N} \mid W_j \mid \quad (6)$$

$$\Omega(f_m)=\lambda_1 \sum_{j=1}^{N} W_j^2 \quad (7)$$

where λ_1 is the regularization parameter and W_j is the Weight (score) assigned to leaf *j*.

3.4.2 Model Evaluation and Validation

Model performance was evaluated using a confusion matrix comprising true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), where landslide and non-landslide samples represented the positive and negative classes, respectively. Classification performance was assessed using accuracy, recall, precision, F1 score, and Cohen's kappa coefficient (Eq. 8). Accuracy measures the overall proportion of correctly classified samples; recall quantifies the model's ability to identify actual landslide occurrences; precision indicates the reliability of predicted landslide instances; and the F1-score provides a balanced assessment by combining precision and recall. Cohen's kappa coefficient measures the agreement between observed and predicted classifications while correcting for chance agreements.

In addition, the predictive capability of the models was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC–ROC). The ROC curve illustrates the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across different classification thresholds (Çorbacıoğlu & Aksel, 2023).

$$\begin{cases} \text{Accuracy}=\dfrac{TP+TN}{TP+TN+FP+FN} \\ \text{Recall}=\dfrac{TP}{TP+FN} \\ \text{Precision}=\dfrac{TP}{TP+FP} \\ \text{F1-score}=2 \times \dfrac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \\ \text{Kappa}=\dfrac{P_o - P_e}{1- P_e} \end{cases} \quad (8)$$

where, P_e is the expected random agreement and P_o observed agreement

3.4.3 Road Vulnerability Assessment

The road vulnerability assessment was conducted by integrating the landslide susceptibility map with the existing road network using a GIS-based spatial overlay technique. The road network was systematically classified according to its functional hierarchy, comprising primary, residential, and unclassified routes. The classified road networks were overlaid on categorized landslide susceptibility maps to identify the segments failing with different susceptibility zones (very low to very high). The exposure to hazard was quantified based on the cumulative length of road segments intersecting each susceptibility class. Vulnerability was expressed as the proportion and total length of roads affected by high and very high susceptibility zones relative to the overall road network. This approach enabled a quantitative assessment of infrastructure exposure to landslide hazards and facilitated comparison across different road classes. The resulting analysis provides a spatially explicit basis for identifying critical road segments and prioritizing mitigation measures in landslide-prone mountainous terrain.

4. Results and Analysis

4.1 Multicollinearity Assessment

Multicollinearity among the landslide causative factors was evaluated using the variance inflation factor (VIF), as shown in Fig. 5. The results indicate that all variables exhibit Variance Inflation Factor (VIF) values below the critical threshold of 5, indicating that multicollinearity is not a significant concern in the dataset. Elevation (VIF = 4.753), Road Proximity (VIF = 4.116), and Soil Type (VIF = 3.849) exhibited comparatively higher VIF values but remained within acceptable limits. The remaining factors showed low VIF values ranging from 1.056 to 2.103, indicating weak intervariable correlations. Therefore, all twelve causative factors were retained for subsequent landslide susceptibility modeling, as they provide independent and non-redundant information for the analysis.

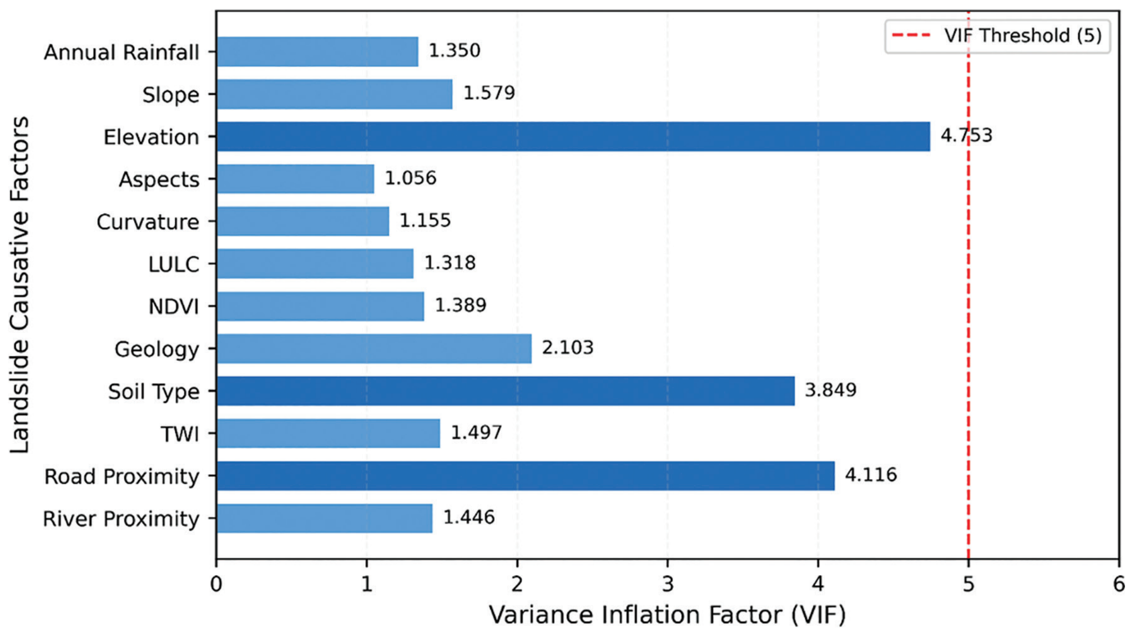

Figure 5: Multicollinearity analysis using the variance inflation factor (VIF) test

