# JOURNAL OF HYDRO - METEOROLOGY

ISSN 2525 - 2208

VIETNAM METEOROLOGICAL AND HYDROLOGICAL ADMINISTRATION

No 23
06-2025




**Publishing licence**
No: 166/GP-BTTTT - Ministry of Information and Communication dated 17/04/2018

**Editorial office**
No 8 Phao Dai Lang, Dong Da, Ha Noi
Tel: 024.39364963
Email: tapchikttv@gmail.com

**Engraving and printing**
Vietnam Agriculture Investment Company Limited
Tel: 0243.5624399

**JOURNAL OF HYDRO-METEOROLOGY**
**Volume 23 - 6/2025**

# TABLE OF CONTENT

JOURNAL OF HYDRO-METEOROLOGY



*Research Article*

# Integrating Dempster-Shafer theory, certainty factors and topographic indices for landslide susceptibility analysis in Ha Quang district, Cao Bang province

**Hoa Mai Thi Phan[1], Hanh Hong Tran[2]*, Phi Quoc Nguyen[2], Hoa Thanh Thi Pham[2]**

[1] Graduate University of Science and Technology, Vietnam Academy of Science and Technology; phanthimaihoa@humg.edu.vn

[2] Hanoi University of Mining and Geology; tranhonghanh@humg.edu.vn; nguyenquocphi@humg.edu.vn; phamthithanhhoa@humg.edu.vn

*Corresponding author: tranhonghanh@humg.edu.vn; Tel.: +84–988150099

Received: 27 January 2025; Accepted: 17 March 2025; Published: 8 June 2025

**Abstract:** Ha Quang District, situated within the Non Nuoc Cao Bang Geopark, is an area of significant geological, ecological, and cultural value and has been recognized by UNESCO as a member of the Global Geoparks Network. However, the region is also prone to various geological hazards, with landslides representing a primary concern. This study aims to apply the Dempster-Shafer (DS) theory and Certainty Factor (CF) to analyze landslide susceptibility in the study area using a Geographic Information System (GIS). A total of 196 landslides were documented using historical records, Google Earth imagery, and field surveys to create a comprehensive inventory map. Seven conditioning factors, including slope, Topographic Roughness Index (TRI), Topographic Wetness Index (TWI), Stream Power Index (SPI), Mass Balance Index (MBI), Normalized Difference Vegetation Index (NDVI), and rainfall, were integrated as thematic layers for analysis. The belief map, representing the most reliable integrated landslide susceptibility model, was assessed using receiver operating characteristic (ROC) analysis and area under the curve (AUC). The evaluation revealed that the model achieved an overall accuracy of 74.5%. To compare performance, the Certainty Factor (CF) model was also applied, obtaining a success rate of 67%. The results indicated that the Dempster-Shafer (DS) theory demonstrated superior predictive capability over the CF model. It addresses a critical gap in landslide susceptibility research by improving predictive accuracy compared to traditional GIS, particularly in handling uncertainty through the Dempster-Shafer theory. These findings are crucial for developing effective landslide risk mitigation strategies and optimizing land-use planning to enhance infrastructure protection and sustainable development.

**Keywords:** Dempster-Shafer; Certainy factors; Topographic indices; Landslide susceptibility; Ha Quang.

## 1. Introduction

Landslides represent a significant natural hazard, frequently leading to substantial property damage and economic losses. These impacts primarily stem from the high costs associated with rebuilding essential infrastructure, including roads, residential structures, and public facilities [1]. Landslides result from a combination of natural processes and anthropogenic activities. Therefore, assessing landslide susceptibility is essential for effectively forecasting these events. Areas predicted to have a high risk of landslides are more likely to experience future occurrences. The key controlling factors include rainfall conditions, slope, and the characteristics of the underlying bedrock and soil [2]. In recent years, natural disasters and

environmental hazards driven by climate change have grown increasingly complex. Seasonal rainfall patterns have become unpredictable, significantly elevating the risk of landslides, flash floods, and debris flows [3]. As a result, disaster prevention and mitigation efforts remain inadequate, causing significant losses for both residents and the government. Over the past decades, landslide susceptibility mapping has been studied using various methods, such as Geographic Information Systems (GIS), probabilistic models, logistic regression models, and other analytical approaches. Some data networks utilize semantic segmentation techniques, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and the integration of high-level object-specific information through data partitioning processes [4–10]. The study [4] utilized three data mining techniques to develop a landslide susceptibility map for Hanyuan County, China. These methods included an adaptive neuro-fuzzy inference system integrated with the frequency ratio (ANFIS-FR), a generalized additive model (GAM), and a support vector machine (SVM). The analysis was conducted using twelve conditioning factors: slope aspect, altitude, slope angle, topographic wetness index (TWI), plan curvature, profile curvature, proximity to rivers, proximity to faults, proximity to roads, land use, normalized difference vegetation index (NDVI), and lithology. The results revealed that all three models exhibited strong predictive performance, with the SVM model achieving the highest accuracy, followed by ANFIS-FR and GAM.

The study [5] utilized the Random Forest (RF) model to evaluate multi-hazard risks, encompassing floods, wildfires, and earthquakes, while also integrating an assessment of social vulnerability within the study area. The study [6] implemented an artificial neural network (ANN) enhanced through particle swarm optimization (PSO) to develop a landslide susceptibility map (LSM). The input dataset comprised a range of geomorphological and environmental variables, including elevation, slope aspect, slope gradient, curvature, soil type, lithology, proximity to roads, rivers, and faults, land use, stream power index (SPI), and topographic wetness index (TWI). The findings indicated that both models demonstrated strong predictive capabilities; however, according to the applied ranking system, the PSO-ANN model achieved superior accuracy compared to the standalone ANN model. A comprehensive study was undertaken to bridge critical gaps in the understanding and prediction of landslide susceptibility in the Naqadeh Region [7]. This research identified several key factors contributing to landslide occurrence, including elevation, aspect, slope gradient, lithology, drainage density, proximity to rivers, weathering, land cover, precipitation, vegetation, proximity to faults, roads, and urban areas. The results demonstrated that the multilayer perceptron (MLP) model achieved the highest overall accuracy in producing landslide susceptibility maps. The study [8] investigated the impact of both dynamic and static factors on landslide prediction by employing the XGBoost machine learning (ML) algorithm to develop landslide susceptibility maps. The findings revealed an expansion in the areas classified as high and very high susceptibility, with the most significant increase observed in scenarios incorporating both dynamic factors.

The study [9] explored two distinct approaches for selecting landslide-free random points the slope threshold method and the buffer-based technique while conducting a comparative analysis of five machine learning models for landslide susceptibility mapping: Support Vector Machine (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The study incorporated fourteen geospatial data layers, including topographic variables, soil characteristics, geological data, and land cover features. The results highlight the effectiveness of machine learning models in evaluating and predicting landslide susceptibility. In a separate study, the researchers utilized both Random Forest (RF) and Support Vector Machine (SVM) to construct landslide susceptibility mapping models, employing a random search optimization technique [10]. The interpretability analysis revealed that the normalized difference vegetation index (NDVI) and proximity to roads were the most influential factors contributing to landslide occurrence.

