VNU Journal of Science: Earth and Environmental Sciences

Journal homepage: https://js.vnu.edu.vn/EES

Original Article

Spatial Assessment of Social and Environmental Vulnerabilities by Landslides in Hoa An Region, Non Nuoc Cao Bang UNESCO Global Geopark, Vietnam

Phan Thi Mai Hoa[1,2], Tran Thi Thu Huong[1], Nguyen Thi Cuc[1], Nguyen Quoc Phi[1,*]

[1]*Hanoi University of Mining and Geology, 18 Vien Street, Dong Ngac, Hanoi, Vietnam*

[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Vietnam*

Received 19th December 2025
Revised 18th May 2026; Accepted 25th June 2026

Abstract: This study develops a GIS-based framework to quantify socio-environmental vulnerability and examine its spatial correspondence with independently derived landslide density (LSD) in Hoa An District, within the Non Nuoc Cao Bang UNESCO Global Geopark, Vietnam. Eight social and environmental indicators were integrated using Principal Component Analysis, explaining 69.80% and 73.39% of variance, respectively, and K-means clustering identified four distinct vulnerability typologies. Landslide density, derived from a validated multi-temporal inventory using kernel density estimation, was used as an independent descriptor of past landslide activity. Results reveal a non-linear relationship between vulnerability and landslide distribution: compound vulnerability hotspots (34.01% of the area) exhibit the highest mean LSD (6.42 events/km²) and largest proportion of high-density zones (35.06%), while transitional areas show localized but elevated landslide concentrations despite moderate vulnerability, indicating the influence of unmodeled geomorphic controls. These findings demonstrate that vulnerability does not directly control landslide occurrence but modulates its spatial concentration through interactions between environmental degradation, human pressure, and terrain conditions, providing a transferable framework for identifying areas prone to disproportionate landslide impacts.

Key words: Principal Component Analysis, K-means clustering, environmental vulnerability, social vulnerability, landslides, UNESCO Global Geopark.

[*] Corresponding author.
E-mail address: nguyenquocphi@humg.edu.vn

https://doi.org/10.25073/2588-1094/vnuees.5471

1. Introduction

Landslides are a major source of mortality and damage in steep terrain, with impacts that are unevenly distributed in space and time [1]. A substantial body of geomorphological and statistical research has therefore focused on landslide susceptibility, i.e., estimating where landslides are more likely to occur under given terrain and environmental conditions [2]. While such approaches are effective in identifying areas prone to slope failure, they do not necessarily explain where landslides are likely to produce the most severe consequences. In practice, disproportionate impacts often occur in locations where exposed populations, critical infrastructure, or sensitive ecosystems have limited capacity to anticipate, withstand, or recover from landslide events [3, 4].

Within widely used risk frameworks, disaster impacts emerge from the interaction of hazard, exposure, and vulnerability [4]. In this context, hazard refers to the potential occurrence of damaging landslides, exposure denotes the presence of people, infrastructure, or ecosystems in areas that may be affected, and vulnerability describes the socio-economic and environmental conditions that increase susceptibility to harm when a hazard occurs [5, 6]. This conceptual distinction highlights the importance of vulnerability-focused assessments, which complement susceptibility mapping by identifying where similar hazard conditions may lead to different impact outcomes. Recent studies further emphasize that vulnerability can vary significantly among communities experiencing comparable levels of hazard exposure, depending on demographic structure, education, livelihood strategies, and adaptive capacity [7].

In Vietnam, recent studies have produced detailed landslide susceptibility maps, including applications in Cao Bang and other northern mountainous regions, reflecting continued methodological advances in predicting where landslides may occur [8]. At the same time, studies on hazard impacts have demonstrated the importance of social conditions and livelihood dependence in shaping community responses to hydro-meteorological hazards in upland environments [9]. However, these research strands are often developed separately, with limited integration between spatial hazard analysis and vulnerability assessment.

Despite these advances, empirical studies that explicitly integrate both social and environmental dimensions of vulnerability within a unified spatial framework remain limited, particularly in geopark settings where environmental conservation and risk management objectives intersect. In addition, the extent to which vulnerability patterns correspond to observed landslide distribution at the landscape scale has not been systematically examined using independent reference indicators [10]. This gap constrains the ability to identify areas where landslides are more likely to generate disproportionate impacts, beyond what can be inferred from susceptibility mapping alone.

Indicator based vulnerability assessments have been widely applied across different hazards and spatial scales, but their implementation depends strongly on data availability and research objectives. In data-rich environments, vulnerability indices are typically constructed using census based areal units and multivariate statistical techniques to summarize correlated indicators [11]. In contrast, in data-scarce contexts, contextualized approaches relying on locally collected information have been developed to better capture vulnerability mechanisms; for example, studies in Rwanda have applied Principal Component Analysis (PCA) to household survey data to identify key vulnerability components under explicit data constraints [7]. Complementary work also shows the value of examining vulnerability jointly with hazard and exposure information to improve the relevance of risk management decisions, while maintaining conceptual separation among these components [12].

Despite the growing application of geospatial and data-driven approaches in landslide studies, most existing research continues to focus on hazard or susceptibility mapping, emphasizing the probability of landslide occurrence based on terrain and environmental factors. Comparatively fewer studies have systematically examined spatial patterns of social and environmental vulnerability as interacting systems, particularly in complex karst landscapes where geomorphological processes and human activities are closely coupled.

Furthermore, previous studies often implicitly assume a direct relationship between vulnerability and landslide occurrence, suggesting that areas with higher vulnerability will necessarily experience higher landslide activity. However, such assumptions overlook the fact that observed landslide distribution primarily reflects past geomorphic processes, whereas vulnerability represents underlying socio-environmental conditions that may amplify or mitigate impacts, rather than directly control landslide initiation.

To address this gap, this study develops a GIS-based analytical framework to identify spatial typologies of social and environmental vulnerability to landsliding in Hoa An District, located within the Non Nuoc Cao Bang UNESCO Global Geopark. Social vulnerability is represented by population density, proportion of disadvantaged groups, education level, and distance to health facilities, while environmental vulnerability is characterized by vegetation condition (NDVI), forest ecosystem density, forest loss intensity, and soil heavy metal contamination. Principal Component Analysis is applied separately to these indicator groups to derive dominant vulnerability dimensions, and K-means clustering is used to classify spatial units into internally consistent vulnerability typologies.

Importantly, landslide density (LSD), derived from an independent landslide inventory using kernel density estimation, is not included as an input variable in the clustering process. Instead, it is employed as an external spatial reference to examine the degree of correspondence between identified vulnerability typologies and observed landslide distribution patterns.

The objective of this study is therefore not only to identify dominant vulnerability structures across the study area, classify these structures into interpretable spatial typologies, and evaluate their spatial correspondence with landslide density, but also to provide a more nuanced understanding of how socio-environmental vulnerability interacts with geomorphic processes and to offer a spatially explicit basis for risk-informed land management and sustainable development planning in karst geopark environments.