4.2 Evaluation Metrics and Classification Report

Model performance was evaluated using multiple statistical metrics, including the confusion matrix, accuracy, precision, recall, F1 score, and Cohen's kappa coefficient. The classification report (Table 2) indicates consistent performance across both classes, with approximately 93% precision, recall, F1-score, and accuracy. The Kappa coefficient of 0.861 further confirms strong agreement between observed and predicted classifications beyond chance level. The Receiver Operating Characteristic (ROC) and confusion matrix analysis (Fig. 6) demonstrate excellent predictive performance, with a mean AUC of 0.98 ± 0.0056 (SD) and a 95% confidence interval of 0.973–0.987 (from stratified 5-fold cross-validation), indicating stable and strong predictive capability. Both training and testing curves closely approach optimal performance, with minimal deviation.

Table 2: Comprehensive landslide susceptibility model evaluation

Label	Recall (%)	F1- score (%)	Accuracy (%)	Support	Kappa
0	93	93	93	1039	0.861
1	93	93	93	1027	

1: landslides and 0: non-landslides

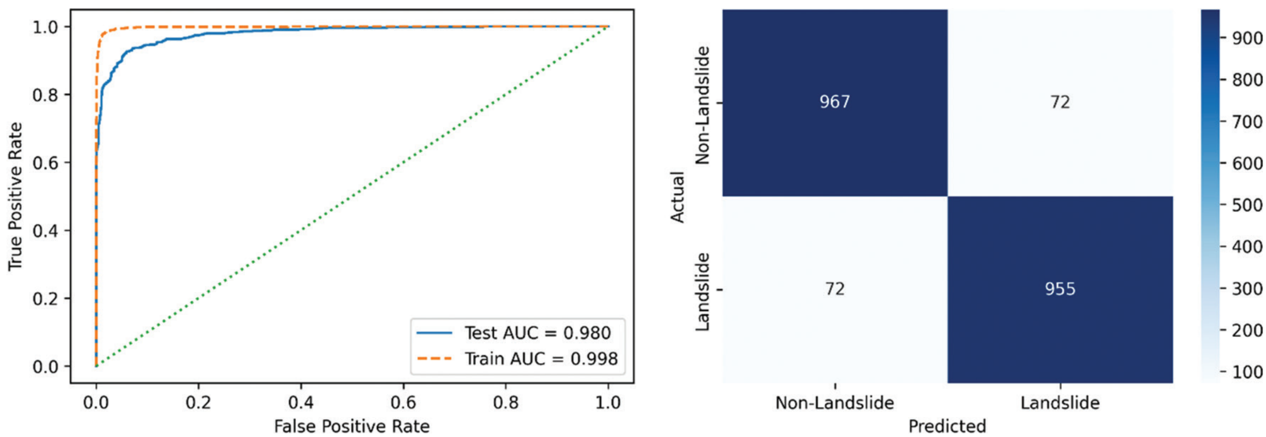

Figure 6:. Comprehensive performance assessment of XGBoost model based on ROC curve analysis and confusion matrix during training and validation phases

4.3 Feature Importance Analysis

The model's feature importance analysis revealed that an integrated set of environmental, topographic, hydrological, and anthropogenic factors primarily governs landslide occurrence. Land use/land cover (LULC) emerged as the most influential factor (13.9%), followed by rainfall (13.5%), elevation (12.5%), NDVI (12.2%), and slope (11.7%), as shown in Fig. 7. The dominance of LULC highlights the significant influence of anthropogenic landscape modifications such as vegetation removal, agricultural expansion, and urban development on slope instability, while rainfall, as the second most influential factor, emphasizes the critical role of intense monsoonal precipitation in elevating pore-water pressure, reducing soil shear strength, and triggering landslides. Topographic variables, particularly elevation and slope, also exhibited strong predictive power. Elevation represents variations in geomorphological conditions, climatic influences, and weathering processes, whereas slope directly governs gravitational forces acting on hillslopes. Similarly, NDVI revealed substantial importance, indicating the role of vegetation cover in reducing erosion susceptibility. Proximity to rivers (9.8%) and roads (8.7%) exhibited moderate influence on landslide occurrence, where river erosion and toe-cutting destabilize adjacent slopes, and road construction alters slope geometry, weakens soil structure, and concentrates surface runoff. Aspect (6.6%) exhibited relatively low influence, affecting solar radiation, soil moisture, and vegetation distribution. In contrast, soil type (3.5%), TWI (2.8%), geology (2.6%), and curvature (2.2%) exhibited comparatively lower importance. Although these factors contribute to localized variations in slope stability, their influence is less pronounced than the dominant climatic, land-cover, and topographic controls. Overall, the results indicate that landslide susceptibility in Kaski is primarily driven by the interaction of land-use dynamics, rainfall patterns, terrain characteristics, and vegetation conditions, which emphasizes the necessity of integrated land management and infrastructure planning in mountainous regions.

