Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Landslide Susceptibility Mapping under Extreme Events: Evidence from the October 2020 Event in Phuoc Son area, Danang city, Vietnam

Viet Long Doan[1*], Chi Cong Nguyen[1], Binh Quang Nguyen[1], Ly Trieu Pham[1], Trung Quan Nguyen[1], Le Huyen Trang[2]
[1]The University of Danang, University of Science and Technology, Danang, Vietnam; dvlong@dut.udn.vn
[2]Geotechnical and Artificial Intelligence research group, University of Transport Technology, Hanoi 100000, Vietnam; lehuyentrang0500@gmail.com

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.2.256-276

***Corresponding author:**
Email address:
dvlong@dut.udn.vn

Received: 24/01/2026
Received in Revised Form: 15/04/2026
Accepted: 25/04/2026

Abstract: Landslides in mountainous regions of Vietnam predominantly occur during the rainy season, with high-intensity rainfall events frequently triggering widespread shallow landslides. On 28 October 2020, an extreme rainfall event, with daily precipitation reaching approximately 350–400 mm, caused numerous landslides in the mountainous Phuoc Son area, Da Nang City. Based on a landslide inventory compiled for this event and a comprehensive set of conditioning factors, this study investigates the influence of these factors on landslide occurrence. In addition, two machine learning algorithms, Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), were employed for landslide spatial prediction as well as landslide susceptibility mapping. Frequency analysis and Frequency Ratio (FR) results indicate that 12-hour accumulated rainfall played a dominant role in triggering this landslide event. Model evaluation shows that both models achieved good predictive performance; however, XGBoost outperformed RF, with the value of AUC up to 0.905. Based on the optimal XGBoost model, landslide susceptibility maps were generated under two scenarios: (1) using the spatial distribution of 12-hour accumulated rainfall and (2) applying a rainfall-triggering threshold of 320 mm/12h across the entire study area. The results indicate that, under Scenario 2, zones of high and very high landslide susceptibility expand markedly throughout the region, providing valuable information for disaster hazard prevention and land-use planning in mountainous areas.

Keywords: cumulative rainfall, XGBoost, Random Forest, landslide susceptibility, rainfall threshold.

1. Introduction

In landslide-related studies, landslide susceptibility maps are considered effective tools for mitigating this particularly dangerous type of natural disaster [1]. In recent years, statistical approaches, including machine learning and deep learning methods, have been recognized as highly effective in predicting landslides. Under statistical-based approaches, these maps are constructed based on historical landslide inventories and a set

of influencing factors [2]. Landslide inventory data in the study area are typically collected over multiple years to ensure a sufficiently large dataset, including information such as location, time of occurrence, landslide type, and landslide magnitude [3]. However, due to limitations in data availability, most studies only provide information on the locations where landslides occurred [4], [5], [6]. Information on the timing of landslide occurrence is addressed in only a limited number of studies on landslide hazard assessment [7], [8]. Regarding the influencing factor data, both natural factors (e.g., topography, geomorphology, geology, hydrology, and land cover) and factors related to human activities (e.g., road construction and deforestation) are collected and evaluated. The study by Reichenbach indicated that as many as 596 influencing variables have been employed in studies assessing landslide susceptibility [1]. Among these variables, rainfall is considered a triggering factor for landslides in numerous study areas worldwide.

In rainfall-induced landslide susceptibility assessment, many study areas lack information on the timing of landslide occurrences. In addition, landslide inventory data are often collected over long time periods spanning multiple years. These limitations make it challenging to select appropriate rainfall data for landslide susceptibility evaluation. This is the main reason why many studies employ average annual precipitation data in landslide spatial prediction [1]. In addition, several studies have employed alternative types of rainfall data in landslide susceptibility assessment. For example, the study by Bui used eight-day maximum rainfall to assess landslide spatial prediction in the mountainous areas of Hoa Binh Province [9]. The study conducted by Su employed two types of precipitation, namely daily rainfall and accumulated rainfall, to assess landslide susceptibility in the Wencheng area, China [10]. The study by Zhang used accumulated rainfall data over periods ranging from one to five days to evaluate landslide susceptibility in Shenzhen, China [11]. The study conducted by Doan employed 1-day to 7-day maximum rainfall data for landslide spatial prediction in Quang Ngai Province and identified the 3-day maximum rainfall as the most appropriate rainfall indicator [12]. However, a common characteristic of these studies is the use of landslide inventory data collected over multiple years while relying on a specific representative rainfall variable. Consequently, these studies have not provided sufficiently convincing justification for the selection of rainfall data in landslide susceptibility assessment. A more reasonable approach may involve using landslide data from a single event together with the corresponding rainfall data for landslide susceptibility mapping. Nevertheless, this approach requires a sufficiently large landslide inventory for machine learning or deep learning models to perform effectively.

It can be seen that significant issues remain regarding the relationship between landslide inventory data and rainfall data in landslide susceptibility assessment for a given area. To address this problem, this study focuses on a large-scale and widespread landslide event that occurred in the mountainous Phuoc Son area of Da Nang City on 28 October 2020 for evaluation. This approach addresses two key issues: (i) the number of landslide sites is sufficiently large to enable effective application of machine learning models; and (ii) the landslide event recorded on 28 October 2020 provides a basis for evaluating and selecting appropriate rainfall data responsible for triggering landslides. Other influencing factors related to geology, topography, hydrology, land use/land cover (LULC), and human activities are also incorporated to comprehensively assess this landslide event. Machine learning algorithms, including Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) [13], [14], [15], [16], are employed to develop landslide spatial prediction models. The predictive performance of the models is evaluated using statistical indices

and the Receiver Operating Characteristic (ROC) method. The best-performing model is then used to generate landslide susceptibility maps under an extreme rainfall scenario applied across the entire study area. These maps provide valuable information on landslide susceptibility levels in the study area under extreme rainfall conditions (such as the event that occurred on 28 October 2020) and serve as useful references for disaster risk reduction and land-use planning.