2. Data and Methodology

2.1. Study Area

Hoa An area, located in Cao Bang Province in northeastern Vietnam, covers approximately 606 km² with a population of about 52,762 inhabitants as of 2019 (Figure 1) [13]. The study area is characterized by steep slopes (commonly >25°), structurally complex lithology, and intense seasonal rainfall (May–September), which are widely recognized as key predisposing and triggering factors for landslides [14]. Local communities, primarily ethnic minorities such as Tay, Nung, H'mong, and Dao, rely heavily on agriculture and forestry. Limited infrastructure and economic constraints further exacerbate their vulnerability to natural hazards, underscoring the importance of assessing both social and environmental dimensions of risk.

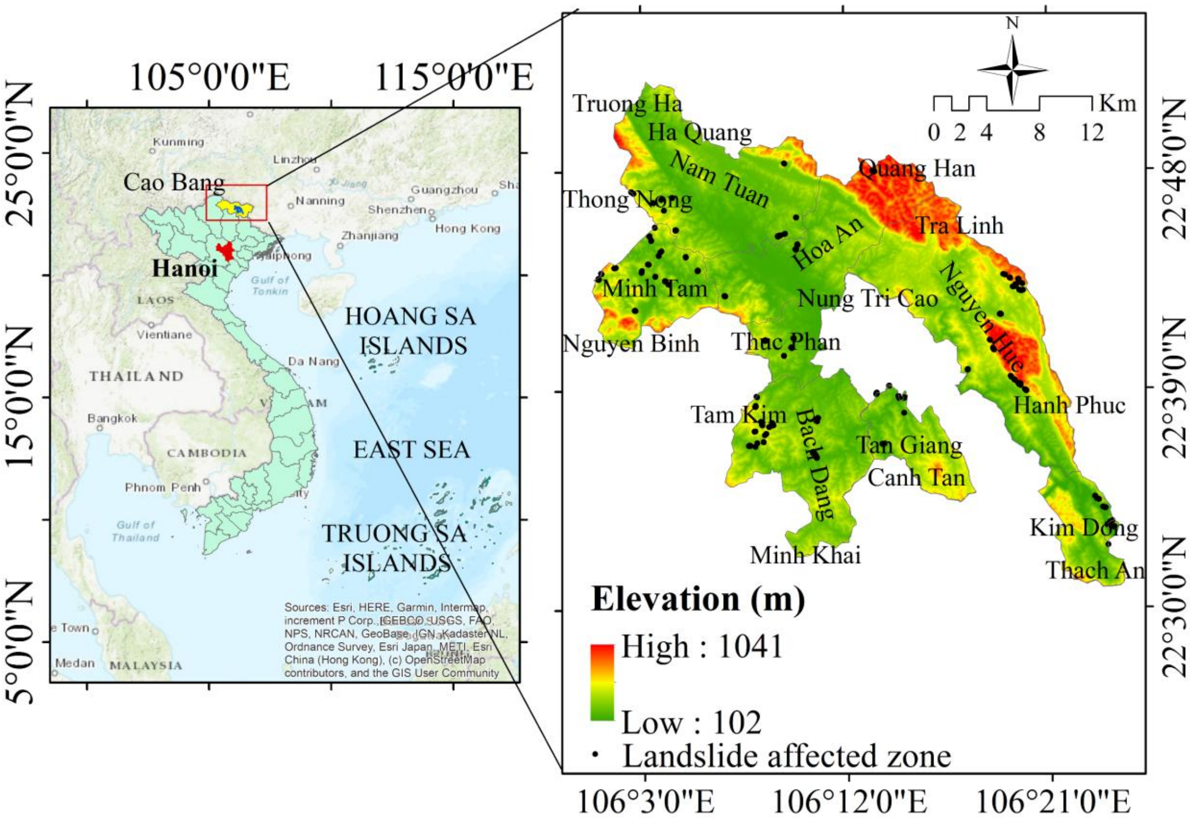

Figure 1. Study area and landslide-affected locations overlaid on the elevation map of Hoa An area (Vietnam).

2.2. Data Properties

Landslides represent a major source of human and economic loss globally [1, 15], yet their impacts depend strongly on underlying social and environmental conditions. In this study, vulnerability is defined as a spatial configuration of such conditions that shapes potential impacts when landslides occur, rather than an event-based measure of damage [16, 17]. Accordingly, indicators are interpreted as proxies of sensitivity and adaptive capacity under exposure and are explicitly distinguished from landslide susceptibility [2].

A total of eight indicators were selected to represent socio-environmental vulnerability. Social vulnerability is characterized by population density (PD), proportion of disadvantaged groups (DGP), education level (EDU), and distance to health facilities (DHF), capturing exposure concentration, social sensitivity, and adaptive capacity (Table 1; Figure 3) [9, 16, 17]. Environmental vulnerability is represented by NDVI, forest ecosystem density (FED), forest loss intensity (FLI), and soil heavy metal contamination (SHMC), which are interpreted as environmental stressors reflecting vegetation degradation, ecosystem disturbance, and contamination processes (Figure 4). These variables do not directly control slope failure but may amplify landslide impacts by altering surface stability, hydrological response, and ecosystem resilience [18, 19]. In this study, environmental indicators are not treated as direct measures of landslide susceptibility but as proxies of environmental conditions that may increase sensitivity to damage when landslides occur.

Therefore, the selected indicators are used to characterize vulnerability as a conditioning factor influencing impact severity rather than as predictors of landslide occurrence.

To ensure consistency in interpretation, all indicators were directionally standardized such that higher values correspond to higher vulnerability. Variables that are inversely related to vulnerability, including NDVI, FED, EDU, and DHF, were transformed accordingly prior to analysis.