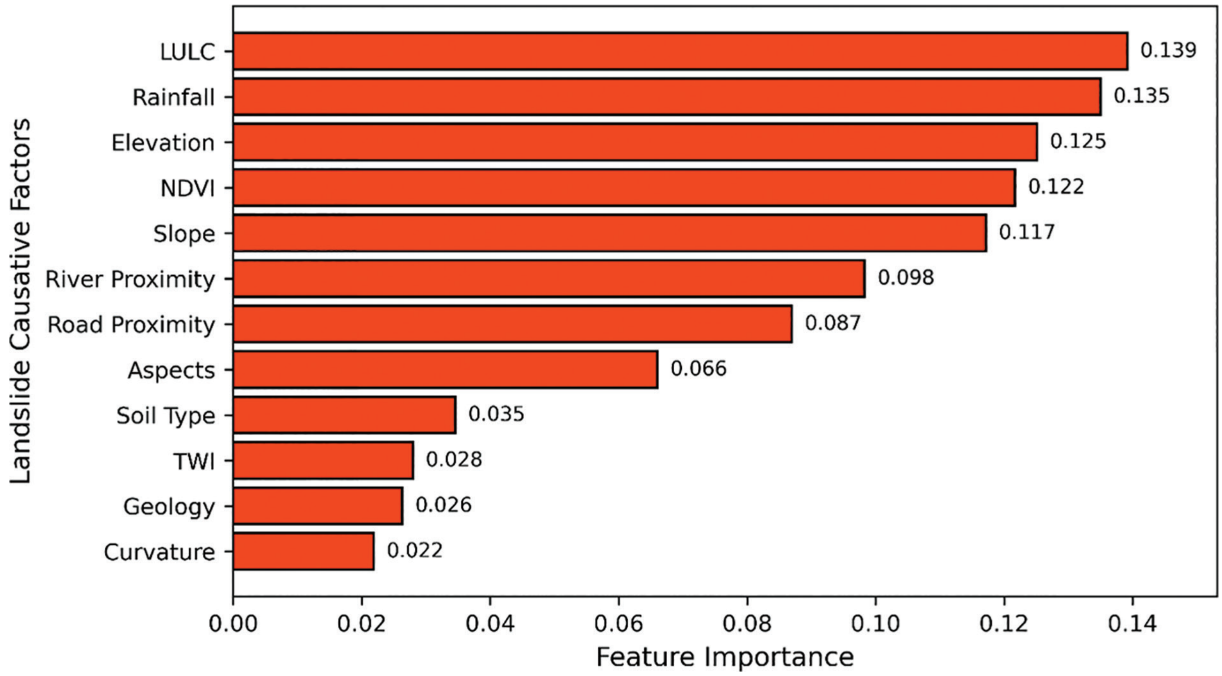

Figure 7: Feature importance analysis of conditioning factors influencing landslide susceptibility

4.4 Landslide Susceptibility Zonation

The XGBoost model was trained using the prepared training dataset. Model hyperparameters were optimized using a grid search approach coupled with 5-fold cross-validation to ensure robust and reliable parameter selection. The optimized parameters include n_estimators = 300, max_depth = 10, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8, gamma = 0.1, reg_alpha = 0.1, reg_lambda = 1, min_child_weight = 3, scale_pos_weight = ratio, random_state = 75, and eval_metric = logloss, selected based on cross-validation performance during hyperparameter tuning. This optimized setup enhanced the model's capability to effectively capture complex non-linear relationships between conditioning factors and landslide occurrence. The trained model was then applied to predict the probability of landslide occurrence for each pixel within the study area; the probability values were used as a susceptibility index to generate the LSM within a Geographic Information System (GIS) environment. The resulting map was further classified into five categories representing varying levels of susceptibility, from very low to very high. This classification was performed using the natural breaks method, and the spatial distribution of landslide susceptibility was visualized using a color gradient, where dark green tones denote areas of very low susceptibility and red represents zones of very high susceptibility (Fig. 8).