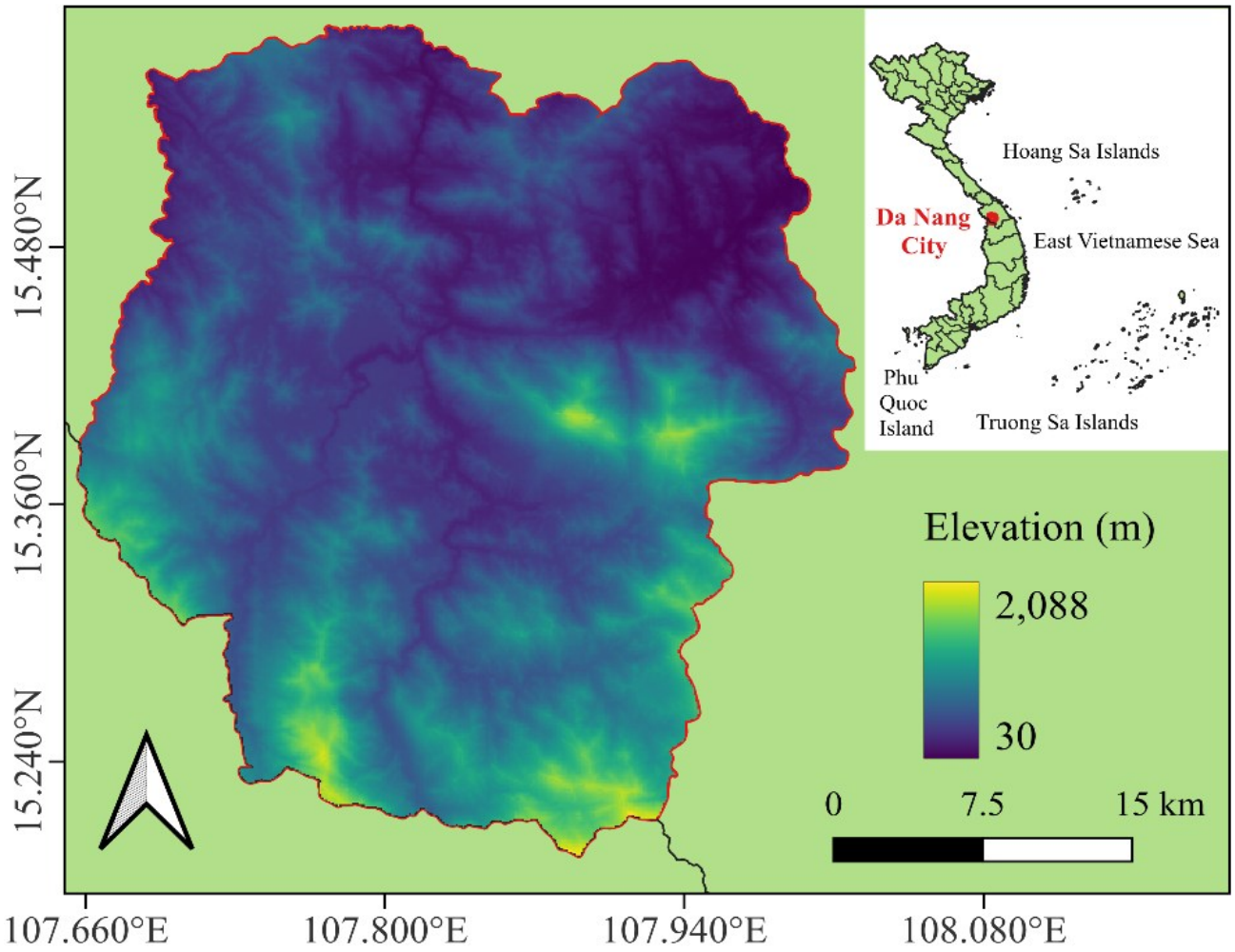


Fig. 1. Location of the Phuoc Son mountainous area, Da Nang City, Vietnam

2. Study area

The study area is the Phuoc Son mountainous region, Danang city (Fig. 1), located between 15°34′50″ and 15°11′35″ N latitude and 107°39′30″ to 108°01′17″ E longitude. The total natural area of the study area is about 1,142 km². The region is predominantly covered by forest, accounting for up to 87% of the total area. The terrain is mainly mountainous and hilly, with most elevations exceeding 200 m and a maximum elevation of over 2,088 m. Slopes are generally steep, with most areas having slope angles ranging from 15° to 40°, and locally exceeding 45°, gradually decreasing in elevation from west to east. This topographic setting forms an ideal windward terrain; therefore, during the rainy season, the occurrence of monsoons or tropical depressions is often accompanied by intense rainfall, leading to frequent landslides [17]. A representative example is the extreme rainfall that occurred on 28 October 2020, which caused widespread landslides in the study region. Landslide locations were then inventoried and collected together with corresponding rainfall data for use in this study.

3. Methodology

Fig. 2 presents the overall methodological

framework adopted in this study, which comprises five main steps. First, landslide inventory data, rainfall data associated with landslide occurrence, and other relevant influencing factors were collected and preprocessed. Second, the collected datasets were analyzed and evaluated to characterize their relationships with landslide occurrence. Third, XGBoost and RF models were developed for landslide spatial prediction using 70% of the available data for model training. Fourth, the predictive performance of the models was tested using the remaining 30% of the data. Finally, landslide susceptibility maps were generated and evaluated based on the best-performing model.

3.1. Landslide inventory collection

The landslide sites for the 28 October 2020 event in the Phuoc Son area were developed using Sentinel-2 multispectral satellite imagery. Sentinel-2 data were selected due to their high spatial resolution (10 m) and free availability, which makes them well-suited for rapid and event-based landslide mapping over large mountainous areas. Pre-event (14 September 2020) and post-event (28 December 2020) Sentinel-2 images were processed to detect landslide occurrences. Normalized Difference Vegetation Index (NDVI) maps were generated using an open-source workflow implemented in Google Earth Engine (GEE). By using the Sentinel-2–based NDVI change detection combined with visual interpretation of high-resolution Google Earth images and field validation using UAV photographs [16], a comprehensive landslide inventory was constructed for the 28 October 2020 event. In total, 2,328 landslide sites were detected, corresponding to an average landslide density of approximately 2 landslides per km². The spatial distribution of the mapped landslides shows a clear clustering pattern, with most landslide occurrences clustered in the southern portion of the study area (Fig. 3). This spatial concentration indicates a strong heterogeneity in landslide occurrence within the region and provides an important basis for subsequent landslide susceptibility modelling and risk assessment.

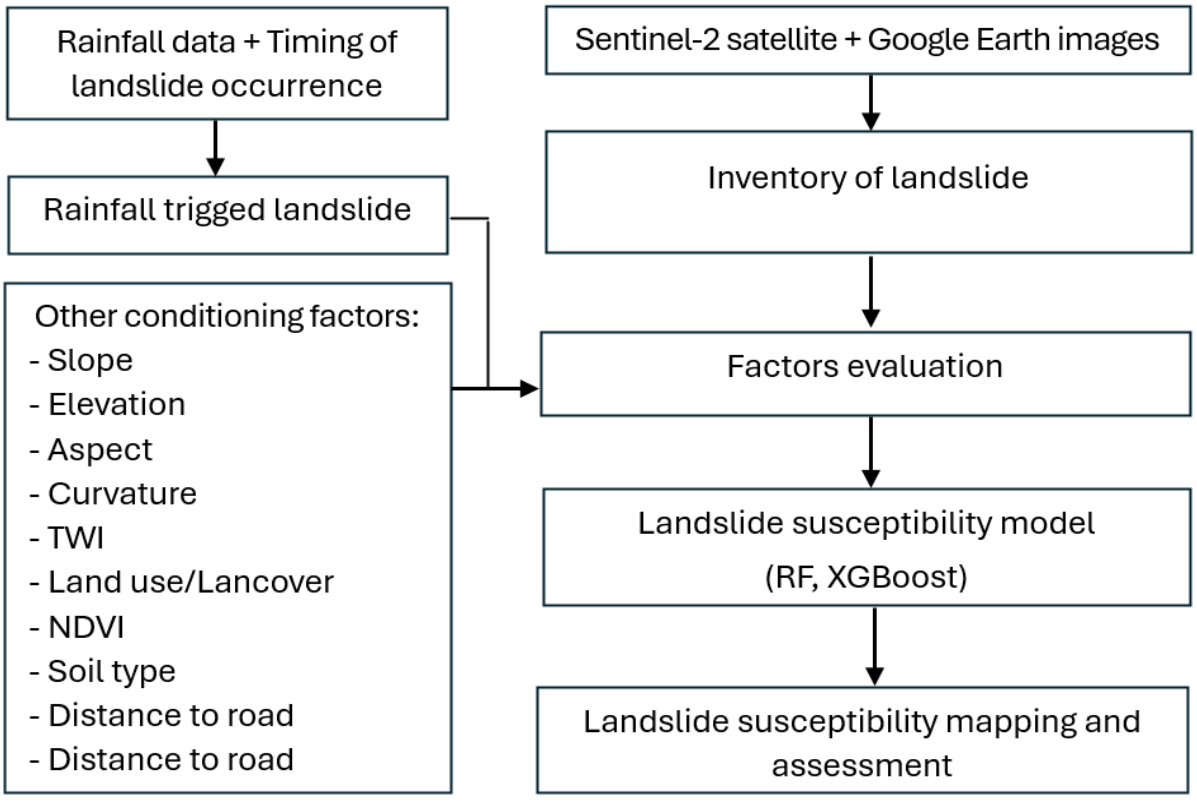


Fig. 2. Overall workflow of the methodology adopted in this study

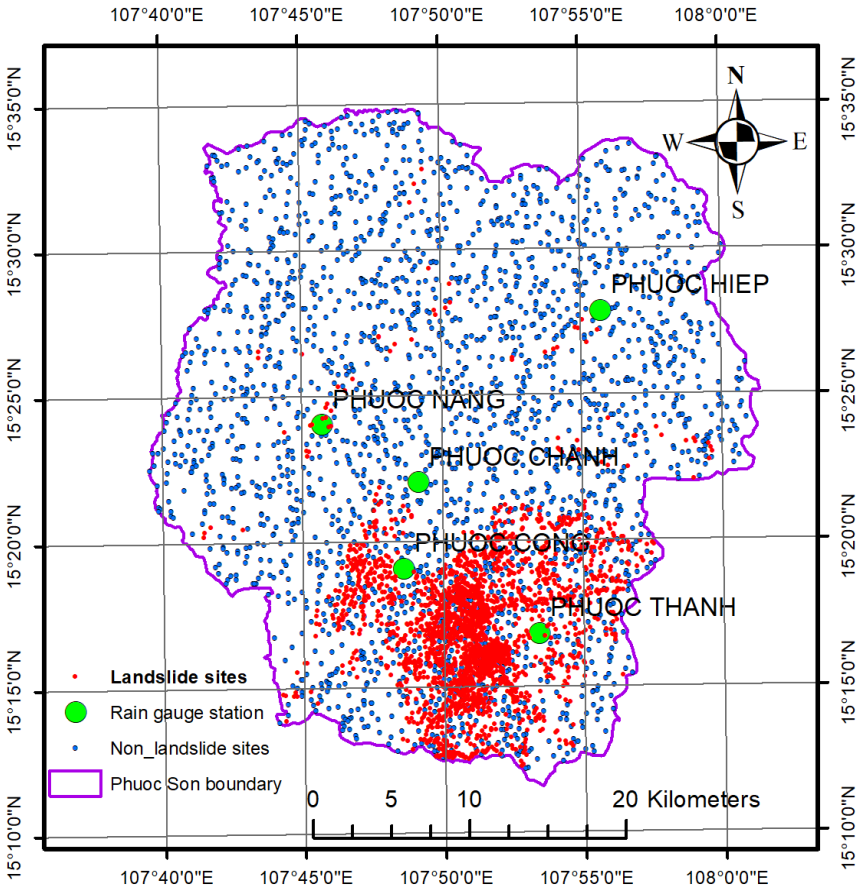


Fig. 3. Spatial distribution of rainfall monitoring stations across the study area