In recent years, the semi-quantitative approach employing the Dempster-Shafer (DS) belief function model has been utilized to represent evidence uncertainty through belief functions, supporting the assessment and zonation of landslide susceptibility [11]. Research findings indicate that the DS model demonstrates high accuracy in predicting areas with a high risk of landslides [12–17]. In the study [12], the methodology comprised two main stages: the fusion of multi-source InSAR data using Dempster-Shafer evidence theory and the implementation of a two-step decision rule. Preliminary findings for the Luoshui-Baini section, derived from both ascending and descending orbit data, identified a total of 68 landslides. The study [13] presents a data integration framework for landslide susceptibility mapping, leveraging the Dempster-Shafer theory of evidence to synthesize multiple geospatial datasets. The case study findings indicate that the proposed approach effectively integrates diverse data sources while outperforming the traditional logistic regression model in predictive accuracy.

The study [14] conducted a landslide susceptibility assessment in the Haraz watershed, Iran, by employing two distinct methodologies: The Dempster-Shafer model and the Weights-of-Evidence model within a GIS framework. The analysis incorporated 11 landslide conditioning factors, including slope gradient, aspect, elevation, proximity to drainage networks, roads, and rivers, lithology, land use, topographic wetness index, stream power index, and slope length. The resulting susceptibility maps offer valuable insights for land use planning and serve as a basis for prioritizing strategies to mitigate and reduce future landslide risks in the case study. The study [15] conducted a comparative evaluation of three GIS-based models Dempster-Shafer (DS), logistic regression (LR), and artificial neural network (ANN) to assess landslide susceptibility in the Shangzhou District of Shangluo City, China. The study utilized 14 landslide conditioning factors, including altitude, slope gradient, slope aspect, topographic wetness index, sediment transport index, stream power index, plan curvature, profile curvature, lithology, rainfall, proximity to rivers, roads, and faults, as well as the normalized difference vegetation index (NDVI), to identify areas most susceptible to landslides.

In a separate study, researchers applied the object-based geons aggregation model to map landslide susceptibility across Austria and evaluate the potential improvement achieved by incorporating the Dempster-Shafer theory (DST) [16]. The analysis considered nine conditioning factors, including elevation, slope, aspect, land cover, rainfall, proximity to drainage networks, faults, and roads, as well as lithology. Ishola et al. utilized the Dempster-Shafer Theory of Evidential Belief Function (DST-EBF) model to develop a groundwater prospectivity zonation (GWPZ) map for the study areas. The analysis incorporated remote sensing, geological, field geophysical, and hydrological datasets, integrating groundwater conditioning factors (GWCFs) within a Geographic Information System (GIS) framework [17]. Machine learning models often achieve high accuracy but require large training datasets and lack interpretability, making them less suitable for regions with limited data availability. In contrast, the Dempster-Shafer (DS) theory is particularly advantageous for handling uncertainty and integrating multiple data sources, making it a more effective approach for landslide susceptibility assessment in data-scarce environments. Ha Quang District, with its steep terrain and fragile geological structures, has experienced frequent landslides, with 196 events recorded via remote sensing. These risks are further amplified by human activities such as deforestation and infrastructure development, while increasing rainfall variability due to climate change acts as a major triggering factor.

Considering these challenges, the DS approach and CF model with topographic indices prove to be an effective solution for generating accurate susceptibility maps in regions with limited data and dynamic environmental conditions. The research seeks to offer critical insights for disaster risk management and support the preservation of biodiversity, as well as archaeological, historical, and environmental assets within the geopark region.

## 2. Materials and Methods

### *2.1. Study area*

Ha Quang District, situated in Cao Bang Province, is a mountainous region within the Non Nuoc Cao Bang Geopark, located approximately 40 kilometers north of Cao Bang City. Geographically, the district is bordered by Trung Khanh District to the east, Bao Lac District to the west, Hoa An and Nguyen Binh Districts to the south, and China to the north (Figure 1). The district covers a total area of 810.96 km² and has a population of 59,467 people [18].

The study area is distinguished by its steep mountainous terrain, intricate geological structures, and unstable vegetation cover. As a highland region, it is highly vulnerable to environmental hazards, particularly landslides and flash floods. Prolonged heavy rainfall, coupled with human activities such as mining, road construction, and agricultural practices, further exacerbates the risk of landslides. These factors have led to substantial damage to property and infrastructure, disruptions to local livelihoods, and the degradation of the natural landscape.

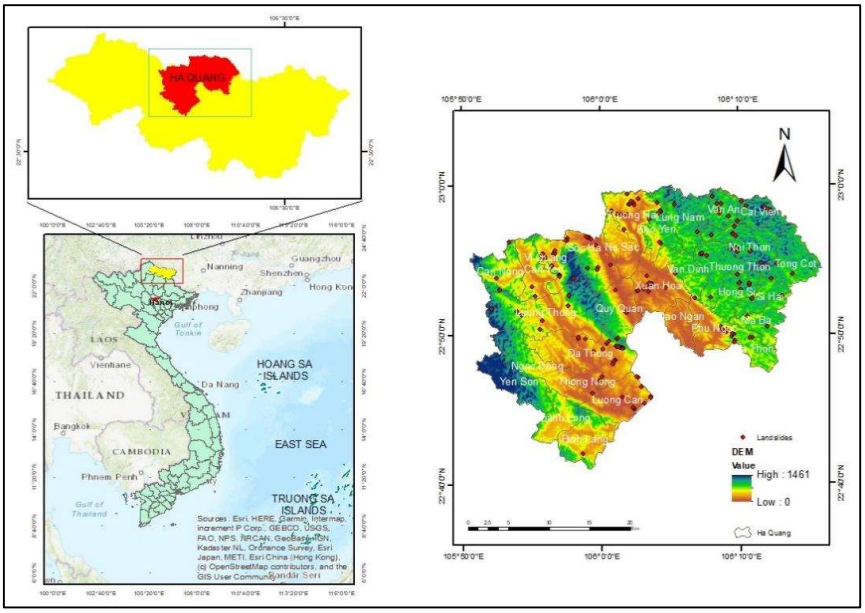


**Figure 1.** Administrative map of Ha Quang district, Cao Bang province.

### *2.2. Data collection*

The study was carried out in five key steps, as follows:

(1) Landslide Identification: Landslide locations were determined through optical remote sensing analysis of Google Earth imagery, supplemented by extensive field surveys.

(2) Selection and Extraction of Factors: Seven influencing factors were selected and extracted using multiple data sources, including Digital Elevation Models (DEM), Landsat 8 (OLI+), and satellite-based rainfall measurements.

(3) Application of the Dempster-Shafer Theory (DST) and CF model: The DST and CF model were employed to compute importance weights for each causative factor map.

(4) Generation of the DS and CF Map: The weighted factor maps were integrated with their corresponding importance weights to produce the DS susceptibility map and CF map.

By utilizing the distinct advantages of each model, this study seeks to conduct a comprehensive comparison of their predictive performance, enhancing the accuracy of landslide susceptibility mapping. A total of 196 landslide occurrences in the study area were identified through visual analysis of Google Earth imagery, utilizing historical statistical data

to develop a comprehensive landslide inventory map. These events are primarily concentrated in Da Thong, Truong Ha and Xuan Hoa, as depicted in Figure 2.

Many landslides are located along major river systems, including the Bang River and Bang River tributaries and on slopes adjacent to the DT204 provincial road. The observed landslides primarily consist of soil slips and minor, shallow-scale slope slides. Based on a comprehensive review of national and international research on landslide susceptibility, seven key conditioning factors were selected due to their established influence on landslide susceptibility.