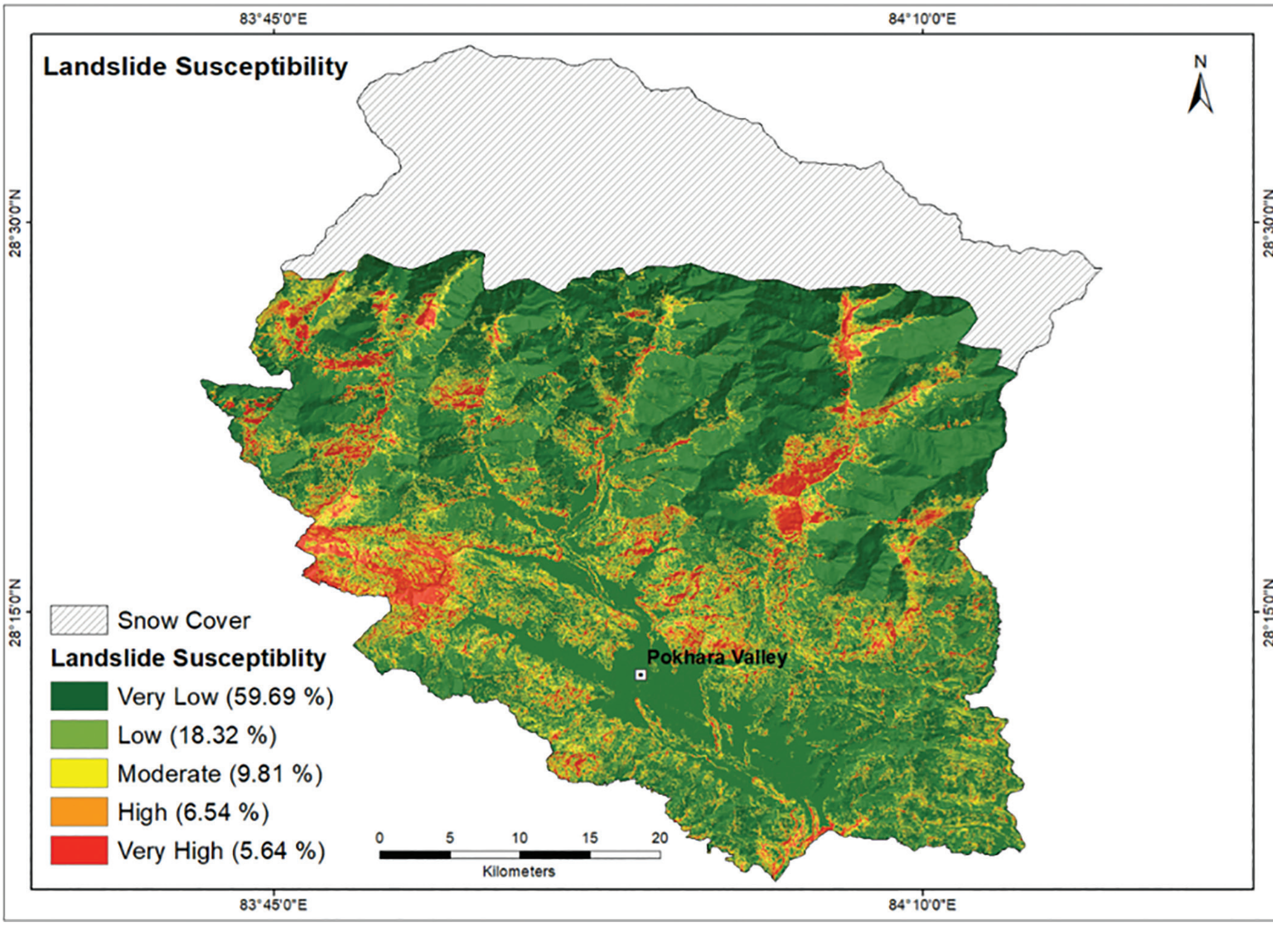

Figure 8: Landslide susceptibility map, categorized into different levels of hazard potential, from very low (dark green) to very high (red).

The resulting landslide susceptibility map revealed that the majority of the area (59.69%) falls under very low susceptibility, while 12.18% are classified as high to very high susceptibility, with localized hotspots. Spatially, the low susceptibility classes are predominantly distributed across stable terrain, particularly around the central and southern parts of the basin, including the Pokhara Valley region. In contrast, high susceptibility zones are mainly concentrated along steep mountain slopes, river valleys, and highly dissected terrain in the northern and western sections of the study area. These areas are characterized by unfavorable geomorphological and environmental conditions, such as steep slopes, intense rainfall influence, and proximity to drainage networks, which increase the likelihood of slope failures. In addition, areas along local roads are particularly susceptible due to poor construction practices and inadequate guidelines.

4.5 Road Network Vulnerability to Landslides

The study area comprises a total road network length of 2,975.7 km, classified as primary roads (640.77 km), residential roads (935.22 km), and unclassified roads (1399.7 km) (Table 3). The spatial overlay of the road network and the landslide susceptibility map shows that a large part of the transportation infrastructure is at risk of landslides at different levels (Table 3 and Fig. 9). Results indicate that a significant proportion of roads fall within low to very low susceptibility zones, accounting for 1,907.75 km (64%) of the total network.

However, a significant portion of the road networks falls within moderate to very high susceptibility classes. Specifically, 342.19 km (11.50%) of roads fall under the high susceptibility zone, while 350.99 km (11.80%) are located in very high susceptibility areas. Together, nearly 23.30% of the total road network is highly vulnerable to landslide hazards. Among road categories, residential roads exhibit the highest relative exposure to high-risk zones, with 16.40% and 17.67% of their total length falling under high and very high susceptibility classes, respectively. Unclassified roads also exhibit substantial exposure, with 16.40% in the high susceptibility zone and 17.67% in the very high category, indicating their significant vulnerability in terrain-sensitive areas. In contrast, primary roads are comparatively less exposed, with 10.62% and 11.07% of their length falling under high and very high susceptibility zones, respectively. Overall, the spatial distribution of vulnerability highlights that unclassified local road networks are more susceptible to landslide hazards than primary roads. These lower-order roads are more exposed to slope instability due to limited engineering standards and steeper terrain, which calls for targeted mitigation and prioritized interventions to improve road resilience in landslide-prone areas.

Table 3: Road network vulnerability under different classes of hazard

Hazard Level	Road Length (km)				% Roads in different hazard levels			
	PR	RR	UR	Total	% PR	% RR	% UR	% Total
Very Low	303.36	712.88	424.12	1440.37	47.34	76.23	30.3	48.40
Low	113.56	84.59	269.21	467.38	17.72	9.05	19.23	15.71
Moderate	84.88	60.39	229.51	374.78	13.25	6.46	16.40	12.59
High	68.03	44.60	229.56	342.19	10.62	4.77	16.40	11.50
Very High	70.94	32.76	247.30	350.99	11.07	3.50	17.67	11.80
Total	640.77	935.22	1399.70	2975.70	-	-	-	-