3.2. Landslide influencing factors mapping

a. Landslide-triggering rainfall

Hourly precipitation data for the study region were collected from five rain gauge stations (Fig. 3), namely Phuoc Thanh, Phuoc Cong, Phuoc Chanh, Phuoc Nang, and Phuoc Hiep (source: https://vrain.vn). Based on the landslide events recorded in the study area on 28 December 2020, which were mainly concentrated around the Phuoc Thanh rain gauge station (Fig. 3), this study focused on analyzing the evolution of cumulative rainfall responsible for triggering the landslide event. The rainfall time series shown in Fig. 4 provides information on daily rainfall and cumulative rainfall at the Phuoc Thanh rain gauge station from 1 October to 30 October 2020. The total cumulative rainfall during this period reached 1305.8 mm, with an average daily rainfall of 43.5 mm/day. From 22 October to 27 October 2020, rainfall was negligible; however, an extreme rainfall event occurred on 28 October 2020, with a daily rainfall amount of 343 mm. This intense rainfall event is considered the primary triggering factor for the landslides in the study area. An analysis of the hourly rainfall pattern on 28 October 2020 (Fig. 5) indicates that rainfall was concentrated within a 12-hour period, with a total accumulated rainfall of 323.6 mm. This short-duration but extremely intense rainfall event, corresponding to an average intensity of approximately 323.6 mm/12h, is considered the primary triggering factor of the landslides. Based on this observation, the present study proposes the use of 12-hour cumulative rainfall derived from five rain gauge stations in the study area (Fig. 3) as the rainfall input for landslide susceptibility assessment. The Inverse Distance Weighting (IDW) interpolation method was applied to create the spatial rainfall distribution map [18]. Subsequently, the rainfall map was classified into six classes by the Natural Breaks classification method [19] (Fig. 6).

b. Other influencing factors

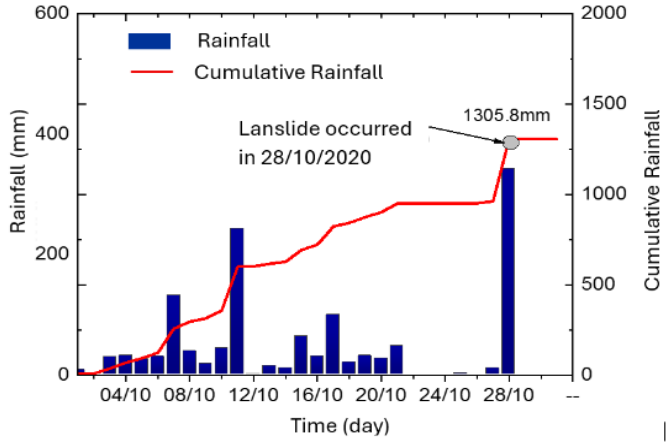


Fig. 4. Daily rainfall at Phuoc Thanh rain gauge station from 01/10-29/10/2020

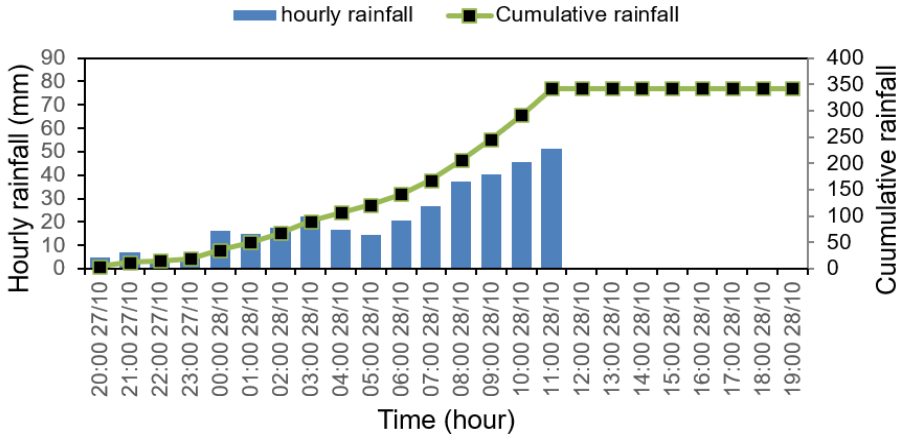


Fig. 5. Rainfall triggered a landslide at the Phuoc Thanh rain gauge station on 28 October 2020

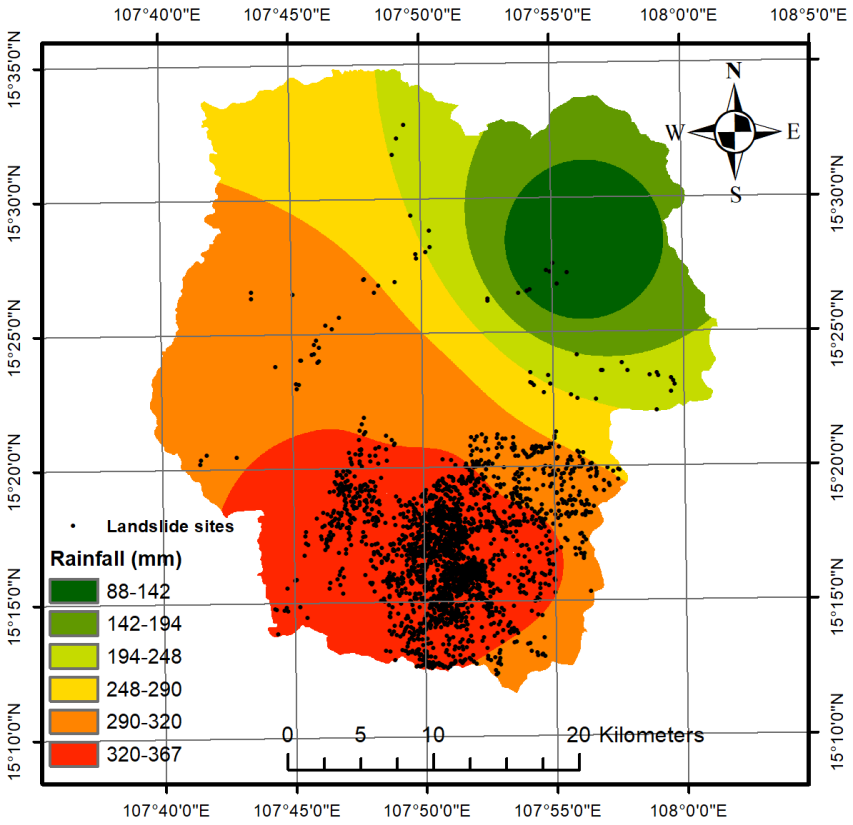


Fig. 6. 12h-cumulative rainfall map in Phuoc Son region

Table 1. Detailed information on other conditioning factors

No.	Landslide conditioning factors	Source	Scale/Resolution
1	Slope	Extracted from NASA DEM	30×30 m
2	Aspect	Extracted from NASA DEM	30×30 m
3	Elevation	Extracted from NASA DEM	30×30 m
4	TWI	Extracted from NASA DEM	30×30 m
5	Curvature	Extracted from NASA DEM	30×30 m
6	Soil type	Departments of Natural Resource and Environment of Quang Nam Province	1/100,000
7	Land use/Land cover	https://landcovermapping.org, data collected in 2020	30×30 m
8	NDVI	Sentinel 2 satellite image, data collected in August 2020	10×10 m
9	Distance to road (m)	Departments of Natural Resource and Environment of Quang Nam Province	1/25,000
10	Distance to stream	Extracted from NASA DEM	30×30 m

In addition to the precipitation triggering factor, this study collected other factors influencing landslide occurrence, including slope, aspect, elevation, curvature, topographic wetness index (TWI), LULC, soil type, distance to roads, and distance to streams. The characteristics and sources of the datasets are summarized in Table 1.