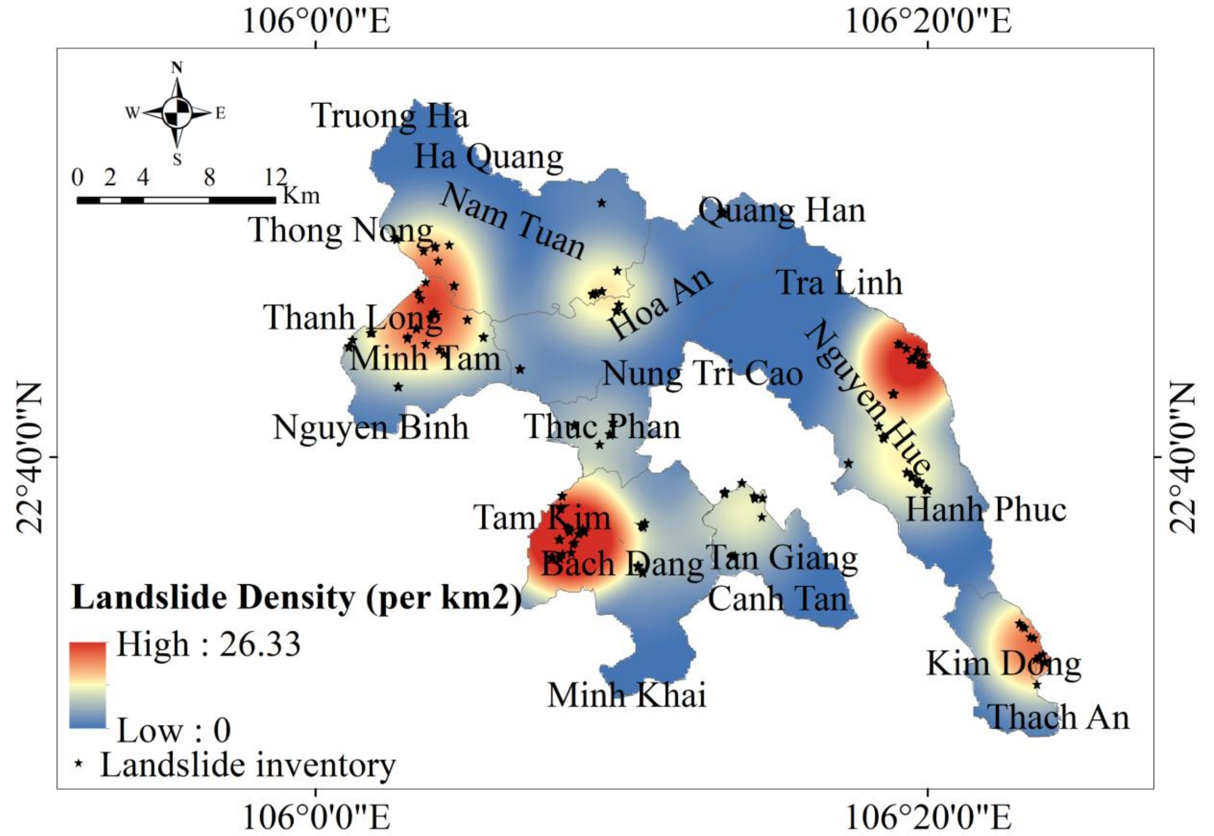

Figure 2. The spatial landslide density (LSD) map.

A landslide inventory comprising 101 locations was compiled through the interpretation of high-resolution Google Earth imagery and validated by field surveys conducted in 2020, 2022, and 2024. Kernel density estimation was applied to derive landslide density (LSD), representing the spatial concentration of past landslides (Figure 2). In this study, LSD is not used as an input variable but serves as an independent proxy of landslide exposure to evaluate the spatial correspondence between vulnerability patterns and observed landslide distribution [20, 21].

Socio-economic data were originally compiled at the commune level and spatially disaggregated to raster format using an area-weighted approach, assuming uniform distribution within each administrative unit. This simplification may introduce spatial smoothing. Environmental variables were resampled to a common grid with a spatial resolution of 30 m to ensure consistency across datasets. All variables were standardized using z-score normalization prior to analysis to ensure comparability and prevent scale dominance in multivariate processing [22].

Table 1. Socio-environmental vulnerability indicators, Data sources and Processing methods

No.	Group	Indicator	Description	Source	Processing
1	Social	Population Density (PD)	People per km²,	2019, Cao Bang census	Interpolated by commune area (GIS calculation)
2		Disadvantaged Group Proportion (DGP)	% pop. <5 or >65	2019, Cao Bang census	Aggregated from census data
3		Education Level (EDU)	Avg. schooling years	2019, Cao Bang census	Aggregated from census data
4		Distance to Health Facility (DHF)	Euclidean dist. to nearest facility	Cao Bang Department of Health + GIS (2024)	GIS-based Euclidean distance (raster analysis)
5	Environmental	Normalized Difference Vegetation Index (NDVI)	Vegetation cover and condition	2020, Sentinel-2 MSI imagery	Calculation of NDVI using satellite image bands (NIR and Red)
6		Forest ecosystem density (FED)	Proportion of forest cover per spatial unit	National forest inventory	Spectral index calculation
7		Forest loss intensity (FLI)	Degree of forest cover reduction over time	2020, LULC maps	Spectral index calculation
8		Soil heavy metal contamination (SHMC)	Composite index of soil contamination (As, Cd, Cr, Hg, Ni, and Pb)	Field sampling; environmental monitoring	calculated using the integrated pollution index (IPI), defined as the mean of contamination factors (CF), where CF is the ratio between measured and background concentrations

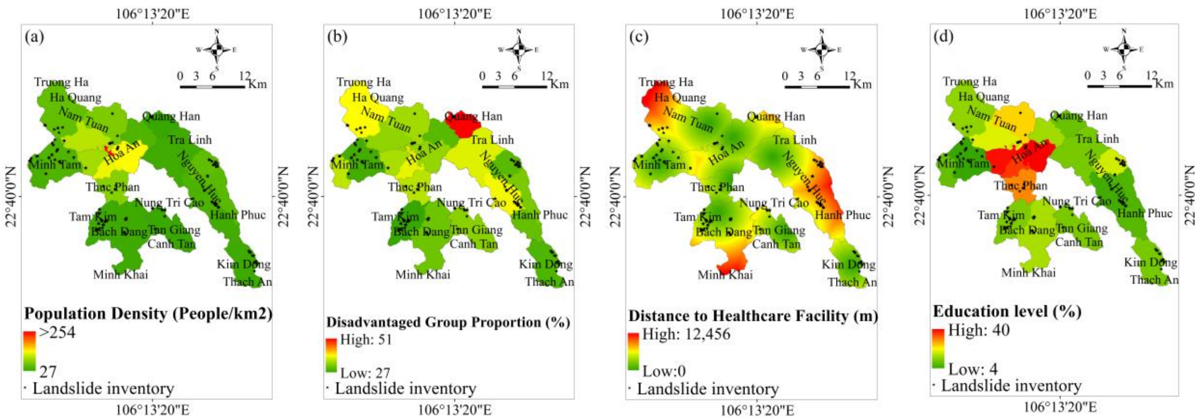

Figure 3. Data layer map of Socio vulnerability indicators: (a) Population Density (PD); (b) Disadvantaged Group Proportion (DGP); (c) Distance to Health Facility (DHF) (d) Education Level (EDU).

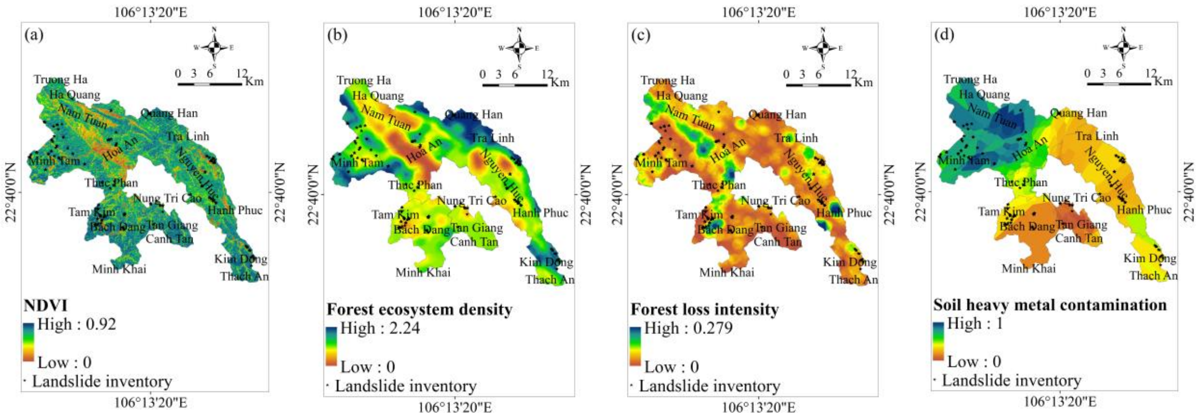

Figure 4. Data layer map of Environmental vulnerability indicators: (a) Normalized Difference Vegetation Index (NDVI); (b) Forest ecosystem density (FED); (c) Forest loss intensity (FLI); (d) Soil heavy metal contamination (SHMC).