PR = Primary roads; RR = Residential roads; UR = Unclassified roads

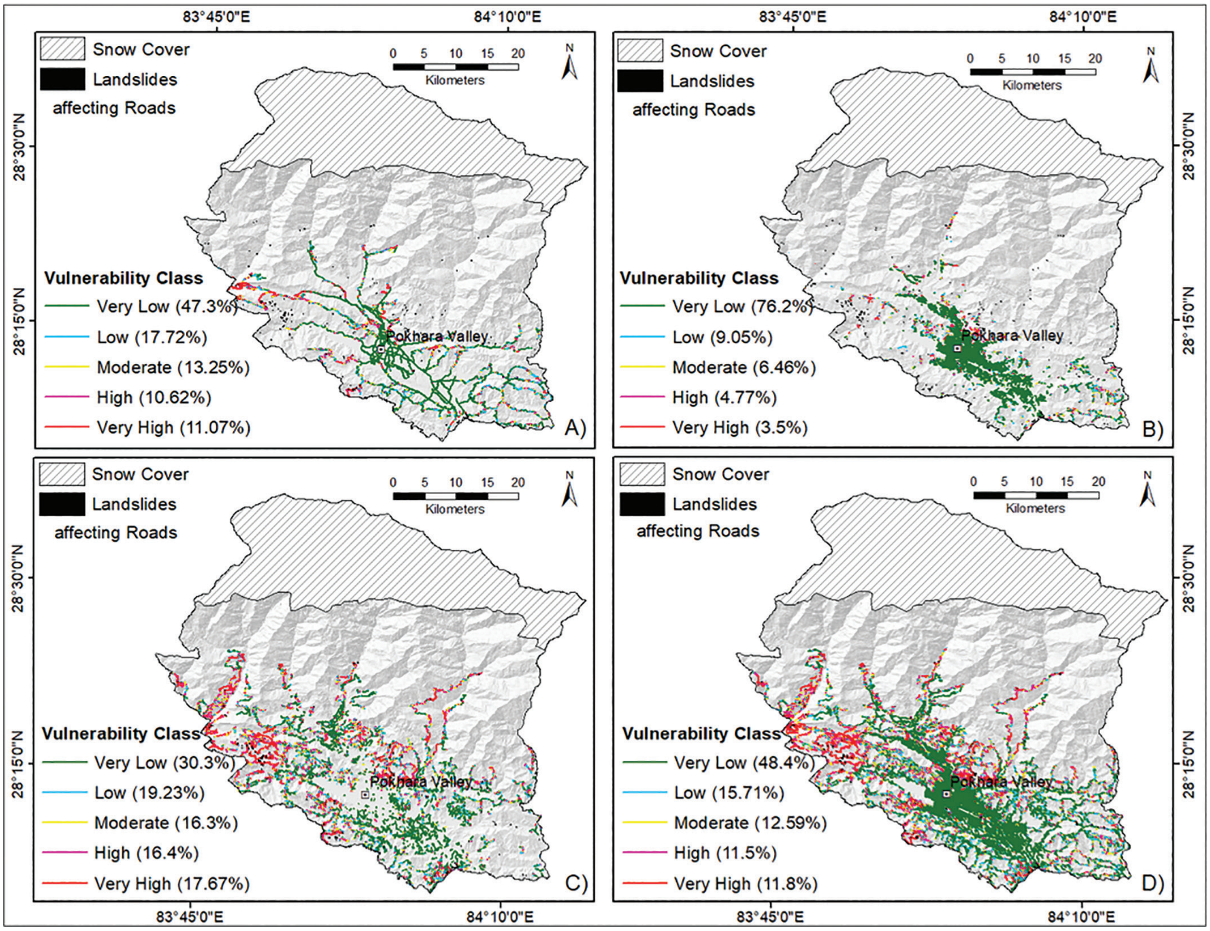

Figure 9: Road vulnerability assessment of road networks: A) Primary routes, B) Residential routes, C) Unclassified roads, and D) Total road network

5. Discussion

The dominance of LULC and rainfall as the primary controlling factors is consistent with established geomorphological principles. Anthropogenic land-use modifications such as deforestation and unregulated urbanization are well-documented drivers of slope destabilization (Chalise et al., 2019; Joshi et al., 2025), while intense monsoonal precipitation elevates pore-water pressure and reduces soil shear strength, triggering shallow landslides and debris flows (Basnet et al., 2020). The significant contribution of NDVI underscores the role of vegetation in reinforcing soil structure through root cohesion and regulating hydrological processes (Asada & Minagawa, 2023). The comparatively lower importance of geology (2.6%), soil type (3.5%), and curvature (2.2%) may reflect the dominance of overlying anthropogenic and climatic signals in this rapidly changing landscape.

The spatial concentration of high- and very-high-susceptibility zones along steep northern and western slopes, river valleys, and areas proximate to road corridors reflects the compounding influence of both natural and anthropogenic processes. These zones are characterized by steep gradients, intense monsoonal precipitation, deeply incised drainage networks, and extensive road construction, collectively elevating the probability of slope failure (Pathak et al., 2025; Robson et al., 2025). Road construction remains one of the most pervasive anthropogenic drivers of landslides in the Nepal Himalaya: slope cutting alters the natural equilibrium of hillslopes, inadequate drainage concentrates surface runoff, and poorly compacted fill materials are prone to rapid saturation and failure (Robson et al., 2025; P. Sharma et al., 2025). Landslide-induced road disruptions have cascading socio-economic consequences, including restricted access to markets and healthcare, increased transportation expenses, and delayed emergency response, effects that are disproportionately severe in remote mountain communities (Adhikari & Dhakal, 2026; Froude & Petley, 2018; Jones, 1998; Kincey et al., 2024). These findings emphasize the necessity of incorporating spatially explicit land-slide susceptibility assessments into infrastructure planning and development policies.