3.3. Data evaluation

a. Variance Inflation Factor method

The Variance Inflation Factor (VIF) is an effective method for assessing multicollinearity [20]. It is calculated as follows:

$$VIF = 1 / (1 - R^2) \quad (1)$$

where R^2 represents the coefficient of determination obtained by regressing a given explanatory variable against all other explanatory variables in the model. In several studies, variables with VIF values greater than 5 are considered to exhibit moderate multicollinearity, while those with VIF values exceeding 10 are typically regarded as highly collinear and should be considered for removal from the model [20].

b. Frequency Ratio method

This study employed the Frequency Ratio (FR) method to evaluate the conditioning factors. The FR values were calculated based on the relationship between historical landslide occurrences and each landslide influencing factor [21], [22]. Each variable was categorized into multiple classes, and the FR value (Frij) associated with each class was calculated according to the following equation [23]:

$$FR_{ij} = \frac{N(L \cap C_{ij})/N(L)}{N(C_{ij})/ N(C)} \quad (2)$$

where:

C_{ij} denotes the *j-th* class of the conditioning factor C_i(i = 1, 2, …, n);

$N(L \cap C_{ij})$ is the number of landslide pixels located within the class C_{ij};

$N(L)$ represents the total number of landslide pixels used for model construction;

$N(C_{ij})$ represents the total number of pixels within the *j-th* class of the *i-th* factor;

$N(C)$ denotes all of the pixels within the study region.

3.4. Landslide susceptibility model

a. Random Forest

Random Forest (RF) is a data mining algorithm developed under the bagging framework, which achieves high classification accuracy by aggregating a large number of Decision Trees

(DTs) [24]. The RF model performance is mainly influenced by two hyperparameters: ntree, which determines the number of decision trees in the ensemble, and mtry, which specifies the number of variables randomly selected at each node split. RF is particularly suitable for landslide susceptibility classification tasks, where samples are categorized as landslide and non-landslide, and it has been widely applied in previous landslide-related studies [14], [25], [26]. In this study, the landslide susceptibility index was calculated as the proportion of decision trees predicting landslide occurrence relative to the total number of trees in the RF model.

b. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a high-performance machine learning algorithm developed by Chen [27]. It is an ensemble learning method that combines multiple decision tree algorithms to construct a more robust predictive model and is derived from the Gradient Tree Boosting framework. XGBoost employs an ensemble of Classification and Regression Trees (CART), which are integrated using a gradient boosting strategy to iteratively minimize the prediction error [27]. Owing to its strong predictive capability, XGBoost has been widely applied in landslide spatial prediction and has demonstrated superior performance in previous studies [13], [14], [28]. However, XGBoost is characterized by a relatively complex structure with multiple hyperparameters, including nrounds, max_depth, eta, gamma, colsample_bytree, min_child_weight, and subsample. Therefore, for a specific study area, it is necessary to optimize these hyperparameters to obtain the most effective predictive model.

3.5. Model evaluation

Multiple quantitative and qualitative criteria were utilized to evaluate the performance of the landslide spatial prediction models. For classification-based models such as XGBoost and RF, commonly used evaluation approaches include the statistical index and the Receiver Operating Characteristic (ROC) method [29]. A set of statistical metrics, including overall accuracy (ACC), positive predictive value (PPV), negative predictive value (NPV), sensitivity (SST), specificity (SPF), and the Kappa coefficient (k), was selected for model evaluation [30], [31], [32].

For the ROC-based evaluation, the area under the ROC curve (AUC) was utilized to assess the model's performance. The AUC value ranges between 0 and 1, with values closer to 1 indicating higher predictive capability. According to the study of Kantardzic, models with AUC values between 0.9 and 1.0 are classified as “excellent”, followed by “good” (0.8–0.9), “fair” (0.7–0.8), “moderate” (0.6–0.7), and “poor” or unreliable (0.5–0.6) [33].

4. Result and Discussion

4.1. Assessment of Input Data

a. Multicollinearity assessment

Table 2. VIF Values of Input Variables

Input variables	Slope	Aspect	Elevation	TWI	Curvature	Soil type	LULC	NDVI	Distance to road	Distance to stream	Rainfall
VIF	1.43	1.02	2.48	1.40	1.68	1.81	1.11	1.10	1.18	1.13	1.36

The results of the multicollinearity assessment using the VIF index are presented in Table 2. It can be observed that all input variables have VIF values less than 5, indicating the absence of multicollinearity within the dataset. Therefore, all 11 input variables are considered suitable for inclusion in the predictive model.

b. Frequency Ratio assessment

Rainfall: The 12-hour accumulated rainfall map was classified into six categories utilizing the Natural Breaks method (Fig. 7). The frequency analysis results (Fig. 7a) indicate that the majority of landslide occurrences are concentrated in areas corresponding to the highest rainfall class (Class

6). Moreover, the Frequency Ratio (FR) value of Class 6 reaches 3.7, which is substantially higher than those of the remaining classes (FR < 0.5). This result clearly demonstrates that intense short-duration rainfall exerts a strong influence on landslides in the study region. These results highlight the effectiveness of incorporating 12-hour accumulated rainfall as a landslide-triggering factor in susceptibility assessment. This approach overcomes key limitations of previous studies conducted in the Phuoc Son mountainous area, which either did not explicitly consider rainfall as a triggering factor [6] or relied on average annual rainfall data [34]. By using event-based accumulated rainfall closely associated with the landslide occurrence, the proposed framework provides a more physically meaningful and reliable representation of rainfall–landslide interactions, thereby improving the predictive capability of landslide susceptibility mapping.

Slope: The slope map was derived from the DEM and classified into six classes, such as 0–10, 10–20, 20–30, 30–40, 40–50, and > 50 degrees (Fig. 7b). Landslides most frequently occurred in areas with slope angles ranging from 20° to 30°, whereas very few landslides were observed in areas with slopes <10° and >40°. However, the Frequency Ratio (FR) analysis shows that the slope class of 40°–50° exhibits the highest landslide density, with an FR value of 2.2.

Aspect: Similar to the slope map, the aspect map was extracted from the DEM and categorized into eight directional classes, including North-East, East, South-East, South, South-West, West, North-West (Fig. 7c). Both the frequency distribution and FR values show only minor differences among aspect classes, indicating that aspect does not exert a significant influence on the appearance of landslides in the study area.

Elevation: The elevation map was categorized into eight classes, such as <200, 200–400, 400–600, 600–800, 800–1000, 1000–1200, 1200–1400, >1400 m (Fig. 7d). The frequency analysis reveals that most landslide occurrences are concentrated at elevations between 600 and 1000 m, while very few landslides occur at elevations below 400 m and above 1200 m. The FR results also indicate a higher landslide density within the 600–1000 m elevation range.

Curvature: The curvature map was generated from the DEM and classified into five classes, including high concave (<-2), concave [(-2) – (-0.05)], flat [(-0.05) – (0.05)], convex [0.05 – 2], and high convex (>2) (Fig. 7e). Landslides in the Phuoc Son area are mainly distributed in both convex and concave terrain. However, the FR analysis shows that concave and strongly concave areas exhibit higher landslide densities than other curvature classes.

Topographic Wetness Index (TWI): The TWI map was also derived from the DEM and classified into eight classes, such as <4.7, 4.7–5.6, 5.6–6.4, 6.4–7.3, 7.3–8.4, 8.4–9.8, 9.8–11.5, >11.5. The frequency distribution (Fig. 7f) indicates that landslides tend to appear in classes with low TWI values, with decreasing frequencies as TWI increases. A comparable trend is observed for the FR values, with the first class (TWI < 4.7) exhibiting the highest FR and the lowest FR in the seventh class (TWI > 9.8).

NDVI: The NDVI map was generated from Sentinel-2 imagery acquired during the dry season of 2020 and classified into eight classes, including < 0.02, 0.02–0.18, 0.18–0.27, 0.27–0.36, and> 0.36. Fig. 7g illustrates that the largest proportion of the study area corresponds to the highest NDVI class (NDVI > 0.36), reflecting extensive vegetation cover. Accordingly, most landslide occurrences are also concentrated in these vegetated areas. The highest FR values are recorded for NDVI classes greater than 0.27; however, the FR values are relatively low and close to 1, suggesting a limited influence of NDVI on landslide occurrence.

Land Use/Land Cover (LULC): The LULC map was obtained from the website

https://landcovermapping.org in 2020, and its classification scheme includes the following categories: Agricultural land, Barren, Deciduous, Evergreen, Other Forest, Grass, Shrub, Urban and Built Up, Surface water. The frequency analysis (Fig. 7h) shows that most of the study area is covered by evergreen forest, and the majority of landslide occurrences are also located within this class. In contrast, the FR analysis indicates that the grassland class exhibits the highest landslide density, with an FR value of 4.