2.3. Analytical Framework

This framework assesses socio-environmental vulnerability based on eight standardized indicators in Figure 5. Principal Component Analysis (PCA) is applied separately to social and environmental indicator groups to extract dominant vulnerability dimensions. The resulting component scores are then used as inputs for K-means clustering to identify spatial typologies of vulnerability.

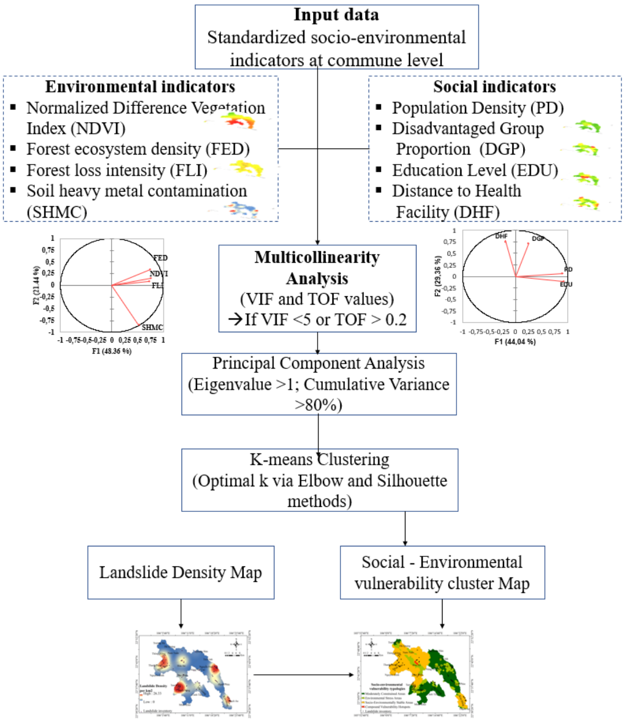

Figure 5. Analytical framework for assessing social and environmental vulnerability to landslide hazard.

PCA was applied separately to social and environmental domains to preserve domain-specific structures and avoid mixing variables with different conceptual meanings.

Landslide density (LSD), derived independently from landslide inventory data, is not included in the modeling process but is used as an external reference variable to evaluate how identified vulnerability typologies correspond to observed landslide concentration.

2.4. Methodology

All analyses were conducted at a spatial resolution of 30 m. The integration of raster-based environmental variables with downscaled socio-economic data introduces potential scale mismatch effects, which are acknowledged as a limitation.

Multicollinearity among indicators was assessed using VIF and Tolerance to ensure variable independence. Multicollinearity was assessed using Variance Inflation Factor (VIF) and Tolerance (TOL), with thresholds of VIF < 2 and TOL > 0.5 adopted to ensure low collinearity and stable PCA performance, consistent with previous studies [22]. Principal Component Analysis was then applied to standardize and reduce the socio-environmental dataset into orthogonal components that captured most of the variance [17]. Components with eigenvalues greater than 1 and cumulative explained variance exceeding 80% were retained to balance dimensionality reduction and information preservation [23].

K-means clustering was applied to the retained PCA scores to classify raster cells into internally coherent vulnerability groups. All analyses were conducted using XLSTAT (Addinsoft, version 2026.1.0) integrated within Microsoft Excel. Prior to PCA, all variables were standardized using z-score normalization. PCA was performed using the XLSTAT Principal Component Analysis module based on the correlation matrix, and components were retained according to the Kaiser criterion (eigenvalues > 1) and cumulative explained variance (> 70%). K-means clustering was then implemented using the XLSTAT Clustering module, with Euclidean distance, k-means++ initialization, and default convergence settings. The optimal number of clusters was determined using the Elbow method (within-cluster sum of squares) and silhouette analysis, ensuring a balance between cluster compactness and separation. The optimal number of clusters was selected using the Elbow criterion and silhouette analysis [23]. To ensure clustering robustness, the K-means algorithm was initialized multiple times with different random seeds, and the solution with the lowest within-cluster variance was retained. Cluster cohesion and separation were verified before mapping the results in GIS to reveal spatial patterns of socio-environmental vulnerability.

To evaluate the spatial relationship between vulnerability typologies and landslide occurrence. Landslide density (LSD) represents the spatial footprint of past landslide activity, and is used as an independent reference variable for comparison. LSD was derived from the landslide inventory using kernel density estimation with consistent parameters across the study area. It was not used as a model input but as an external descriptor to compare exposure levels among vulnerability typologies. Overlay analysis was conducted by calculating summary. statistics of LSD within each vulnerability typology, including mean, maximum, and proportion of high-density area.

The results of this overlay analysis are presented in Section 3, where variations in landslide density across vulnerability typologies are analyzed to identify areas more likely to experience severe impacts.

3. Results

3.1. Principal Component Structure of Social and Environmental Vulnerability

Multicollinearity diagnostics for social and environmental indicators based on Variance Inflation Factor (VIF) and Tolerance (TOL) in Table 2. All variables satisfy recommended thresholds (VIF < 2; TOL > 0.5), indicating low collinearity and suitability for Principal Component Analysis.

Table 2. Multicollinearity diagnostics of Social – Environmental indicators (VIF and TOL)

Group	Variable	R^2	TOL	VIF
Social	PD	0.490	0.510	1.961
	DGP	0.052	0.948	1.055
	EDU	0.495	0.505	1.980
	DHF	0.071	0.929	1.077
Environmental	NDVI	0.259	0.741	1.349
	FED	0.250	0.750	1.333
	FLI	0.199	0.801	1.248
	SHMC	0.086	0.914	1.094

Principal Component Analysis (PCA) was applied separately to the environmental and social indicator sets to extract dominant dimensions of vulnerability.

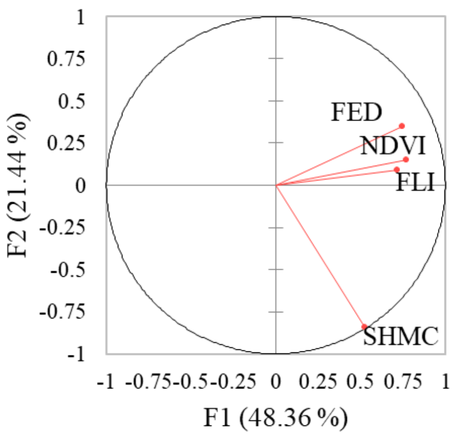

Figure 6. PCA of environmental indicators (F1–F2 correlation circle).