The road vulnerability assessment indicates that a significant portion of the Kaski road network faces substantial landslide hazards. Specifically, 21.67% of primary roads, 8.27% of residential roads, and 34.07% of unclassified roads intersect zones of high or very high landslide susceptibility, with 23.3% of the total road network (693.18 km out of 2,975.7 km) traversing high-risk terrain. Unclassified roads are more likely to be damaged because there are no formal engineering design standards and they are often built on steep, geologically sensitive land without proper slope stabilization or drainage systems. These findings are broadly consistent with recent vulnerability assessments in comparable Himalayan settings: Sharma et al. (2025) identified analogous exposure patterns along the Narayanghat-Kathmandu Road corridor, and Adhikari & Dhakal (2026) documented similarly high landslide susceptibility along mountain road sections in western Nepal. The concentration of at-risk roads in mid-elevation zones amplifies the potential societal impact of landslide-induced disruptions. Priority-based maintenance programs should allocate resources to the most exposed road segments, focusing on slope stabilization, retaining wall construction, drainage system upgrades, and real-time displacement monitoring for the 693.18 km of high-risk roads. Such a risk-prioritized approach maximizes the effectiveness of constrained maintenance budgets and systematically reduces network-wide exposure (Bhattarai & Bhandary, 2026a, 2026b).

Collectively, the findings reinforce the imperative of adopting a risk-informed, evidence-based approach to infrastructure development and management in mountainous terrain. Maintenance strategies should be risk-stratified, allocating proportionally more resources to the high- and very-high- susceptibility segments identified through this spatial overlay analysis. Engineering interventions, including slope reinforcement, retaining structures, surface and subsurface drainage improvements, and bioengineering techniques such as vetiver grass stabilization, are essential tools for reducing landslide risk along vulnerable road segments (Reichenbach et al., 2018). Embedding susceptibility maps within formal road alignment planning and environmental impact assessment processes can systematically steer infrastructure away from particularly unstable terrain, reducing future liability and long-term maintenance expenses. Coupling probabilistic susceptibility models with real-time rainfall monitoring, satellite-based displacement detection, and community-based early warning protocols represents a comprehensive approach to improving disaster preparedness and reducing landslide-induced road disruptions (Khan et al., 2025; Pradhan et al., 2025).

6. Conclusion

This study presents a comprehensive geospatial machine learning framework for landslide susceptibility mapping and assessing the vulnerability of the road network in the Kaski district (a mountainous terrain), Nepal. The model demonstrated superior predictive performance, achieving an overall accuracy of 93%, a Kappa coefficient of 0.861, and a mean AUC of 0.98 ± 0.0056. Multicollinearity diagnostics confirmed that all twelve conditioning factors retained for modeling provided independent and non-redundant information, strengthening the analytical integrity of the approach. About 59.69% of the study area falls under very low susceptibility, mainly in the stable Pokhara Valley and southern basin. Around 12.18% is highly susceptible, concentrated on steep northern and western slopes influenced by rainfall, drainage proximity, rugged terrain, and road construction activities. The interplay of anthropogenic and natural drivers primarily governs landslides in Kaski, as revealed by the feature importance analysis. Land use changes, rainfall, and slope emerged as the key influencing factors, highlighting the impacts of vegetation removal, agricultural expansion, and unregulated urban growth on slope stability. Proximity to rivers and roads also exerted a moderate influence, indicating the combined role of infrastructure development and fluvial erosion in increasing slope instability in the region. An overlying analysis of the road network showed that 693.18 km (23.3%) of the total road network (2,975.7 km) are in high-susceptibility zones. The result shows that road networks are very vulnerable. Due to poor engineering standards, unclassified local roads are more vulnerable to landslides, necessitating prioritized mitigation to enhance resilience. Collectively, the results indicate that landslide risk in Kaski is spatially uneven and concentrated in areas where intense rainfall, land use change, and unfavorable topographic conditions coincide. The findings underscore the necessity for targeted slope stabilization measures, improved road construction standards, and integrated disaster risk management strategies in mountainous regions to enhance transportation resilience.

Future research may further enhance road vulnerability assessment by incorporating additional influencing factors such as traffic intensity, road design standards, maintenance conditions, and structural characteristics of road segments. Integrating these parameters with geospatial and hazard-based indicators could provide a more comprehensive and realistic representation of infrastructure vulnerability, thereby improving decision-making for resilient road network planning and management.

Acknowledgements: All authors acknowledge all data sources such as Google Earth, SOTER Database for Nepal, the Department of Mines and Geology (DMG) of Nepal, Department of Hydrology and Meteorology (DHM) of Nepal, and the European Space Agency for providing crucial datasets like ASTER DEM datasets for a successful run of this study.

Supplementary Materials: No supplementary material is available for this study; however, we have utilized data from various sources, including ASTER DEM and other causative factors such as roads, soil, geology, lineaments, and streams, which can be accessed upon request from the authors.