The data used in this study was partially obtained from existing information and maps prepared by various organizations, while the remaining data was generated through field investigations. The rainfall data corresponds to the maximum daily precipitation recorded in August 2023, was provided by the US National Oceanic and Atmospheric Administration (NOAA). It was estimated using interpolation techniques based on data from meteorological stations. The DEM (30m tiff) and Landsat 8 (OLI+) imagery, captured on December 22, 2023, was sourced from the USGS [19], with a spatial resolution of 30 meters for multispectral images and 15 meters for the panchromatic band.

The selection of the seven conditioning factors: Slope, Topographic Ruggedness Index (TRI), Topographic Wetness Index (TWI), Stream Power Index (SPI), Mass Balance Index (MBI), Normalized Difference Vegetation Index (NDVI), and Rainfall was based on their well-documented influence on landslide occurrence in previous studies [20, 21] and their suitability for the study area's environmental and geological conditions. These factors represent a comprehensive combination of topographic, hydrological, and vegetation-related parameters that directly impact slope stability.

Slope plays a fundamental role in gravitational instability, while TRI, TWI, SPI, and MBI account for terrain roughness, water retention, erosive potential, and mass balance dynamics, respectively. NDVI reflects vegetation cover, which contributes to slope reinforcement, whereas Rainfall serves as a primary triggering factor for landslides. These seven factors were prioritized over other potential variables, such as soil depth or fault density, due to their strong correlation with landslide susceptibility, their proven reliability in similar geographic settings [22]. The Dempster-Shafer and CF theory model is integrated with terrain indices to assess landslide susceptibility. The landslide-triggering factors used in the sensitivity analysis are classified into three primary categories in Table 1.

**Table 1.** The processed data and their sources.

| Data | Factor | Source | Processing Method |
|---|---|---|---|
| Topological | Slope (Raster) | Derived from DEM | Calculated using GIS tools to determine slope angle |
| | SPI (Stream Power Index) [23] | Derived from DEM | Computed from DEM using GIS-based hydrological analysis |
| | TRI (Terrain Ruggedness Index) (Raster) [24] | Derived from DEM | Derived using GIS tools to assess terrain roughness |
| | TWI (Topographic Wetness Index) (Raster) [25] | Derived from DEM | Calculated using DEM and flow accumulation data |
| | MBI (Mass Balance Index) (Raster) [26] | Derived from DEM | Estimated using DEM in combination with soil production and erosion data |
| Hydro-Meteorological | Rainfall (Raster) | CMORPH precipitation dataset of US National Oceanic and Atmospheric Administration (NOAA) | Interpolated from meteorological station data |
| Anthropogenic Hazard-Enhancing | NDVI (Normalized Difference Vegetation Index) (Raster) | Satellite imagery (Landsat 8) | Computed using NIR and Red bands of satellite images |

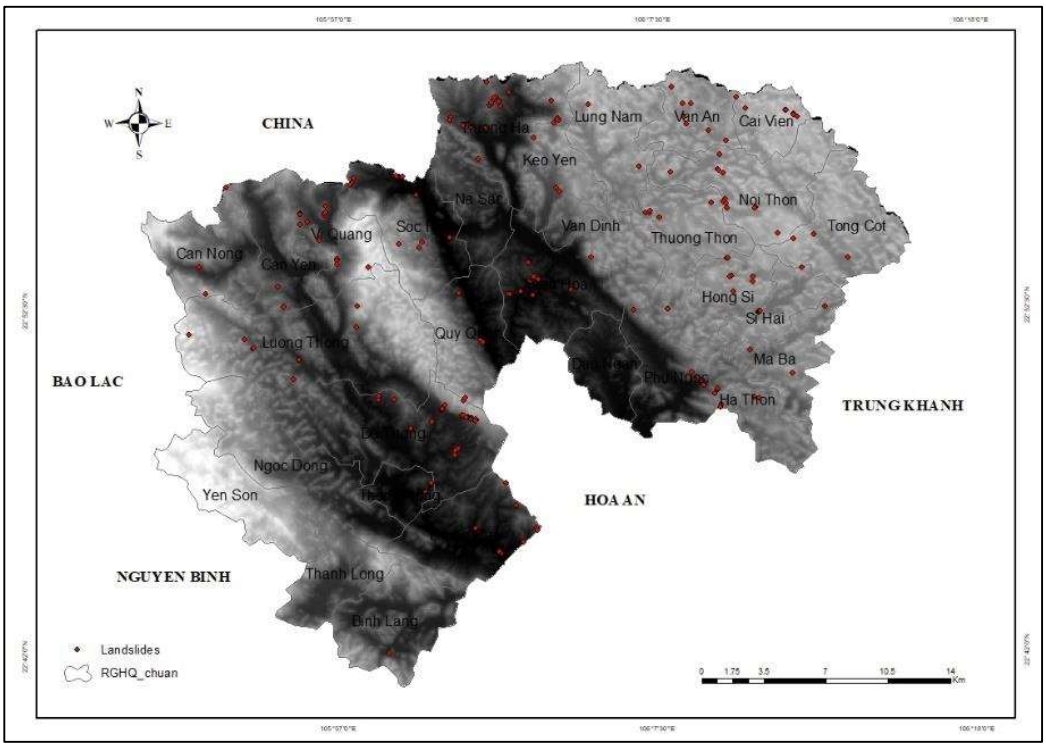


**Figure 2.** The landslide points in the study area.

## *2.3. Methodology*

### 2.3.1. Dempster-Shafer Theory (DS)

The Dempster-Shafer Theory (DS) is a powerful tool for handling uncertainty and integrating multiple data sources [27, 28]. Originally formulated by Arthur P. Dempster in 1967, the method was subsequently expanded and refined by Glenn Shafer in 1976.

This method is based on the theory of belief functions, where different data layers are represented as Evidence-Based Functions (EBF). These functions consist of four fundamental components [20].

- Belief (Bel): The lower bound of confidence that an area is prone to landslides.
- Plausibility (Pls): The upper bound, representing the maximum possible support for landslide occurrence.
- Uncertainty (Unc): The gap between Bel and Pls, calculated as Unc = Pls - Bel.
- Disbelief (Dis): The degree of opposition to landslide occurrence, given by Dis = 1 - Pls.

These components satisfy Bel + Unc + Dis = 1, ensuring a complete representation of uncertainty. To compute EBF values, seven spatial evidence maps are overlaid with a binary landslide inventory map. For each evidence class, the number of pixels intersecting landslides ($N(Cij \cap L)$) and non-landslide areas ($N(Cij) - N(Cij \cap L)$) is determined.

Integration of multiple evidence layers follows Dempster's rule of combination, refining landslide susceptibility predictions by merging individual EBF maps. This allows for a probabilistic yet flexible approach to hazard assessment.

$$Bel_{X_1X_2} = \frac{Bel_{X_1}Bel_{X_2} + Bel_{X_1}Unc_{X_2} + Bel_{X_2}Unc_{X_1}}{\beta} \quad (1)$$

$$Dis_{X_1X_2} = \frac{Dis_{X_1}Dis_{X_2} + Dis_{X_1}Unc_{X_2} + Dis_{X_2}Unc_{X_1}}{\beta} \quad (2)$$

$$\mathrm{Unc}_{X_1X_2} = \frac{\mathrm{Unc}_{X_1}\mathrm{Unc}_{X_2}}{\beta} \tag{3}$$

$$\mathrm{Pls}_{X_1X_2} = 1 - \mathrm{Bel}_{1-X_1X_2} \tag{4}$$

$$\beta = 1 - \mathrm{Bel}_{X_1}\mathrm{Dis}_{X_2} - \mathrm{Dis}_{X_1}\mathrm{Bel}_{X_2} \tag{5}$$

The landslide susceptibility map is developed based on Equation (6):

$$\mathrm{LSI} = \sum_{\mathrm{Bel}=1}^{n}\left(\mathrm{Bel} \times \mathrm{X}\right) \tag{6}$$

where LSI (Landslide Susceptibility Index) represents the index measuring the likelihood of landslides occurring; Bel (Belief Function) is a belief function based on Dempster-Shafer theory, indicating the confidence level of a specific conditioning factor in triggering landslides; X is the value of a variable or conditioning factor (e.g., slope, aspect, etc.); n is the total number of conditioning factors considered in the model.