For the environmental indicators, the first component (F1; eigenvalue = 1.934) explains 48.36% of the total variance, while the first two components cumulatively explain 69.80%, indicating that environmental conditions can be effectively represented by a reduced two-dimensional structure (Figure 6). Prior to PCA, all indicators were directionally standardized such that higher values correspond to higher environmental vulnerability. In particular, NDVI and forest ecosystem density were inverted to reflect reduced vegetation condition and lower ecosystem integrity. As a result, the component loadings can be interpreted consistently in terms of increasing environmental stress.

F1 shows consistently positive and relatively high loadings for the Normalized Difference Vegetation Index (NDVI; 0.767), forest ecosystem density (0.748), forest loss intensity (0.717), and soil heavy metal contamination (0.522). This pattern indicates that F1 captures a generalized environmental vulnerability gradient, where degraded vegetation condition, reduced ecosystem integrity, forest disturbance, and soil-related stressors co-vary spatially.

Although only F1 satisfies the Kaiser criterion, component retention was guided by explained variance and interpretability. The second component (F2), accounting for 21.44% of the variance, is strongly dominated by soil heavy metal contamination (loading = −0.841), indicating a distinct contamination-related dimension that is partially independent from vegetation and forest-related processes.

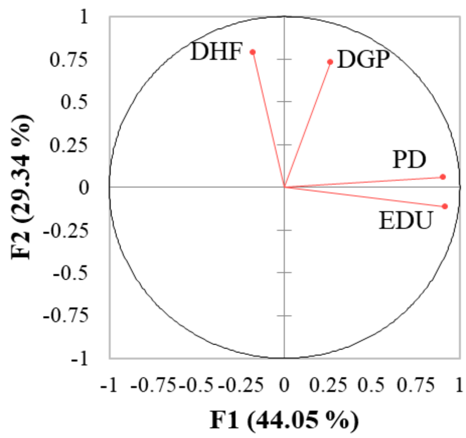

Figure 7. PCA of social indicators (F1–F2 correlation circle).

For the social indicators, two principal components with eigenvalues greater than 1 (F1 = 1.762; F2 = 1.174) jointly explain 73.39% of the total variance, suggesting a well-defined two-dimensional structure (Figure 7). All social indicators were standardized to ensure that higher values represent higher social vulnerability. In particular, education level was inverted so that lower educational attainment corresponds to higher vulnerability.

The first component (F1; 44.05%) is strongly associated with population density (0.908) and education level (0.915). Under the adopted standardization, this component reflects a socio-demographic vulnerability gradient characterized by high population pressure combined with reduced human capital. The second component (F2; 29.34%) is dominated by distance to health facilities (0.788) and the proportion of disadvantaged groups (0.733), representing a dimension of social sensitivity and limited access to essential services.

High population density may increase exposure concentration, but its contribution to vulnerability depends on adaptive capacity; in this study, its combination with low education levels suggests increased sensitivity rather than development advantage.

Overall, the PCA results indicate that environmental vulnerability is primarily structured by a dominant gradient integrating vegetation degradation, ecosystem disturbance, and contamination processes, with a secondary dimension specifically associated with soil contamination. Social vulnerability, in turn, is organized along two orthogonal axes reflecting (i) population pressure combined with reduced adaptive capacity and (ii) social sensitivity and service accessibility.

Because all indicators were directionally standardized prior to analysis, the extracted components provide a consistent and interpretable representation of vulnerability gradients. These components form a robust basis for subsequent clustering analysis and spatial typology identification.

3.2. Clustering of Vulnerability Typologies

K-means clustering was applied to the principal component scores derived from both social and environmental domains to delineate spatial typologies of vulnerability under landslide impacts. The low inter-correlation among the PCA components, with only a moderate association between the primary social and environmental axes (r = 0.507), indicates that these dimensions capture largely independent yet complementary aspects of vulnerability, justifying their integration within a unified clustering framework.

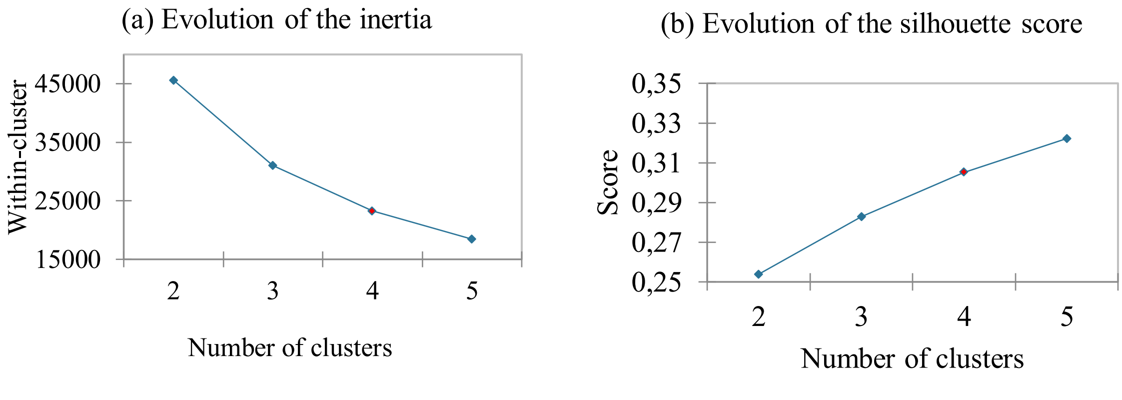

Figure 8. Determination of the optimal number of clusters based on (a) within-cluster inertia (elbow method) and (b) silhouette analysis.

The optimal number of clusters was determined using both inertia and silhouette metrics. Although clustering performance improved progressively from K = 2 to K = 5, the increase in silhouette score beyond K = 4 was marginal (0.305 at K = 4 versus 0.322 at K = 5) (Figure 8). The K = 5 solution introduced a small and spatially fragmented cluster with limited interpretability, which did not provide meaningful additional insight into vulnerability patterns. Therefore, K = 4 was retained as the most appropriate solution, ensuring a balance between statistical robustness and spatial interpretability while avoiding over-segmentation.

The resulting four clusters represent distinct socio-environmental vulnerability typologies, characterized by contrasting configurations of social and environmental conditions (Table 3; Figure 9). Cluster 3 (Moderately stable transitional areas) exhibits moderately low scores in both social and environmental dimensions (Social PC1 = −0.071; Environmental PC1 = −0.370), representing intermediate conditions. These areas function as transitional zones, where vulnerability is not pronounced but may locally increase due to spatial heterogeneity in environmental or social factors.