Soil Type: The soil type map of the study area is classified into the following groups: Acrisol, Ferasol, Gleysol, Fluvisol, Leptosols, and Water bodies. Two dominant soil types in the study area are Acrisols and Ferralsols (Fig. 7i), and most landslides occur within these soil classes. The FR results indicate that landslides are more likely to occur in areas characterized by Acrisol soils.

Distance to Stream: The distance to stream map was derived from DEM data and classified into distance-based categories relative to stream branches as follows: 0–50, 50–100, 100–150, 150–200, 200–250, >250 m. The frequency distribution (Fig. 7j) shows that most landslides occur at distances greater than 250 m from streams. The FR analysis reveals no significant differences among distance classes, indicating that distance to stream does not have a pronounced influence on landslide appearance in the study region.

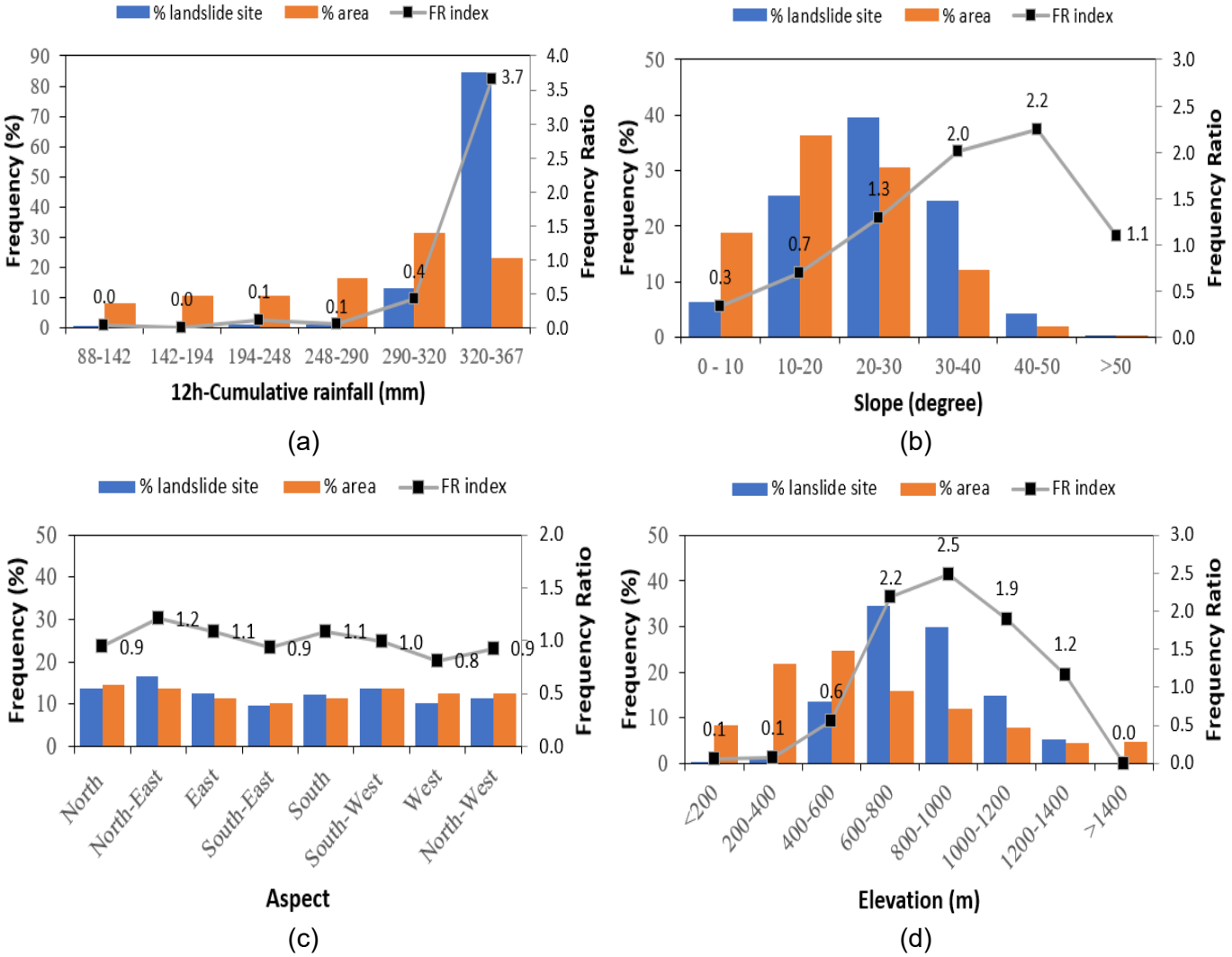
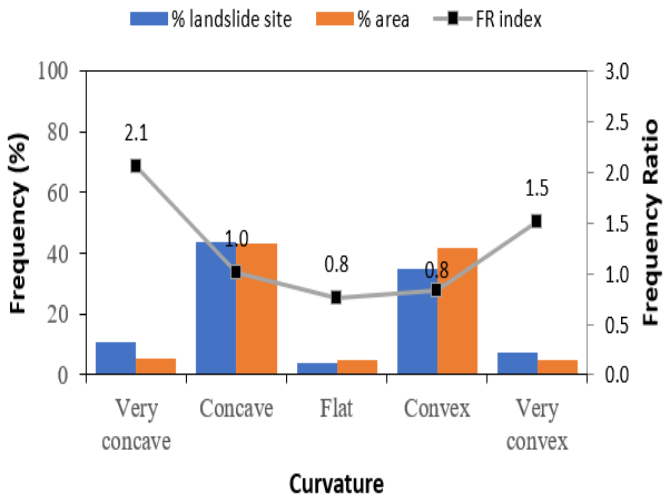
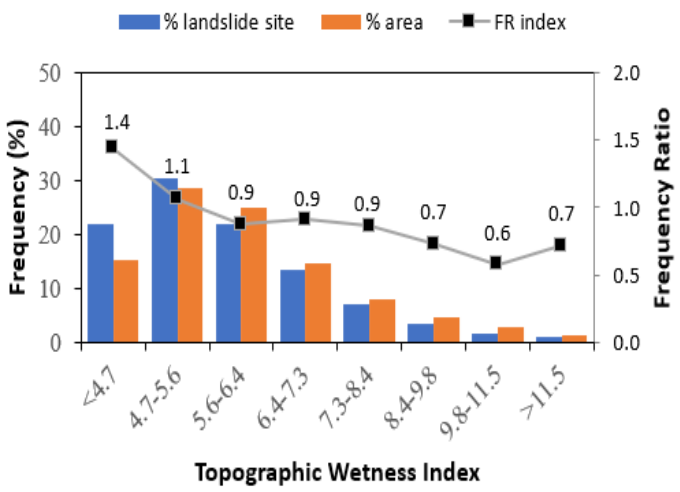


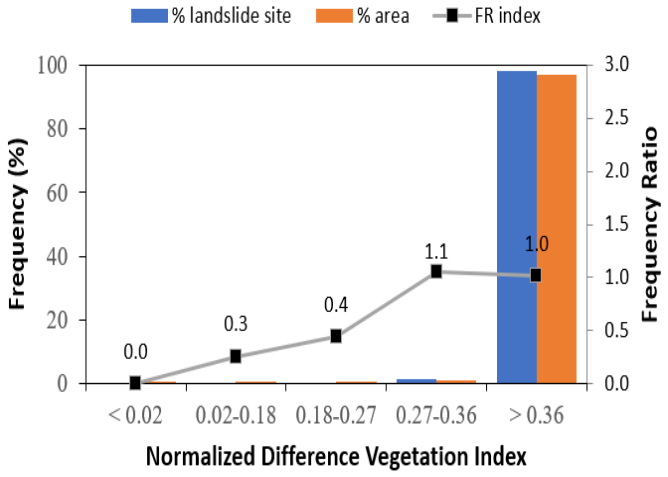
Fig. 7. The frequency distributions of the influencing factors: (a) rainfall, (b) slope, (c) aspect, (d) elevation, (e) curvature, (f) TWI, (g) NDVI, (h) LULC, (i) soil type, (j) Distance to stream, (k) Distance to road



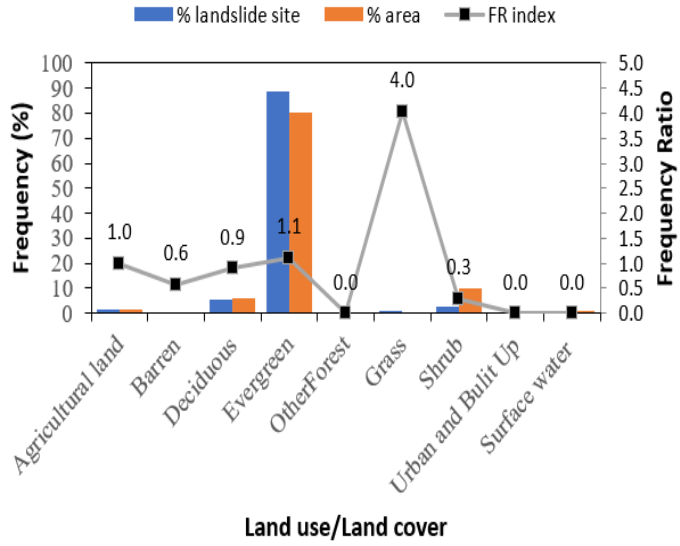
(e)