Declaration of generative AI and AI-assisted technologies in the writing process: During the manuscript preparation, generative AI and AI-assisted technologies were also used for paraphrasing, grammar, and English editing to improve the readability of the work. All authors reviewed and edited the text as needed after using the AI tools. The authors take full responsibility for the published content, including all ideas, concepts, insights, scientific conclusions, and recommendations.

Author Contributions: B.R.J. and C.B. performed data collection, data preparation, data management, methodology development, landslide susceptibility mapping and analysis, and manuscript draft production. N.K.C. contributed to model verification, data updating, data verification, and improving the text. The research approach was subsequently improved, and the manuscript was amended with comments from I.P.A. N.P.B. supervised with general oversight, modification, and finalization of the text. All authors have read and approved the final version of the test.

Conflicts of Interest: The authors do not have any personal or institutional conflicts of interest that could have influenced or might influence the research findings.

Funding: This research has not received any external funding.

References

Adhikari, S., & Dhakal, S. (2026). Landslide susceptibility assessment and slope stability analysis of cut slopes along Dumre-Paudhi Bazar road section, western Nepal. *Geosystems and Geoenvironment*, *5*(2). https://doi.org/10.1016/j.geogeo.2026.100499

Asada, H., & Minagawa, T. (2023). Impact of Vegetation Differences on Shallow Landslides: A Case Study in Aso, Japan. *Water (Switzerland)*, *15*(18). https://doi.org/10.3390/w15183193

Basnet, K., Shrestha, A., Joshi, P. C., & Pokharel, N. (2020). Analysis of Climate Change Trend in the Lower Kaski District of Nepal. *Himalayan Journal of Applied Science and Engineering*, *1*(1), 11–22. https://doi.org/10.3126/hijase.v1i1.33536

Bhandari, C., Niraj, K. C., Acharya, I. P., Subedi, M., & Buddhi, R. (2026). Machine learning-based landslide risk assessment in Doti district, western Nepal. *Bulletin of Engineering Geology and the Environment*, *85*(7). https://doi.org/10.1007/s10064-026-05040-9

Bhattarai, T. R., & Bhandary, N. P. (2025). Machine Learning Ensemble Methods for Co-Seismic Landslide Susceptibility: Insights from the 2015 Nepal Earthquake. *Applied Sciences (Switzerland)*, *15*(15). https://doi.org/10.3390/app15158477

Bhattarai, T. R., & Bhandary, N. P. (2026a). Integration of machine learning model and CMIP6 analysis for climate change impact-led landslide susceptibility and population exposure assessments in the Nepal Himalaya. *Geoenvironmental Disasters*, *13*(1), 30. https://doi.org/10.1186/s40677-026-00382-8

Bhattarai, T. R., & Bhandary, N. P. (2026b). Slope unit-based evaluation of climate change impacts on landslide susceptibility in the Nepal Himalaya. *Geomatics, Natural Hazards and Risk*, *17*(1). https://doi.org/10.1080/19475705.2026.2649615

Chalise, D., Kumar, L., & Kristiansen, P. (2019). Land degradation by soil erosion in Nepal: A review. In *Soil Systems* (Vol. 3, Issue 1, pp. 1–18). MDPI AG. https://doi.org/10.3390/soilsystems3010012

Chowdhuri, I., Pal, S. C., Saha, A., Roy, P., Chakrabortty, R., & Shit, M. (2022). Application of novel framework approach for assessing rainfall induced future landslide hazard to world heritage sites in Indo-Nepal-Bhutan Himalayan region. *Geocarto International*, *37*(27), 17742–17776. https://doi.org/10.1080/10106049.2022.2134464

Chowdhury, M. S. (2023). A review on landslide susceptibility mapping research in Bangladesh. In *Heliyon* (Vol. 9, Issue 7). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e17972

Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. In *Turkish Journal of Emergency Medicine* (Vol. 23, Issue 4, pp. 195–198). Wolters Kluwer Medknow Publications. https://doi.org/10.4103/tjem.tjem_182_23

DMG. (2022). *'Annual Report of Department of Mines and Geology, Annual Report No. 13 DMG."*

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, *18*(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Gautam, P., Kubota, T., & Aditian, A. (2021). Evaluating underlying causative factors for earthquake-induced landslides and landslide susceptibility mapping in Upper Indrawati Watershed, Nepal. *Geoenvironmental Disasters*, *8*(1). https://doi.org/10.1186/s40677-021-00200-3

Huang, S., & Chen, L. (2024). Landslide susceptibility mapping using an integration of different statistical models for the 2015 Nepal earthquake in Tibet. *Geomatics, Natural Hazards and Risk*, *15*(1). https://doi.org/10.1080/19475705.2024.2396908

Jones, K. H. (1998). A comparison of algorithms used to compute hill slope as a property of the DEM. *Computers and Geosciences*, *24*(4), 315–323. https://doi.org/10.1016/S0098-3004(98)00032-6

Joshi, B. R., Bhandary, N. P., Acharya, I. P., KC, N., & Bhandari, C. (2025). Landslide susceptibility mapping optimization for improved risk assessment using multicollinearity analysis and machine learning technique. *Applied Sciences*, *15*(22), 12152.