In contrast, Cluster 2 (Environmental stress areas) is defined by a high Environmental PC1 score (1.973), while social pressure remains moderate (Social PC1 = 0.177). This typology reflects areas where vulnerability is primarily driven by environmental degradation processes, including vegetation disturbance and ecosystem instability, rather than socio-demographic constraints.

Cluster 1 (Low-pressure environmentally stable areas) is characterized by low values of both Social PC1 (−0.625) and Environmental PC1 (−0.844), indicating low population pressure and relatively intact environmental conditions. These areas represent the least vulnerable typology, where both social and environmental stressors are minimal.

Cluster 4 (Compound vulnerability hotspots) is distinctly separated in PCA space, with exceptionally high values for both Social PC1 (5.246) and Environmental PC1 (2.795). These hotspots represent areas where environmental degradation and human pressure co-occur, potentially reinforcing slope instability through combined effects on land cover, drainage alteration, and surface disturbance. Although limited in spatial extent, these areas are of particular importance due to their potential for severe impacts.

The distribution of clusters in PCA space further illustrates the structural differences among typologies. The environmental PCA space (Figure 9a) shows a clear gradient along Environmental PC1, with strong separation between environmentally stressed areas (Cluster 2) and compound vulnerability hotspots (Cluster 4). In contrast, the social PCA space (Figure 9b) exhibits a more discrete and vertically aligned pattern, reflecting the aggregation of social variables at the commune level, which results in repeated values and linear clustering structures.

Despite this structural constraint, meaningful differentiation remains evident. Cluster 4 is clearly isolated with high Social PC1 values, indicating concentrated population pressure, while Cluster 1 is associated with higher Social PC2 values, reflecting limitations in accessibility and adaptive capacity. Partial overlap between Clusters 2 and 3 suggests that social variables alone provide limited discrimination in intermediate conditions.

Table 3. Cluster centroids in PCA space representing socio-environmental vulnerability typologies.

Cluster	Name	Social PC1	Social PC2	Environmental PC1	Environmental PC2
1	Low-pressure environmentally stable areas	-0.625	0.606	-0.844	0.634
2	Environmental Stress Areas	0.177	-0.326	1.973	0.231
3	Moderately stable transitional areas	-0.071	-0.533	-0.370	-0.761
4	Compound Vulnerability Hotspots	5.246	0.366	2.795	-0.113

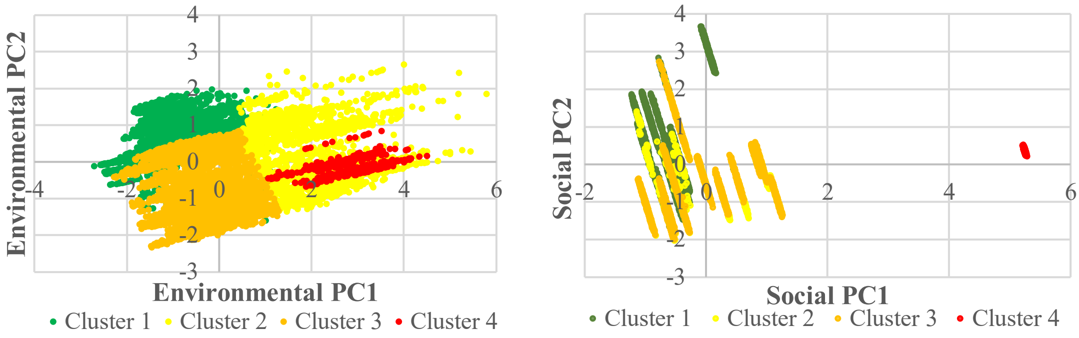

Figure 9. Distribution of K-means clusters in PCA space representing socio-environmental vulnerability typologies in Hoa An: (a) environmental PCA and (b) social PCA.

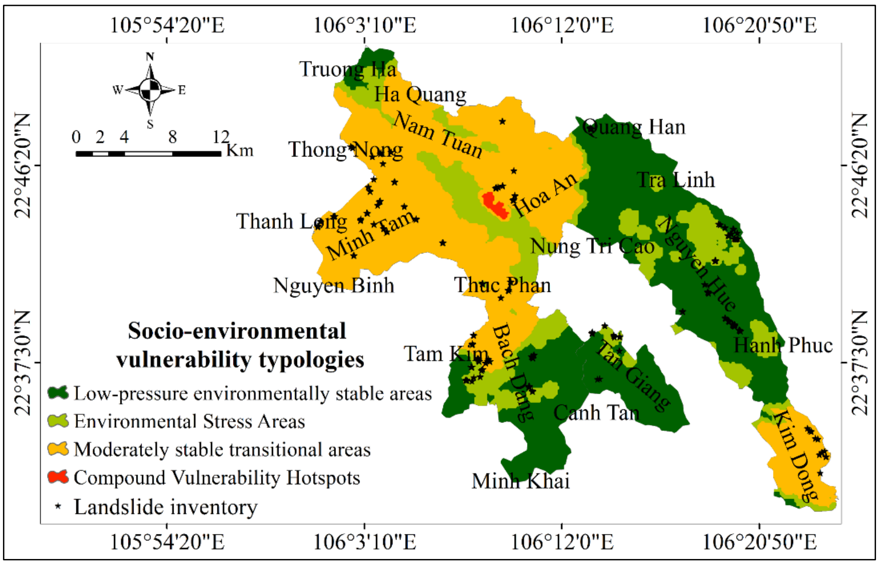

Figure 10. Spatial distribution of socio-environmental vulnerability typologies in Hoa An area.

The spatial distribution of these typologies (Figure 10) highlights their geographic expression across the study area. Cluster 1 dominates relatively undisturbed regions, while Cluster 2 and Cluster 3 occupy intermediate zones with varying degrees of environmental and social influence. Cluster 4 appears as spatially concentrated hotspots.

Overall, the contrast between PCA space representation (Figure 8) and mapped typologies (Figure 9) highlights the complementary roles of environmental gradients and socially structured constraints in shaping vulnerability patterns. In particular, the clear separation of Cluster 4 and the prominence of Cluster 2 emphasize the dominant role of environmental degradation in structuring vulnerability at the landscape scale.

3.3. Relationship between Vulnerability Typologies and Landslide Density

The relationship between socio-environmental vulnerability typologies and landslide density (LSD) reveals a structured but non-linear pattern across the study area (Table 4; Figure 8). Mean LSD is lowest in Cluster 1 (4.23 events/km^2) and highest in Cluster 4 (6.42

events/km²), indicating a clear contrast between relatively stable areas and compound vulnerability hotspots. This pattern confirms that vulnerability does not act as a controlling factor of landslide initiation but rather modulates the spatial concentration and persistence of landslide activity.

Cluster 4 (compound vulnerability hotspots), which covers 34.01% of the study area, exhibits both the highest mean LSD and the largest proportion of high-LSD zones (35.06%). This pattern indicates a spatially extensive concentration of landslide occurrence, where environmental degradation indicators and population pressure co-occur. Rather than reflecting a single dominant driver, this cluster represents areas where multiple vulnerability factors coincide spatially.