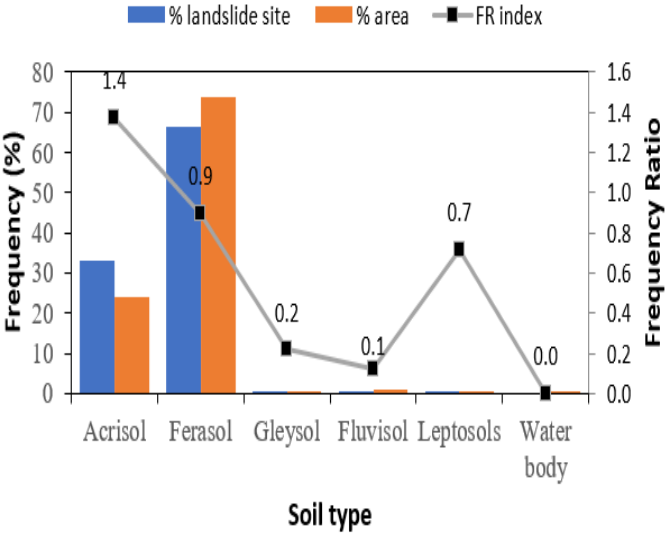
(f)



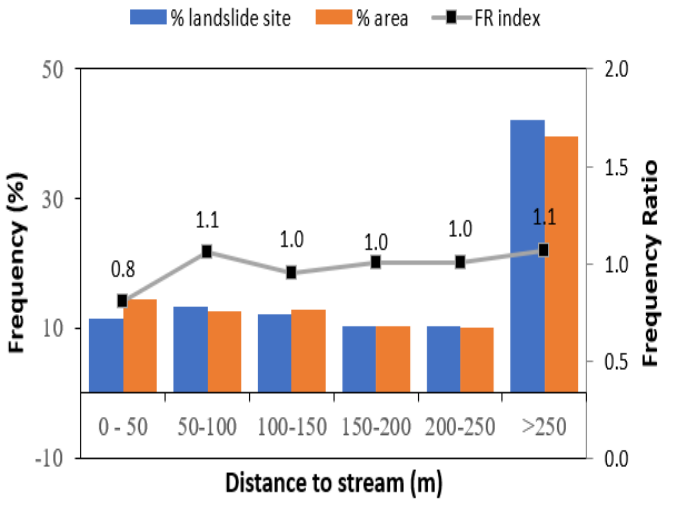
(g)



(h)

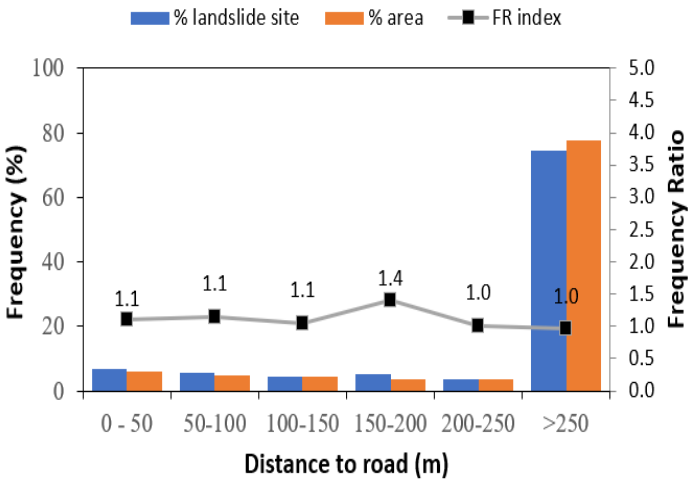


(i)



(j)

Fig. 7. (continued)



(k)

Fig. 7. (continued)

Distance to Road: The distance-to-road map provides information on the distance from a given location to transportation networks and is classified into the following categories: 0–50, 50–100, 100–150, 150–200, 200–250, >250 m. Most of the study area is located more than 250 m away from roads (Fig. 7k), reflecting a sparse road network in the region. Consequently, most landslide occurrences are also recorded at distances greater than 250 m from roads. The FR analysis shows no notable differences among distance classes, suggesting that distance to road does not significantly influence the landslide event that occurred on 28 October 2020.

Overall, the results of the Frequency Ratio analysis reveal that rainfall exerts the strongest control on landslide occurrence in the study area. Landslides are highly concentrated in areas experiencing extreme 12-hour accumulated rainfall, with the highest rainfall class exhibiting FR values significantly greater than those of other classes. In addition, topographic factors, particularly slope and elevation, also play important roles in landslide susceptibility. Steep slopes (40°–50°) and elevations ranging from 600 to 1000 m show relatively high FR values, indicating favorable conditions for slope instability. Furthermore, soil type contributes to landslide occurrence, with Acrisol soils exhibiting higher susceptibility compared to other soil classes. In contrast, factors such as aspect, distance to streams, and distance to roads show limited influence, as reflected by their relatively uniform FR values across classes. These results emphasize the dominant role of intense rainfall acting in combination with terrain and soil conditions in triggering the landslide event of 28 October 2020 in the Phuoc Son area.

c. Feature importance

This study employs the XGBoost method to evaluate the importance of influencing factors. The results, illustrated in Fig. 8, indicate that the influence of these factors on landslides is quantified using the Gain parameter. Features with higher gain values have a more significant impact on model predictions. Specifically, the 12-hour cumulative rainfall ranks first, with a markedly higher Gain value compared to the other factors, followed by elevation, slope, soil type, aspect, distance to road, distance to stream, TWI, and curvature. Vegetation-related factors, such as NDVI and LULC, show no significant influence on landslide occurrence. This can be explained by the fact that the study area is characterized by a forest cover exceeding 90% of its total area. The landslide event occurred over a wide area, predominantly in regions with high forest cover density (Fig. 7h).

4.2. Performance of landslide prediction models

The landslide susceptibility models were

developed using 70% of the available dataset for training. To mitigate overfitting, this study employs the k-fold cross-validation method during the model training process. The training dataset is partitioned into two subsets: one subset for model development (four folds) and the remaining subset for model validation (one-fold). Additionally, a fine-tuning technique was applied to identify the optimal hyperparameter sets for each model. The details of the corresponding results are provided in Table 3. To validate the predictive capability of the models and to assess the potential risk of overfitting, the performance of the models was assessed on both the training and testing datasets. The evaluation results based on statistical indices are summarized in Table 4.

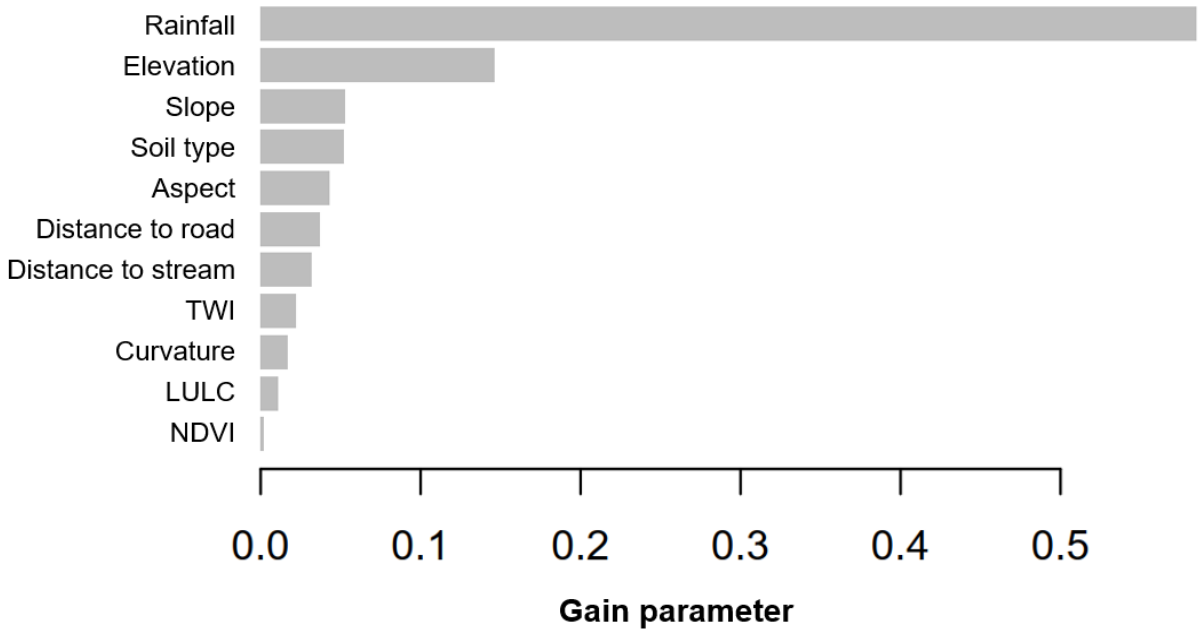


Fig. 8. Feature importance with the XGBoost method

Table 3. The optimal hyperparameter sets of the XGBoost and Random Forest (RF) models