### 2.3.2. Certainty Factor model

The Certainty Factor (CF) method is a statistical approach for assessing landslide susceptibility [29]. CF quantifies the relationship between landslide occurrence and influencing factors, with values ranging from -1 to 1: CF > 0: Strong positive correlation; CF < 0: Weak or negative correlation; CF = 0: No clear relationship. It is calculated as:

$$\mathrm{CF}_{ij} = \begin{cases} \dfrac{\mathrm{f}_{ij} - \mathrm{f}}{\mathrm{f}_{ij}\left(1 - \mathrm{f}\right)}, \text{if} \quad \mathrm{f}_{ij} > \mathrm{f} \\ \dfrac{\mathrm{f}_{ij} - \mathrm{f}}{\mathrm{f}_{ij}\left(1 - \mathrm{f}\right)}, \text{if} \quad \mathrm{f}_{ij} \le \mathrm{f} \end{cases} \tag{7}$$

where $f_{ij}$ and f represent landslide densities in specific classes and the overall study area, respectively.

### 2.3.3. Evaluation and comparison method

a) Statistical Validation of Factor Contributions

Multicollinearity, or the correlation among independent variables, is assessed in regression analysis using two key indicators: Variance Inflation Factor (VIF) and Tolerance Factor (TOF). The values of VIF and TOL are determined using Equations (8) and (9).

$$\mathrm{VIF} = \frac{1}{1 - \mathrm{R}_i^2} \tag{8}$$

$$\mathrm{TOF} = \frac{1}{\mathrm{VIF}} \tag{9}$$

where $\mathrm{R}_i^2$ represents the coefficient of determination for the independent variable $i$ when regressed against the other predictors in the model.

VIF measures how much the variance of a regression coefficient is inflated due to multicollinearity. A VIF < 2 indicates low multicollinearity, 2-5 suggests moderate correlation, and > 5 requires corrective action [30].

TOF, the inverse of VIF, evaluates variable independence. A TOF > 0.5 indicates low multicollinearity, 0.2-0.5 suggests moderate correlation, and < 0.2 signals a potential issue [30].

Since VIF and TOF are reciprocals, a lower VIF corresponds to a higher TOF, ensuring better model stability. Monitoring these values helps optimize regression models by eliminating or transforming highly correlated variables, thereby improving reliability and predictive accuracy.

b) Evaluation and comparison method

There are several methods for evaluating the accuracy of landslide susceptibility maps, with field validation being the most reliable. However, this approach is often difficult and resource-intensive. To address this, various statistical and mathematical techniques are commonly used, including landslide density analysis, Success Rate Curve (SRC), Prediction Rate Curve (PRC), Chi-square validation, and Receiver Operating Characteristic (ROC) analysis.

In this study, the ROC curve was applied to assess the predictive performance of the models. The ROC curve is generated by plotting sensitivity against 100-specificity, which in this paper corresponds to the cumulative percentage of landslide occurrences versus the percentage of the landslide susceptibility index [31]. The Area Under the Curve (AUC) measures the model's accuracy, with values ranging from 0 to 1:

AUC < 0.5: The model performs worse than random prediction.

$0.5 \leq$ AUC < 0.7: The model has low predictive capability.

$0.7 \leq$ AUC < 0.9: The model has moderate to good predictive performance.

AUC $\geq$ 0.9: The model is highly accurate.

AUC $\approx$ 1.0: The model provides nearly perfect predictions.

## 3. Results and Discussions

a) Analysis of landslide susceptibility based on EBF and FR values

The landslide susceptibility map is estimated based on the scale, completeness, and accuracy of the landslide inventory map, as well as the maps of various landslide conditioning factors. The susceptibility maps were developed using the Dempster-Shafer (DS) models, incorporating key conditioning factors such as slope, Stream Power Index (SPI), Topographic Wetness Index (TWI), Mass Balance Index (MBI), and Terrain Ruggedness Index (TRI). The maps categorize susceptibility into five classes: very low, low, moderate, high, and very high. The results for the factors are listed in Table 2.

**Table 2.** Spatial relationships between the seven factors and landslide distribution.

| Factors | Class | N(Cij) | Ratio (%) | N(CijD) | Ratio (%) | Freq.ration (FR) | BelCij | DisCij | UncCij | PlsCij |
|---|---|---|---|---|---|---|---|---|---|---|
| Slope (°) | <10 | 770707 | 20.119 | 592 | 11.348 | 0.564 | 0.117 | 0.200 | 0.683 | 0.800 |
| | 10-20 | 1166297 | 30.446 | 1720 | 32.969 | 1.083 | 0.197 | 0.201 | 0.602 | 0.799 |
| | 20-30 | 914351 | 23.869 | 1535 | 29.423 | 1.233 | 0.245 | 0.200 | 0.554 | 0.800 |
| | 30-45 | 712500 | 18.599 | 1053 | 20.184 | 1.085 | 0.230 | 0.200 | 0.570 | 0.800 |
| | >45 | 266897 | 6.967 | 317 | 6.076 | 0.872 | 0.210 | 0.199 | 0.591 | 0.801 |
| MBI | <-0.5 | | | | | | | | | |
| | -0.5-0 | 1375135 | 35.897 | 2090 | 40.061 | 1.116 | 0.188 | 0.202 | 0.610 | 0.798 |
| | 0-0.5 | 667723 | 17.431 | 749 | 14.357 | 0.824 | 0.177 | 0.200 | 0.624 | 0.800 |
| | >0.5 | 377022 | 9.842 | 388 | 7.437 | 0.756 | 0.177 | 0.199 | 0.624 | 0.801 |
| SPI | 0-500 | 1410872 | 36.830 | 1990 | 38.145 | 1.036 | 0.172 | 0.202 | 0.626 | 0.798 |
| | 500-1000 | | | | | | | | | |
| | 1000-2000 | 1828326 | 47.728 | 2180 | 41.786 | 0.876 | 0.120 | 0.204 | 0.675 | 0.796 |
| | 2000-5000 | 573754 | 14.978 | 836 | 16.025 | 1.070 | 0.237 | 0.199 | 0.564 | 0.801 |
| | >5000 | 536571 | 14.007 | 821 | 15.737 | 1.124 | 0.251 | 0.199 | 0.549 | 0.801 |
| TRI | 0-2 | 473570 | 12.362 | 755 | 14.472 | 1.171 | 0.267 | 0.199 | 0.534 | 0.801 |
| | 2-4 | 418531 | 10.926 | 625 | 11.980 | 1.097 | 0.254 | 0.199 | 0.547 | 0.801 |
| | 4-6 | | | | | | | | | |
| | 6-12 | 472231 | 12.327 | 283 | 5.425 | 0.440 | 0.100 | 0.199 | 0.701 | 0.801 |
| | >12 | 634860 | 16.573 | 721 | 13.820 | 0.834 | 0.181 | 0.200 | 0.620 | 0.800 |
| TWI | <3 | 794959 | 20.752 | 1248 | 23.922 | 1.153 | 0.238 | 0.200 | 0.562 | 0.800 |
| | 3-6 | 1409737 | 36.801 | 2321 | 44.489 | 1.209 | 0.201 | 0.202 | 0.597 | 0.798 |
| | 6-9 | 518965 | 13.547 | 644 | 12.344 | 0.911 | 0.205 | 0.199 | 0.596 | 0.801 |