Cluster 2 and Cluster 3 represent intermediate vulnerability conditions but display contrasting spatial patterns of landslide occurrence. Cluster 2 shows a mean LSD of 5.03 events/km² and a relatively lower proportion of high-LSD areas (16.62%), indicating a more diffuse distribution of landslide activity. This suggests that landslides in Cluster 2 are spatially widespread but moderate in intensity.

In contrast, Cluster 3 exhibits a slightly higher mean LSD (5.21 events/km²) despite covering only 5.73% of the study area and contains a disproportionately high share of high-LSD zones (24.72%). The presence of relatively high landslide density in transitional areas suggests that localized geomorphic controls, such as lithology or structural discontinuities, may dominate over aggregated socio-environmental conditions. Given that this cluster represents transitional socio-environmental conditions. This pattern suggests that localized geomorphic controls, such as lithological contrasts, structural discontinuities, or slope morphology, may dominate landslide occurrence in these areas, overriding aggregated socio-environmental conditions.

This contrast highlights two distinct modes of landslide manifestation, reflecting the interaction between regional-scale environmental degradation processes and local-scale geomorphic controls: (i) diffuse and spatially extensive landslide activity associated with broader vulnerability patterns (Cluster 4 and partly Cluster 2), and (ii) localized, high-intensity concentrations occurring within transitional conditions (Cluster 3).

Table 4. Landslide density (LSD) characteristics across vulnerability typologies

Cluster	Area (%)	Mean LSD	Max LSD	High-LSD Area (%)
1	46.97	4.23	26.24	20.611
2	13.29	5.03	26.33	16.615
3	5.73	5.21	26.29	24.718
4	34.01	6.42	26.22	35.062

Cluster 1, which dominates the study area (46.97%), exhibits the lowest mean LSD and a relatively low proportion of high-LSD zones (20.61%), indicating generally stable conditions. However, the presence of scattered landslide occurrences within this cluster suggests that landslide activity is not entirely constrained by vulnerability typologies and may still occur under specific local conditions not represented in the current indicator set.

Despite clear differences in mean LSD and spatial patterns, maximum LSD values remain relatively consistent across all clusters (approximately 26 events/km²). This indicates that extreme landslide concentrations are governed by localized triggering mechanisms rather than broader vulnerability conditions. Therefore, maximum LSD likely reflects localized triggering processes, whereas mean LSD provides a more robust indicator of the broader vulnerability structure captured in this study.

4. Discussion

The results demonstrate that socio-environmental vulnerability exhibits a structured but non-linear association with landslide density, reflecting the interaction between geomorphic controls and human-environment processes rather than a direct causal relationship. The observed patterns indicate that vulnerability does not directly control landslide occurrence, but rather influences how landslides are spatially distributed and concentrated across the landscape.

The highest levels of landslide density observed in Cluster 4 highlight the importance of compounded vulnerability conditions, where environmental degradation and population pressure co-occur. These areas represent zones where multiple stressors overlap, potentially amplifying the spatial extent of landslide occurrence. However, this relationship does not imply a direct control of vulnerability on landslide initiation, as slope failure is primarily governed by terrain properties and triggering mechanisms such as rainfall.

A key finding of this study is the distinction between diffuse and localized landslide patterns. While Cluster 4 and, to some extent, Cluster 2 exhibit spatially extensive and moderate landslide activity, Cluster 3 shows localized but relatively intense concentrations of landslides. This suggests that areas with intermediate vulnerability conditions can still experience high landslide density, depending on localized processes that are not captured by aggregated socio-environmental indicators. Such behavior reinforces the idea that vulnerability shapes the spatial manifestation of landslides rather than their occurrence alone.

The absence of a strictly increasing relationship between vulnerability typologies and landslide density further emphasizes the complexity of landslide systems. While general trends can be identified, the spatial variability observed across clusters indicates that landslide occurrence is influenced by multiple interacting factors operating at different scales. In this context, vulnerability typologies provide a useful framework for identifying areas where landslide impacts may be more pronounced, even if they do not fully explain the underlying physical processes.

By decoupling vulnerability assessment from landslide occurrence, this framework enables an independent evaluation of their spatial relationship, thereby avoiding circular reasoning commonly associated with susceptibility-based analyses. Using landslide density (LSD) as an external reference rather than a model input further supports an objective assessment of spatial correspondence and reinforces a clearer conceptual distinction between vulnerability and hazard components.

However, several limitations should be acknowledged. First, the analysis does not include terrain-based variables or triggering factors such as rainfall intensity, which are known to influence landslide processes. As a result, localized variations in landslide density cannot be fully explained within the current framework. Second, landslide density represents past occurrences and does not capture temporal dynamics or event-specific conditions. In addition, the integration of commune-level socio-economic data with raster-based environmental variables introduces scale mismatch, which may affect the spatial precision of vulnerability patterns.

Future research should integrate terrain parameters and dynamic triggering factors to better explain localized landslide concentrations and improve the linkage between vulnerability patterns and landslide processes. Incorporating such variables would enhance the interpretability of vulnerability–landslide relationships. It would also clarify the role of vulnerability as a spatial modifier of landslide impacts rather than a predictor of landslide occurrence.

5. Conclusion

This study reveals that landslide occurrence in the Hoa An karst landscape is not uniformly distributed but exhibits clear spatial differentiation associated with underlying socio-environmental conditions. Areas characterized by compound vulnerability, where environmental degradation and population pressure co-occur, show the highest landslide density and the most spatially extensive concentration of landslide activity. In contrast, relatively stable areas display low and scattered landslide occurrence, indicating that vulnerability gradients correspond to distinct modes of landslide manifestation across the landscape.

Importantly, the relationship between vulnerability and landslide density is non-linear. While high vulnerability areas tend to concentrate landslide activity, transitional zones with moderate vulnerability can also exhibit localized but relatively high landslide densities. This pattern suggests that landslide occurrence is jointly influenced by regional-scale vulnerability structures and local geomorphic controls, rather than being directly determined by vulnerability alone.

Methodologically, the integration of PCA and K-means clustering proves effective in reducing complex socio-environmental data into interpretable spatial typologies, allowing the identification of distinct vulnerability configurations that are not evident from individual indicators. This typology-based approach provides a more nuanced understanding of how vulnerability structures modulate the spatial concentration of landslides, rather than predicting their occurrence.

By explicitly separating vulnerability assessment from landslide occurrence and using landslide density as an independent reference, the study avoids circular reasoning and enables a clearer interpretation of vulnerability–landslide relationships.

The analysis does not include terrain or triggering variables, which limits the explanation of localized landslide patterns. Future work should integrate geomorphological and dynamic factors to improve process understanding and support more comprehensive risk assessment.

Author Contributions

Phan Thi Mai Hoa: Conceptualisation, Methodology, Data curation, Visualization Fieldwork, Writing – review & editing. Nguyen Quoc Phi: Validation, Supervision. Nguyen Thi Cuc: Investigation, Formal analysis. Tran Thi Thu Huong: Fieldwork, Data curation.