Model	Parameter	Parameter tuning	Optimal hyperparameter
XGBoost	nrounds	100, 200, 500	200
	eta	0.01, 0.025, 0.05, 0.3	0.025
	max.depth	2, 4, 6	6
	colsample_bytree	0, 0.5, 1	1
	min_child_weight	0, 0.5, 1	0.5
	gamma	0, 0.1, 0.2	0
	subsample	0.8, 1	0.8
Random Forest	ntree	100–500	274
	mtry	2–5	2
	maxnodes	100–200	120

The prediction results obtained from both the training and testing datasets (Table 4) demonstrate that the XGBoost model consistently outperforms the RF model across all statistical metrics, including kappa (k), ACC, SST, SPF, PPV, and NPV. More importantly, the relatively small discrepancies between the training and testing results for both models indicate a high level of robustness and good generalization capability, with no clear evidence of overfitting. For the kappa

index, the reduction from the training to the testing dataset is limited to approximately 10% for the XGBoost model and 6% for the RF model. A similar pattern is observed for the ACC index, with decreases of only 4.6% for XGBoost and 2.5% for RF. These minor variations suggest that both models maintain stable predictive performance when applied to unseen data, reflecting their reliability in landslide susceptibility assessment.

Evaluation based on the testing dataset further confirms the superior and more stable performance of the XGBoost model. The kappa value of XGBoost reaches 0.688, exceeding that of the RF model (k = 0.671), while its ACC value attains 0.842 compared to 0.836 for RF. Additionally, the small differences observed between the paired indices SST and SPF, as well as PPV and NPV, for both models indicate a balanced and stable prediction of landslide and non-landslide classes. Overall, these results highlight the robustness and consistency of the XGBoost model, making it more suitable for landslide susceptibility prediction in the study area.

Model predictive performance was further evaluated through ROC curve analysis, as shown in Fig. 9. The results in Fig. 9a indicate a noticeable difference in AUC values between the two models when using the training dataset, with XGBoost achieving an AUC of 0.950, compared to 0.921 for the RF model. This suggests a stronger discriminative capability of XGBoost during the model training phase. For the testing dataset, the difference in AUC values between the two models becomes smaller (Fig. 9b), indicating comparable generalization performance. Nevertheless, XGBoost still exhibits a slightly superior predictive ability, with an AUC of 0.905, which is classified as “excellent”, while the RF model attains an AUC of 0.895, corresponding to a “good” prediction level. The limited reduction in AUC values from the training to the testing dataset for both models further confirms their robustness and indicates the absence of severe overfitting.

Overall, the above analyses demonstrate that, for the mountainous Phuoc Son area, the XGBoost model provides more accurate and robust landslide spatial predictions than the RF model. This finding is consistent with previous studies that compared the performance of XGBoost and RF in similar geomorphological settings, such as in Quang Ngai Province [12], [16]. Based on its superior predictive performance and stability, the XGBoost model was therefore selected to generate the landslide susceptibility map for the Phuoc Son mountainous region.

4.3. Landslide susceptible index

Table 4. The performance of the XGBoost and RF models

Statistical indexes	Training data		Testing data	
	XGBoost model	RF model	XGBoost model	RF model
k	0.765	0.714	0.688	0.671
ACC	0.883	0.857	0.842	0.836
SST	0.944	0.913	0.887	0.877
SPF	0.821	0.800	0.800	0.793
PPV	0.840	0.821	0.821	0.814
NPV	0.937	0.901	0.872	0.861
AUC	0.950	0.921	0.905	0.895

The landslide susceptibility index (LSI) for each pixel within the study region was calculated using the established XGBoost-based landslide susceptibility model and the set of conditioning factor maps, including slope, elevation, aspect, curvature, TWI, LULC, soil type, distance to road,

and distance to stream. For the rainfall-triggering factor, two rainfall scenarios were considered. In the first scenario, the observed 12-h accumulated rainfall associated with the landslide event on 28 October 2020 (Fig. 6) was used to assess the predictive performance of the susceptibility map for this specific event. In the second scenario, a uniform rainfall threshold of 320 mm/12 h, identified as the critical triggering rainfall, was applied across the entire study region to evaluate potential landslide susceptibility in other locations.

Fig. 10 illustrates the spatial distribution of LSI values under the two rainfall scenarios. Under the first scenario (Fig. 10a), the majority of the study area is characterized by very low landslide susceptibility, with LSI values lower than 0.2. In contrast, when the uniform rainfall threshold of 320 mm/12h is applied to the whole area, a clear increase in LSI values is observed (Fig. 10b), particularly in zones with high susceptibility (LSI > 0.6). This result highlights the strong control of extreme short-duration precipitation on landslide susceptibility levels in the mountainous Phuoc Son region.

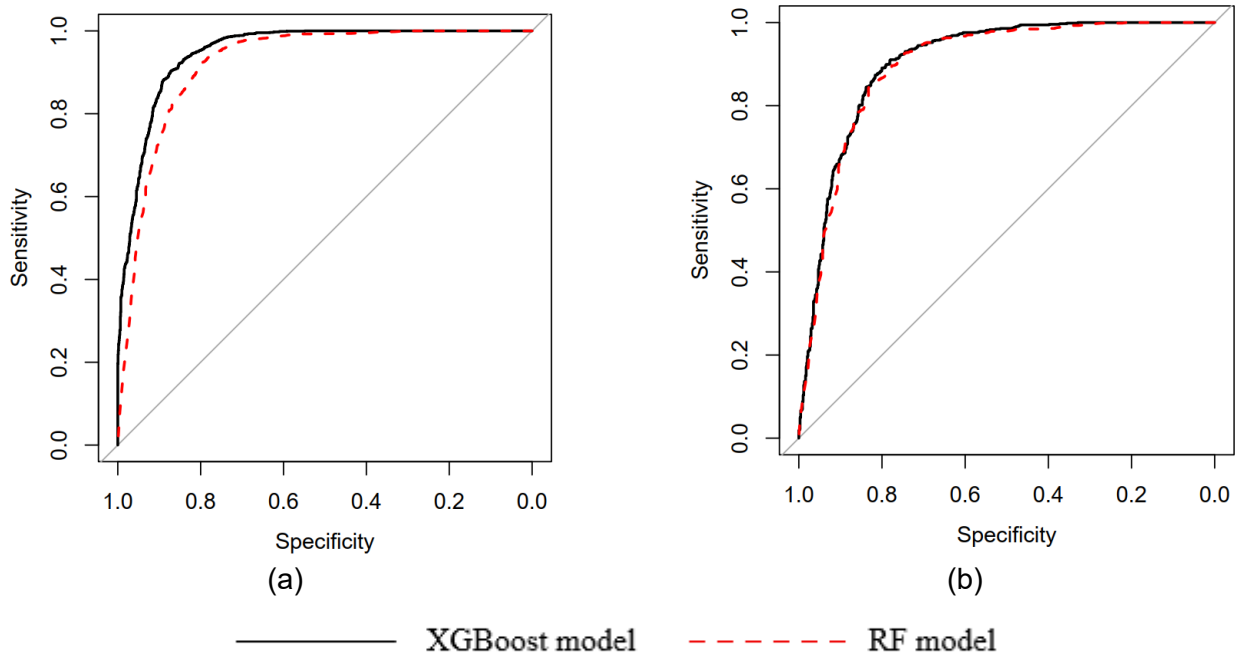


Fig. 9. The ROC result of the training model (a) and the testing models (b)