| Fac-tors | Class | N(Cij) | Ratio (%) | N(CijD) | Ratio (%) | Freq.ration (FR) | BelCij | DisCij | UncCij | PlsCij |
|---|---|---|---|---|---|---|---|---|---|---|
| Rainfall | 9-12 | | | | | | | | | |
| | >12 | 91008 | 2.376 | 130 | 2.492 | 1.049 | 0.265 | 0.198 | 0.537 | 0.802 |
| | <600 | 2006834 | 52.387 | 2854 | 54.706 | 1.044 | 0.131 | 0.206 | 0.663 | 0.794 |
| | 600-900 | 923349 | 24.104 | 1366 | 26.184 | 1.086 | 0.215 | 0.200 | 0.584 | 0.800 |
| | 900-1200 | 421327 | 10.999 | 489 | 9.373 | 0.852 | 0.197 | 0.199 | 0.604 | 0.801 |
| | 1200-1500 | 388234 | 10.135 | 378 | 7.246 | 0.715 | 0.166 | 0.199 | 0.634 | 0.801 |
| | >1500 | | | | | | | | | |
| NDVI | <0.1 | 55309 | 1.444 | 33 | 0.633 | 0.438 | 0.112 | 0.199 | 0.690 | 0.801 |
| | 0.1-0.2 | 629432 | 16.431 | 1485 | 28.465 | 1.732 | 0.379 | 0.199 | 0.422 | 0.801 |
| | 0.2-0.3 | 1213660 | 31.682 | 2118 | 40.598 | 1.281 | 0.230 | 0.201 | 0.569 | 0.799 |
| | 0.3-0.4 | 1023042 | 26.706 | 829 | 15.890 | 0.595 | 0.113 | 0.201 | 0.686 | 0.799 |
| | >0.4 | 909309 | 23.737 | 752 | 14.414 | 0.607 | 0.120 | 0.200 | 0.679 | 0.800 |

Table 2 presents the calculated Evidential Belief Function (EBF) and Frequency Ratio (FR) values for different environmental factors influencing landslide susceptibility. The results highlight significant relationships between topographic, hydrological, and vegetation-related parameters and landslide occurrence patterns.

b) Slope influence on landslide susceptibility

Landslide occurrences are predominantly concentrated in areas with slopes ranging from 10-45 degrees, with the 20-30° category showing the highest belief function (Bel = 0.245) and a relatively strong FR value (1.233). Moderate slopes tend to accumulate soil and weathered material, forming a transitional zone between bedrock and the surface This zone is particularly prone to failure, especially during heavy rainfall when increased pore water pressure reduces soil strength. In contrast, slopes steeper than 45° demonstrate lower susceptibility (Bel = 0.210, FR = 0.872). Extremely steep slopes are typically characterized by exposed rock formations rather than thick soil deposits. Since landslides predominantly occur in areas with substantial soil cover, regions with minimal or absent soil layers are less prone to landslides [32].

c) Mass Balance Index (MBI) and landslide susceptibility

The MBI values suggest that areas with negative mass balance (MBI < 0) are more prone to landslides, with the -0.5 to 0 category displaying the highest FR (1.116) and belief function (Bel = 0.188). This indicates that regions experiencing net material loss due to erosion and weathering are more susceptible to failure [25]. Conversely, areas with a positive mass balance (MBI > 0.5) show the lowest susceptibility (Bel = 0.177, FR = 0.756), implying greater slope stability due to material accumulation.

d) Stream Power Index (SPI) and Terrain Ruggedness Index (TRI)

SPI values indicate that high-energy flow regions (SPI > 5000) exhibit the highest susceptibility (Bel = 0.251, FR = 1.124), suggesting that intense surface runoff contributes significantly to slope instability. Similarly, terrain ruggedness (TRI) follows a comparable trend, with moderate ruggedness (TRI = 2-4) exhibiting the highest landslide likelihood (Bel = 0.254, FR = 1.097). This finding aligns with previous studies, where increased surface roughness correlates with higher landslide frequency due to structural weaknesses in highly dissected terrains [24].

e) Topographic Wetness Index (TWI) and Rainfall Contributions

TWI values indicate that moderate to high wetness areas (TWI = 3-6) have the highest landslide susceptibility (Bel = 0.201, FR = 1.209), reinforcing the role of soil moisture accumulation in triggering failures. Additionally, rainfall influence is most pronounced in regions with precipitation below 600 mm (Bel = 0.131, FR = 1.044), indicating that landslides are not solely driven by extreme rainfall events but also by cumulative effects of prolonged wet conditions.

f) Normalized difference vegetation index (NDVI) and Landslide Susceptibility

NDVI values reveal an inverse relationship between vegetation density and landslide occurrence. The 0.1-0.2 NDVI range demonstrates the highest susceptibility (Bel = 0.379, FR = 1.732), suggesting that sparse vegetation cover fails to provide sufficient root reinforcement for slope stability. Conversely, areas with higher NDVI (> 0.4) exhibit lower susceptibility (Bel = 0.120, FR = 0.607), highlighting the protective role of dense vegetation in stabilizing slopes by reducing surface erosion and improving soil cohesion.

To evaluate and address potential multicollinearity among landslide predisposing factors, variance inflation factors (VIF) were computed following data preprocessing, as presented in Table 3. The results indicate that all factors exhibit VIF values below 2, suggesting that multicollinearity is not a significant issue and that the model variables maintain statistical independence.

**Table 3.** The influence and independence of contributing factors.

| No | Factor | Importance | VIF | TOF |
|---|---|---|---|---|
| 1 | Slope | 0.427 | 1.746 | 0.5729 |
| 2 | TRI | 0.450 | 1.818 | 0.5501 |
| 3 | TWI | 0.068 | 1.073 | 0.9320 |
| 4 | MBI | 0.009 | 1.009 | 0.9908 |
| 5 | SPI | 0.037 | 1.039 | 0.9627 |
| 6 | NDVI | 0.163 | 1.195 | 0.8367 |
| 7 | Rainfall | 0.068 | 1.072 | 0.9325 |

Among the evaluated factors, TRI (0.450) emerged as the most critical driver of landslide susceptibility in the study area, emphasizing the significant role of terrain ruggedness in slope instability. Its VIF (1.818) and TOF (0.5501) indicate moderate multicollinearity, suggesting a correlation with Slope (0.427, VIF: 1.746, TOF: 0.5729). These finding highlights that highly rugged terrains, rather than steep slopes alone, are a primary trigger for landslides in this region an important distinction from studies in other geographical contexts where slope inclination is often the dominant factor.

Another noteworthy result is the influence of NDVI (0.163, VIF: 1.195, TOF: 0.8367), which underscores the stabilizing effect of vegetation cover. Unlike in many mountainous regions where hydrological factors outweigh vegetation influence, this study reveals that areas with reduced vegetation density are significantly more prone to landslides, indicating that deforestation or land-use changes could exacerbate slope failures.