Joshi, B. R., Bhandary, N. P., Acharya, I. P., KC, N., & Bhandari, C. (2026). Integration of information value with machine learning method for an enhanced predictive performance in landslide susceptibility mapping. *Discover Geoscience*, *152*, 1–36. https://doi.org/https://doi.org/10.1007/s44288-026-00517-2

Khan, D., Akram, W., & Ullah, S. (2025). Enhancing landslide susceptibility predictions with XGBoost and SHAP: a data-driven explainable AI method. *Geocarto International*, *40*(1). https://doi.org/10.1080/10106049.2025.2514725

Kincey, M. E., Rosser, N. J., Swirad, Z. M., Robinson, T. R., Shrestha, R., Pujara, D. S., Basyal, G. K., Densmore, A. L., Arrell, K., Oven, K. J., & Dunant, A. (2024). National-Scale Rainfall-Triggered Landslide Susceptibility and Exposure in Nepal. *Earth's Future*, *12*(2). https://doi.org/10.1029/2023EF004102

Mejia-Manrique, S. A., Ramos-Scharrón, C. E., Hughes, K. S., Gonzalez-Cruz, J. E., & Khanbilvardi, R. M. (2025). Dynamic landslide susceptibility for extreme rainfall events using an optimized convolutional neural network approach. *Natural Hazards*, *121*(13), 15383–15411. https://doi.org/10.1007/s11069-025-07396-9

Niraj Baral, Akhilesh Kumar Karna, & Suraj Gautam. (2021). Landslide Susceptibility Assessment Using Modified Frequency Ratio Model in Kaski District, Nepal. *International Journal of Engineering and Management Research*, *11*(1), 167–177. https://doi.org/10.31033/ijemr.11.1.23

Pathak, L., Baral, B., Joshi, K., Basnet, D. R., & Godone, D. (2025). Landslides in the Himalayas: The Role of Conditioning Factors and Their Resolution in Susceptibility Mapping. *Geosciences (Switzerland)*, *15*(4). https://doi.org/10.3390/geosciences15040131

Pawłuszek-Filipiak, K., & Lewandowski, T. (2025). The Impact of Feature Selection on XGBoost Performance in Landslide Susceptibility Mapping Using an Extended Set of Features: A Case Study from Southern Poland. *Applied Sciences (Switzerland)*, *15*(16). https://doi.org/10.3390/app15168955

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, *40*(10), 927–930. https://doi.org/10.1130/G33217.1

Pradhan, A. M. S., Ghimire, P., Shrestha, S., Lee, J. S., Lee, J. H., & Park, H. J. (2025). Extreme gradient boosting with Shapley Additive Explanations for landslide susceptibility at slope unit and hydrological response unit scales. *Geoscience Frontiers*, *16*(4), 102081. https://doi.org/10.1016/j.gsf.2025.102081

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. In *Earth-Science Reviews* (Vol. 180, pp. 60–91). Elsevier B.V. https://doi.org/10.1016/j.earscirev.2018.03.001

Robson, E. B., Dahal, B. K., & Toll, D. G. (2025). A participatory approach to determine the use of road cut slope design guidelines in Nepal to lessen landslides. *Natural Hazards and Earth System Sciences*, *25*(3), 949–973. https://doi.org/10.5194/nhess-25-949-2025

Roubal, R. (2014). *Impacts of the Beni-Jomsom Road (Kali Gandaki Valley) on traditional social structures, settlements, and tourism in the Kali Gandaki Valley, Nepal (Master Thesis).*

Searle, M. P., Law, R. D., Godin, L., Larson, K. P., Streule, M. J., Cottle, J. M., & Jessup, M. J. (2008). Defining the Himalayan Main Central Thrust in Nepal. *Journal of the Geological Society*, *165*(2), 523–534. https://doi.org/10.1144/0016-76492007-081

Sharma, K. K., Bhandary, N. P., Subedi, M., & Rajan, K. C. (2025). Integration of Landslide Susceptibility and Road Infrastructure Vulnerability for Risk Assessment and Mountain Road Resilience Enhancement. *Indian Geotechnical Journal*. https://doi.org/10.1007/s40098-025-01222-6

Sharma, P., Chandel, V. B. S., Kahlon, S., & Thakur, S. N. (2025). Assessing landslide susceptibility in the Upper Ravi river catchment, Himachal Pradesh, India: a comprehensive analysis using the logistic regression model. *Geoenvironmental Disasters*, *12*(1). https://doi.org/10.1186/s40677-025-00327-7

Singh, A., Dhiman, N., Niraj, K. C., & Shukla, D. P. (2024). "Ensembled transfer learning approach for error reduction in landslide susceptibility mapping of the data scare region." *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-76541-4

Ullah, I., Reicherter, K., Pánek, T., Tibaldi, A., Al-Najjar, H., Kalantar, B., Ueda, N., & Havenith, H. B. (2025). Landslide susceptibility mapping based on data mining models in Lesser Caucasus and Kura foreland basin (Armenia and Azerbaijan). *Geomatics, Natural Hazards and Risk*, *16*(1). https://doi.org/10.1080/19475705.2025.2537221

UNDRR. (2015). *Sendai Framework for Disaster Risk Reduction 2015 - 2030.* https://www.undrr.org/media/16176/download?startDownload=20260622

Wang, J., Zang, M., Peng, J., Xu, C., Su, Z., Liu, T., & Li, M. (2025). Identification of Dominant Controlling Factors and Susceptibility Assessment of Coseismic Landslides Triggered by the 2022 Luding Earthquake. *Remote Sensing*, *17*(16). https://doi.org/10.3390/rs17162797

Yalcin, A. (2008). GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *CATENA*, *72*(1), 1–12. https://doi.org/https://doi.org/10.1016/j.catena.2007.01.003