Acknowledgement

This research was funded by the 2026 institutional science and technology project at Hanoi University of Mining and Geology, entitled “Application of Clustering Algorithms to Assess Economic Development and Landslide-Induced Resilience in the Non Nuoc Cao Bang Geopark” (Project Code: T26-37). The authors gratefully acknowledge this support.

References

[1] M. J. Froude, D. N. Petley, Global Fatal Landslide Occurrence from 2004 to 2016, Natural Hazards and Earth System Sciences, Vol. 18, No. 8, 2018, pp. 2161-2181, https://doi.org/10.5194/nhess-2018-49.

[2] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, F. Guzzetti, A Review of Statistically-based Landslide Susceptibility Models, Earth-Science Reviews, Vol. 180, 2018, pp. 60-91, https://doi.org/10.1016/j.earscirev.2018.03.001.

[3] C. Idukunda, C. Michellier, F. D. Longueville, E. Twarabamenye, S. Henry, Assessing Community Vulnerability to Landslide and Flood in Northwestern Rwanda, International Journal of Disaster Risk Reduction, Vol. 123, 2025, pp. 105329, https://doi.org/10.1016/j.ijdrr.2025.105329.

[4] H. Lee et al., Synthesis Report of the IPCC Sixth Assessment Report (AR6), Longer Report, IPCC, Intergovernmental Panel on Climate Change (IPCC), 2023.

[5] D. Amaratunga et al., United Nations Office for Disaster Risk Reduction Regional Assessment Report on Disaster Risk Reduction 2023: Europe and Central Asia, 2023.

[6] UNDRR 2017, The Sendai Framework Terminology on Disaster Risk Reduction, Vulnerability, [Online] Available: https://www.undrr.org/terminology/vulnerability (accessed on: September 1st, 2025).

[7] C. Idukunda, C. Michellier, F. D. Longueville, E. Twarabamenye, S. Henry, Assessing Community

Vulnerability to Landslide and Flood in Northwestern Rwanda, International Journal of Disaster Risk Reduction, Vol. 123, 2025, pp. 105329, https://doi.org/10.1016/j.ijdrr.2025.105329.

[8] N. T. Kien, V. T. H. Lien, P. L. H. Linh, N. Q. Thanh, Landslide Susceptibility Mapping Based on the Combination of Bivariate Statistics and Modified Analytic Hierarchy Process Methods: A Case Study of Tinh Tuc Town, Nguyen Binh District, Cao Bang Province, Vietnam, Journal of Disaster Research, Vol. 16, No. 4, 2021, pp. 521-528, https://doi.org/10.20965/jdr.2021.p0521.

[9] N. T. T. Pham, D. Nong, A. R. Sathyan, M. Garschagen, Vulnerability Assessment of Households to Flash Floods and Landslides in the Poor Upland Regions of Vietnam, Climate Risk Management, Vol. 28, 2020, pp. 100215, https://doi.org/10.1016/j.crm.2020.100215.

[10] E. Tate, S. Rufat, M. A. Rahman, S. Hoover, Profiles of Social Vulnerability for Flood Risk Reduction, International Journal of Disaster Risk Reduction, Col. 118, 2025, pp. 105250, https://doi.org/10.1016/j.ijdrr.2025.105250.

[11] M. Englund, M. Vieira Passos, K. André, Å. Gerger Swartling, L. Segnestam, K. Barquet, Constructing a Social Vulnerability Index for Flooding: Insights from A Municipality in Sweden, Frontiers in Climate, Vol. 5, 2023, pp. 1038883, https://doi.org/10.3389/fclim.2023.1038883.

[12] E. E. Koks, B. Jongman, T. G. Husby, W. J. Botzen, Combining Hazard, Exposure and Social Vulnerability to Provide Lessons for Flood Risk Management, Environmental Science & policy, Vol. 47, 2015, pp. 42-52, https://doi.org/10.1016/j.envsci.2014.10.013.

[13] Cao Bang Provincial Statistics Office, 2021. Complete results of the 2019 Population and Housing Census in Cao Bang Province, Cao Bang Provincial Statistics Office, 2021, [Online], Available: https://thongkecaobang.gso.gov.vn (accessed on: September 1st, 2025).

[14] H. D. Vu, N. T. P. Nguyen, Y. T. H. Ngo, T. D. Le, Geotourism Current State and Future Prospects: A Case Study in the Cao Bang UNESCO Global Geopark, Vietnam, GeoJournal of Tourism and Geosites, Vol. 43, No. 3, 2022, pp. 1063-1070, https://doi.org/10.30892/gtg.43327-921.

[15] N. J. M. Rodríguez, J. Vega, O. B. Zanabria, J. D. G. Ruiz, S. Botero, Towards an Understanding of Landslide Risk Assessment and its Economic Losses: A Scientometric Analysis, Landslides, Vol. 21, No. 8, 2024, pp. 1865-1881, https://doi.org/10.1007/s10346-024-02272-2.

[16] S. L. Cutter, The Vulnerability of Science and the Science of Vulnerability, Annals of the Association of American Geographers, Vol. 93, No. 1, 2023, pp. 1-12, https://doi.org/10.1111/1467-8306.93101.

[17] J. Birkmann, Measuring Vulnerability to Promote Disaster-resilient Societies: Conceptual Frameworks and Definitions, Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies, Vol. 1, No. 9, 2006, pp. 3-7.

[18] G. Tóth, T. Hermann, G. Szatmári, L. Pásztor, Maps of Heavy Metals in the Soils of the European Union and Proposed Priority Areas for Detailed Assessment, Science of the Total Environment, Vol. 565, 2016, pp. 1054-1062, https://doi.org/10.1016/j.scitotenv.2016.05.115.

[19] P. I. Gomes, U. Aththanayake, W. Deng, A. Li, W. Zhao, T. Jayathilaka, Ecological Fragmentation Two Years after a Major Landslide: Correlations between Vegetation Indices and Geo-Environmental Factors, Ecological Engineering, Vol. 153, 2020, pp. 105914, https://doi.org/10.1016/j.ecoleng.2020.105914.

[20] A. L. Handwerger, E. J. Fielding, S. S. Sangha, and D. P. Bekaert, Landslide Sensitivity and Response to Precipitation Changes in Wet and Dry Climates, Geophysical Research Letters, Vol. 49, No. 13, 2022, pp. e2022GL099499, https://doi.org/10.1029/2022GL099499.

[21] C. J. V. Westen, E. Castellanos, S. L. Kuriakose, Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview, Engineering Geology, Vol. 102, No. 3-4, 2008, pp. 112-131, https://doi.org/10.1016/j.enggeo.2008.03.010.

[22] I. T. Jolliffe, J. Cadima, Principal Component Analysis: A Review and Recent Developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 374, No. 2065, 2016, pp. 20150202, https://doi.org/10.1098/rsta.2015.0202.

[23] P. J. Rousseeuw, Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, Journal of Computational and Applied Mathematics, Vol. 20, 1987, pp. 53-65, https://doi.org/10.1016/0377-0427(87)90125-7.