While Rainfall (0.068, VIF: 1.072, TOF: 0.9325) and TWI (0.068, VIF: 1.073, TOF: 0.9320) are traditionally key contributors to landslide susceptibility, their relatively lower importance in this study suggests that rainfall-induced slope failures are less dominant compared to terrain-driven instabilities. This deviation from global patterns highlights the necessity of localized models for effective landslide prediction.

Conversely, MBI (0.009, VIF: 1.009, TOF: 0.9908) and SPI (0.037, VIF: 1.039, TOF: 0.9627) were the least influential factors, indicating that mass balance dynamics and stream power play a minimal role in triggering landslides in this specific landscape. This suggests that landslides in the region are not significantly influenced by erosion or sediment transport but are more structurally driven by terrain configuration.

These findings challenge conventional assumptions about landslide susceptibility by revealing that terrain ruggedness surpasses slope steepness as the dominant factor, while vegetation loss emerges as a key destabilizing element. The results underscore the importance of integrating localized environmental characteristics into predictive modeling, ensuring that landslide susceptibility assessments align with the unique geomorphological and ecological conditions of the study area.

g) Measurement of performance and comparison of susceptibility models

The performance evaluation of landslide susceptibility models reveals significant differences in predictive capability between the CF and EBF models, as measured by the area under the receiver operating characteristic curve (AUC) (Figure 3).

The CF model achieves an AUC of 0.670, which falls within the 0.6-0.7 range, classifying it as "poor" according to standard accuracy benchmarks. In contrast, the EBF model attains a higher AUC of 0.745, positioning it within the 0.7-0.8 range, indicating "fair to good" predictive performance. These results suggest that the EBF model demonstrates superior classification capability compared to the CF model in identifying landslide-prone areas.

Beyond AUC values, an analysis of the ROC curves reveals additional insights. Both models exhibit relatively wide confidence intervals, particularly in the lower region of the ROC curve, indicating some degree of uncertainty in classification at specific thresholds. However, the EBF model shows greater stability at higher true positive rates (TPR), suggesting that it consistently performs better in detecting actual landslide occurrences. This highlights the robustness of the EBF model and it was employed to develop a detailed landslide hazard zoning map for the study area.

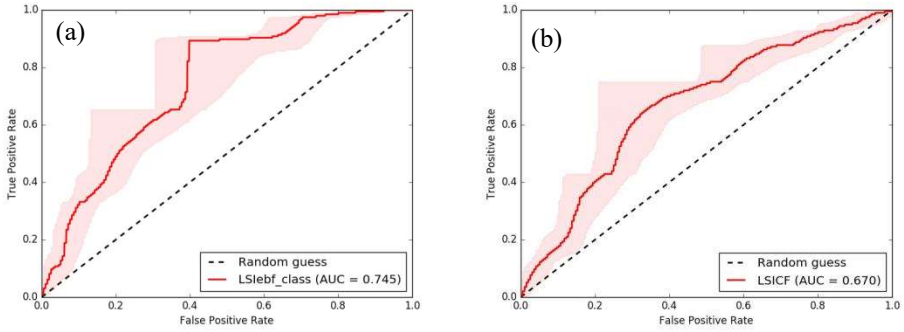


**Figure 3.** ROC curves for validation testing: (a) AUC of the landslide susceptibility map generated using the EBF model; (b) AUC of the landslide susceptibility map derived from the FR model.

Based on Table 4, Figure 4, landslide-prone areas are classified into different risk levels as follows: Very high landslide risk areas cover an area of 91.1 km², accounting for 11,23% of the total study area. This region has a high density of landslides, with 65 recorded landslides; High landslide risk areas span 210.3 km², making up 25,93% of the total study area. This category has the largest proportion of land and the highest number of recorded landslides, totaling 63 events; Moderate landslide risk areas occupy 284.6 km², representing 35,73% of the study area. In this region, 63 landslides have been recorded; Low landslide risk areas cover 140.8 km², comprising 17,36% of the study area. The number of recorded landslides in this category is 4.

The areas with the highest landslide risk are primarily concentrated in the communes of Noi Thon va Phu Ngoc. Additionally, some communes in the communes of Truong Ha, Vi Quang, are also highly susceptible to landslides (Figure 5). These areas require special attention in disaster risk management and infrastructure development planning.

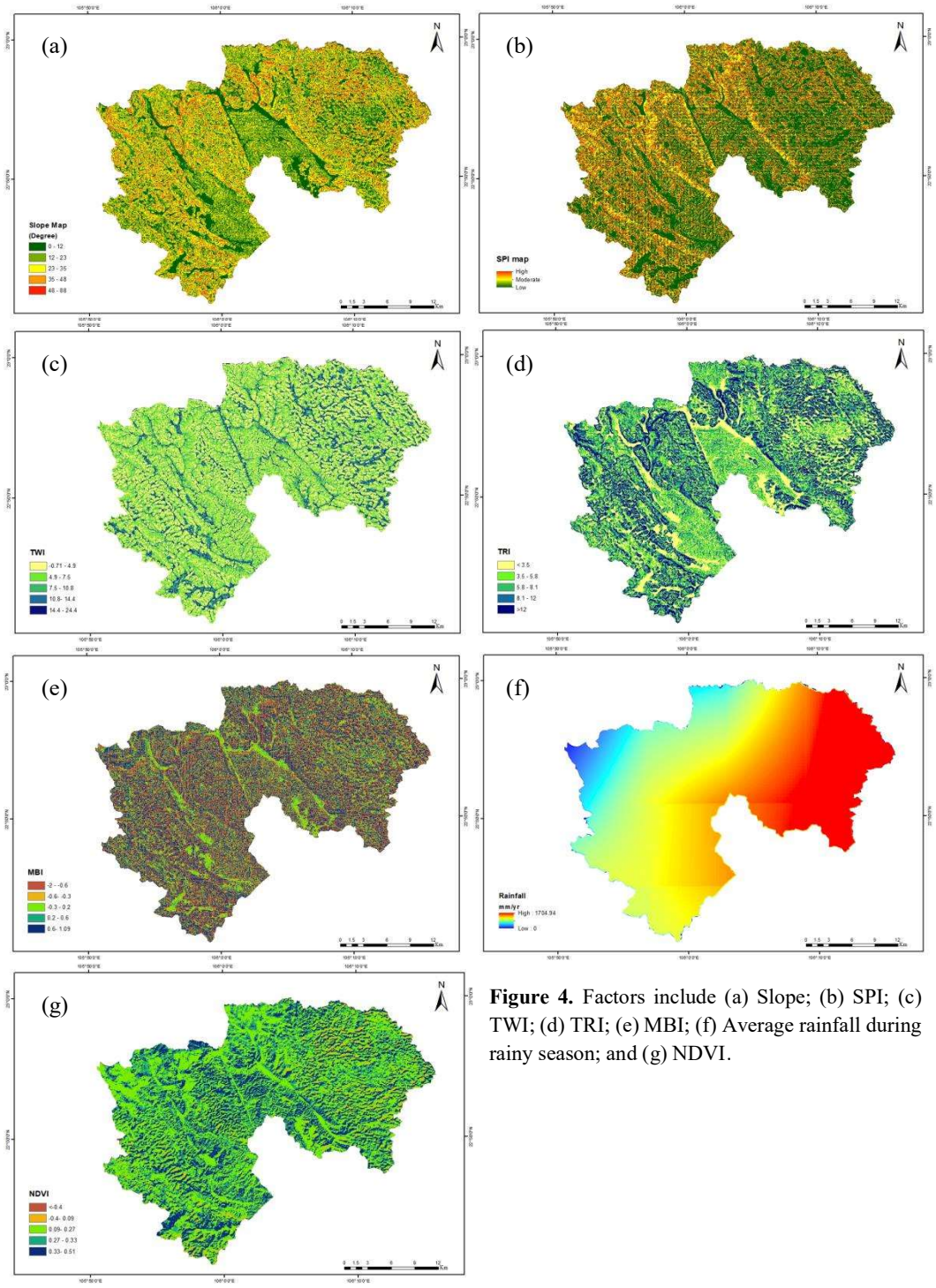


**Figure 4.** Factors include (a) Slope; (b) SPI; (c) TWI; (d) TRI; (e) MBI; (f) Average rainfall during rainy season; and (g) NDVI.

**Table 4.** The results of the reliability analysis of the methods.

| Risk level | Area ($km^2$) | Pixel | Ratio of pixels to the total of pixels (%) | Actual landslide points | Ratio of landslides to the total number of landslides (%) |
|---|---|---|---|---|---|
| Very low | 84.2 | 210,087 | 10.38 | 1 | 2.09 |
| Low | 140.8 | 351,258 | 17.36 | 4 | 4.91 |
| Moderate | 284.6 | 709,869 | 35.09 | 63 | 35.73 |
| High | 210.3 | 524,657 | 25.93 | 63 | 31.12 |
| Very High | 91.2 | 227,145 | 11.23 | 65 | 26.15 |
| **Total** | **810.96** | **2,023,016** | **100** | **196** | **100** |

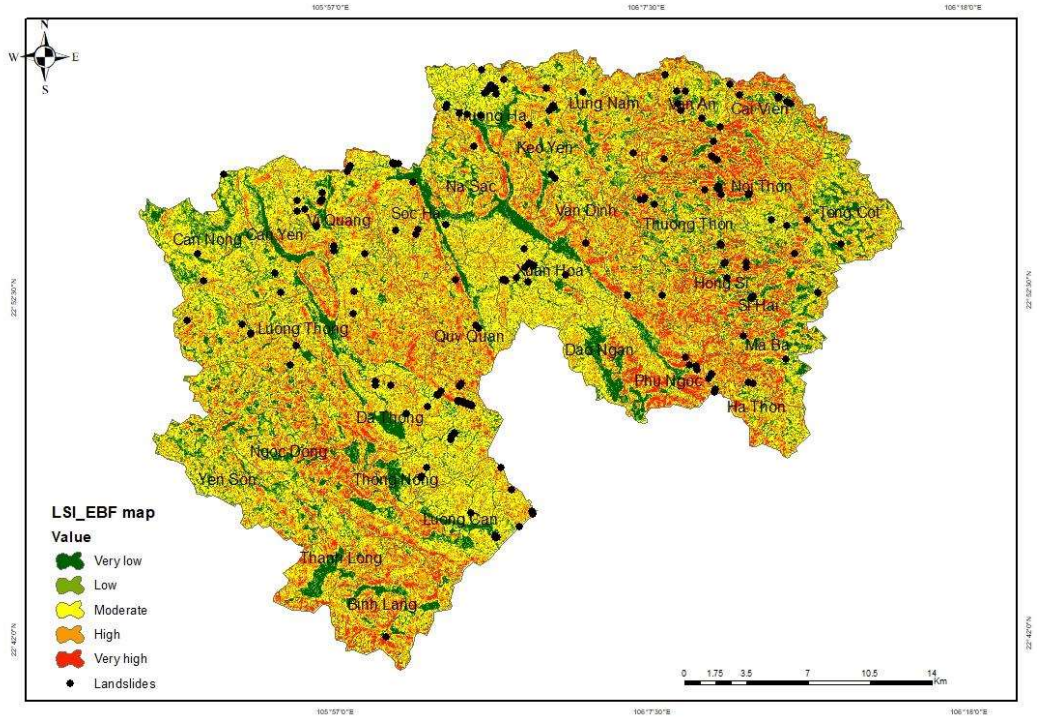


**Figure 4.** The landslide susceptibility map was developed using the DS model.

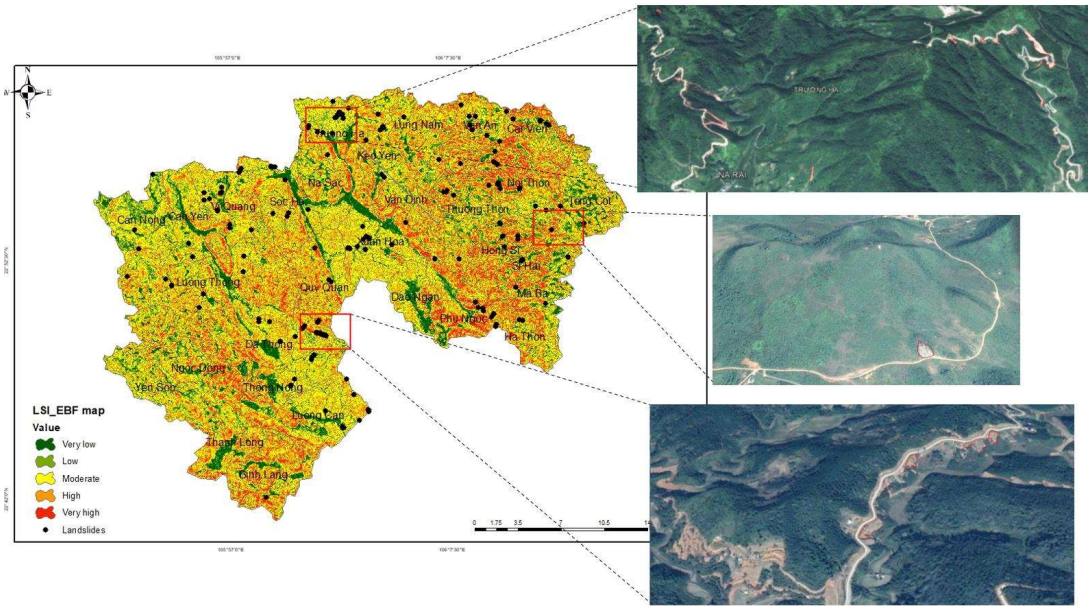


**Figure 5.** The landslide susceptibility map using the EBF model and several landslide blocks identified in Google Earth imagery.

## 4. Conclusion

In conclusion, this study identifies seven key factors contributing to landslides in the study area: slope, SPI (Stream Power Index), TWI (Topographic Wetness Index), TRI (Terrain Ruggedness Index), MBI (Mass Balance Index), rainfall, and NDVI (Normalized Difference Vegetation Index). The ROC method was applied to evaluate and validate the performance of the obtained results, indicating success rates of 74.5% for EBF and 67% for FR. The results

demonstrate the feasibility, strong predictive capability, and effectiveness of the Dempster-Shafer modeling landslide susceptibility mapping in Ha Quang with high landslide occurrence rates observed in the eastern and parts of the southwestern regions.

The results underscore the complex interplay between topography, hydrology, and vegetation cover in landslide susceptibility. Moderate slopes (20-30°), negative mass balance, high surface runoff (SPI > 5000), and sparse vegetation (NDVI 0.1-0.2) emerge as the most critical factors influencing landslide occurrence. These findings emphasize the need for integrated hazard assessment models that incorporate multiple environmental parameters to enhance landslide prediction accuracy and support risk mitigation strategies. The relationship between slope gradient and landslide susceptibility in study area is not strictly linear; instead, it follows a parabolic pattern. Therefore, landslide susceptibility models must distinguish between soil-driven landslides (common in moderate slopes) and rockfalls or structural failures (more typical on very steep terrain). And the TRI index emerges as a crucial factor in enhancing the accuracy of landslide risk prediction, offering valuable insights into terrain instability.

This study acknowledges certain limitations, primarily related to data constraints and the need for a broader model comparison. While the analysis provides valuable insights into landslide susceptibility, its predictive capability remains limited due to the exclusion of key socio-environmental factors such as land use patterns, population density, and human activities. Additionally, the model's performance could be further validated by applying it to different case studies with diverse geological and climatic conditions.

To enhance the reliability and applicability of the proposed approach, future research should focus on expanding the dataset and integrating a wider range of influencing factors. Incorporating machine learning techniques and high-resolution remote sensing data may improve predictive accuracy. Furthermore, testing the model in various geographical regions with complex topographies would help develop a more robust and transferable framework for landslide risk assessment.

**Author contribution statement:** Designed, conducted the research: H.M.T.P., H.H.T.; analyzed: H.M.T.P.; interpreted the data: H.M.T.P., H.H.T.; wrote and edited the draft manuscript: H.H.T., H.M.T.P., P.Q.N., H.T.T.P.

**Acknowledgements:** This research was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2024.TS.109. We would like to thank the anonymous reviewers for their valuable comments, which significantly enhanced the quality of this publication.

**Competing interest statement:** The authors declare no conflict of interest.

**References**

1. Kumar, P. Social and economic impact in the landslide prone zones and related policies. Landslides in the Himalayan Region: Risk Assessment and Mitigation Strategy for Sustainable Management. Springer Nature Singapore, 2024, pp. 499-529.
2. Tangestani, M.H. A comparative study of dempster-shafer and fuzzy models for landslide susceptibility mapping using a GIS: an experience from Zagros Mountains, SW Iran. *J. Asian Earth Sci.* **2009**, *35(1)*, 66–73.
3. Sana, E.; Kumar, A.; Robson, E.; Prasanna, R.; Kala, U.; Toll, D.G. Preliminary assessment of series of landslides and related damage by heavy rainfall in Himachal Pradesh, India, during July 2023. *Landslides* **2024**, *21*, 919–931.
4. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85.

5. Zhang, T.; Wang, D.; Lu, Y. Machine Learning-Enabled Regional Multi-Hazards Risk Assessment Considering Social Vulnerability. *Sci. Rep.* **2023**, *13(1)*, 13405.
6. Moayedi, H.; Mehrabi, M.; Mosallanezhad, M.; Rashid, A.S.A.; Pradhan, B. Modification of landslide susceptibility mapping using optimized PSO-ANN technique. *Eng. Comput.* **2019**, *35*, 967–984.
7. Teng, F.; Mao, Y.; Li, Y.; Qian, S.; Nanehkaran, Y. A. Comparative models of support-vector machine, multilayer perceptron, and decision tree predication approaches for landslide susceptibility analysis. *Open Geosci.* **2024**, *16(1)*, 20220642.
8. Lamichhane, S.; Kansakar, A.R.; Devkota, N.; Dahal, B.K. Integrating dynamic factors for predicting future landslide susceptibility. *Environ. Earth Sci.* **2025**, *84(3)*, 89.
9. Agboola, G.; Beni, L.H.; Elbayoumi, T.; Thompson, G. Optimizing landslide susceptibility mapping using machine learning and geospatial techniques. *Ecol. Inf.* **2024**, *81*, 102583.
10. Qiu, H.; Xu, Y.; Tang, B.; Su, L.; Li, Y.; Yang, D.; Ullah, M. Interpretable landslide susceptibility evaluation based on model optimization. *Land* **2024**, *13*(5), 639.
11. Sentz, K.; Ferson, S. Combination of evidence in Dempster-Shafer theory, Citeseer 2002, pp. 1-96.
12. Yang, Q.; Yunjun, Z.; Wang, Y.; Aoki, Y.; Wang, R. Monitoring landslides along the Jinsha river basin based on dempster-shafer evidence theory with multi-source INSAR time series. IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, 2024.
13. Park, N.W. Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis. *Environ. Earth Sci.* **2011**, *62*, 367–376.
14. Pourghasemi, H.; Pradhan, B.; Gokceoglu, C.; Moezzi, K.D. A comparative assessment of prediction capabilities of Dempster–Shafer and weights-of-evidence models in landslide susceptibility mapping using GIS. *Geomatics Nat. Hazards Risk* **2013**, *4(2)*, 93–118.
15. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32(4),* 367–385.
16. Gudiyangada Nachappa, T.; Tavakkoli Piralilou, S.; Ghorbanzadeh, O.; Shahabi, H.; Blaschke, T. Landslide susceptibility mapping for Austria using geons and optimization with the Dempster-Shafer theory. *Appl. Sci.* **2019**, *9(24)*, 5393.
17. Ishola, K.S.; Bakare, M.O.; Hamid-Mosaku, A.I.; Okolie, C.J.; Olagunju, K.T.; Oshikoya, O.M. Fusion of GIS, remote sensing, geophysics and Dempster Shafer theory of evidence for mapping groundwater prospectivity: A case study of the central parts of Lagos State, Nigeria. *Solid Earth Sci.* **2024**, *9(3)*, 100196.
18. General Statistics Office. Online available: https://www.gso.gov.vn/.
19. USGS. Online available: https://earthexplorer.usgs.gov/
20. Pradhan, A.M.S.; Kim, Y.T. Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea. *Bull. Eng. Geol. Environ.* **2017**, *76*, 1263–1279.
21. Guo, W.; Ye, J.; Liu, C.; Lv, Y.; Zeng, Q.; Huang, X. An approach for predicting landslide susceptibility and evaluating predisposing factors. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *135*, 104217.
22. Wang, G.; Chen, X.; Chen, W. Spatial prediction of landslide susceptibility based on GIS and discriminant functions. *ISPRS Int. J. Geo-Inf.* **2020**, *9(3)*, 144.

23. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Processes* **1991**, *5(1)*, 3-30.
24. Riley, S.; Degloria, S.; Elliot, S.D. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *Int. J. Sci.* **1999**, *5*, 23–27.
25. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24(1)*, 43–69.
26. Möller, M.; Volk, M.; Friedrich, K.; Lymburner, L. Placing soil-genesis and transport processes into a landscape context: a multiscale terrain-analysis approach. *J. Plant Nutr. Soil Sci.* **2008**, *171(3)*, 419–430.
27. Fei, L.; Li, T.; Ding, W. Dempster–Shafer theory-based information fusion for natural disaster emergency management: A systematic literature review. *Inf. Fusion* **2024**, *112*, 102585.
28. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **1967**, *38(2)*, 325–339.
29. Chung, C.J.F.; Fabbri, A.G. The representation of geoscience information for data integration. *Nonrenewable Resour.* **1993**, *2*, 122–139.
30. Oke, J.; Akinkunmi, W.B.; Etebefia, S.O. Use of correlation, tolerance, and variance inflation factor for multicollinearity test. *Global Sci. J.* **2019**, *7(5)*, 652–659.
31. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365.
32. Qiu, H.; Cui, Y.; Hu, S.; Yang, D.; Pei, Y.; Ma, S.; Liu, Z. Size distribution and size of loess slides in response to slope height and slope gradient based on field survey data. *Geomatics Nat. Hazards Risk* **2019**, *10(1)*, 1443–1458.