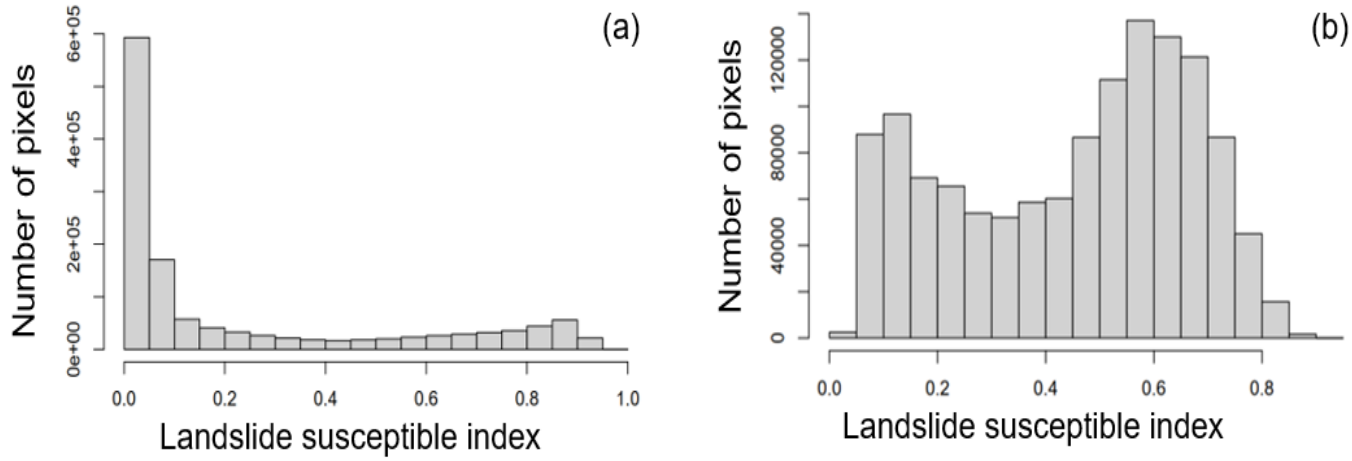


Fig. 10. The landslide susceptible index in two cases: (a) using 12-cumulative rainfall, (b) using a rainfall threshold of 320 mm

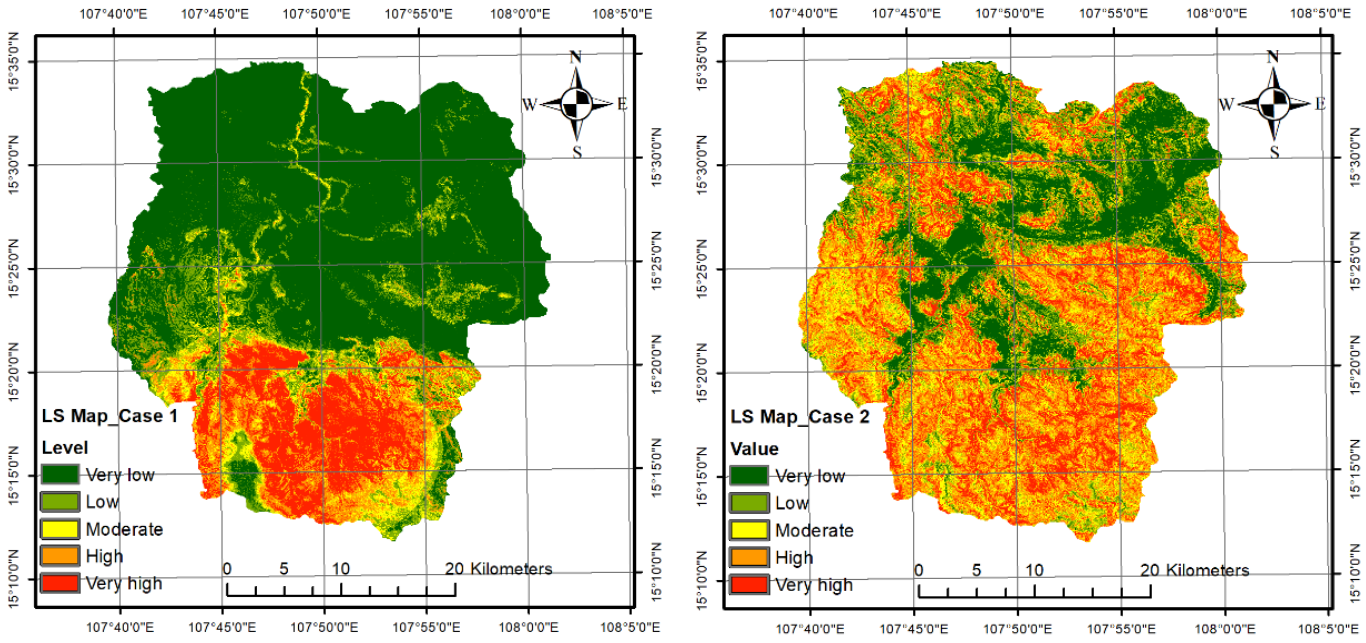


Fig. 11. The landslide susceptibility map for two cases: (a) using 12-hour cumulative rainfall, (b) using a rainfall threshold of 320 mm

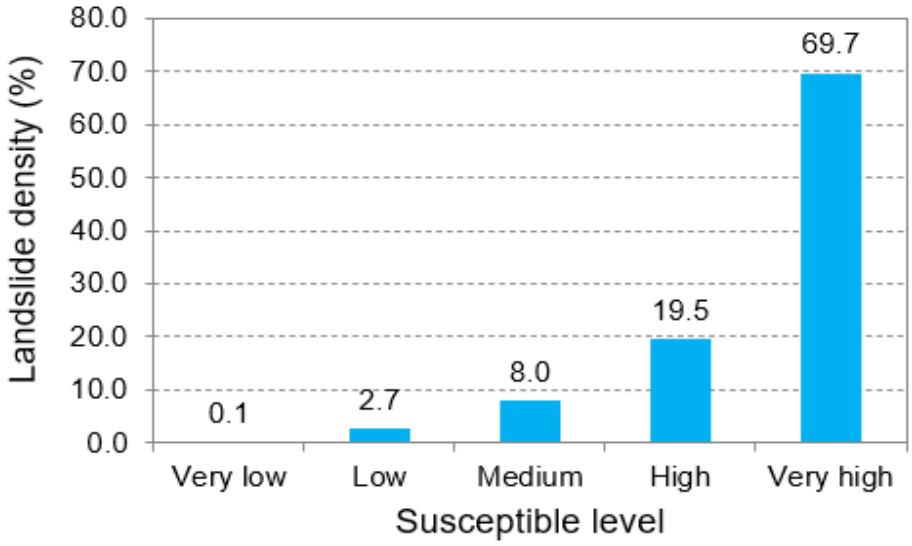


Fig. 12. The landslide density in the case using a 12-hour cumulative rainfall map

4.4. Landslide susceptibility mapping

At this stage, the LSI map was classified into five susceptibility levels: very low, low, moderate, high, and very high, using a color gradient ranging from blue to red. The Natural Breaks classification method was adopted to categorize the continuous LSI values into discrete susceptibility classes [19].

For Scenario 1, which employs the spatial distribution of 12-h accumulated rainfall (Fig. 11a), areas with a very high susceptibility level (red class) are mainly concentrated in the southern part of the study region. This spatial pattern corresponds closely with the zones experiencing the highest rainfall accumulation, as shown in the rainfall distribution map (Fig. 6). In contrast, the northern part of the study region is dominated by very low and low susceptible levels, which is consistent with the relatively low rainfall amounts observed in these areas. This strong spatial agreement confirms the dominant role of short-duration extreme rainfall in controlling landslide occurrence during the 28 October 2020 event. To

quantitatively evaluate the performance of these maps, a landslide density-based validation approach was applied [20]. The landslide density results (Fig. 12) indicate that 69.7% of the recorded landslide points are located within the very high susceptibility class, while an additional 19.5% fall within the high susceptibility class. In contrast, less than 3% of the landslide points occur in very low and low classes. These results demonstrate the high predictive accuracy and reliability of the proposed susceptibility map for the mountainous Phuoc Son region under the observed rainfall conditions.

In Scenario 2, a uniform extreme rainfall threshold of 320 mm/12h was applied across the entire study area (Fig. 11b) to evaluate potential landslide susceptibility under a spatially extensive extreme rainfall event. Under this scenario, areas classified as high and very high susceptibility expand noticeably toward the northern part of the study area, indicating a substantial increase in landslide susceptibility under extreme rainfall conditions. Fig. 13 quantitatively summarizes the areal proportions of susceptibility classes for both scenarios. Compared with Scenario 1, the proportion of low-susceptibility areas decreases markedly from 60.8% to 20.2%. In contrast, high-susceptibility areas increase from 8.7% to 26.0%, while very high susceptibility areas expand from 12.6% to 19.4%. The moderate susceptibility class also shows a significant increase in area, from 6.5% to 19.0%. These changes highlight the strong sensitivity of landslide susceptibility patterns to rainfall intensity and underscore the potential for widespread susceptibility escalation under extreme short-duration rainfall events.

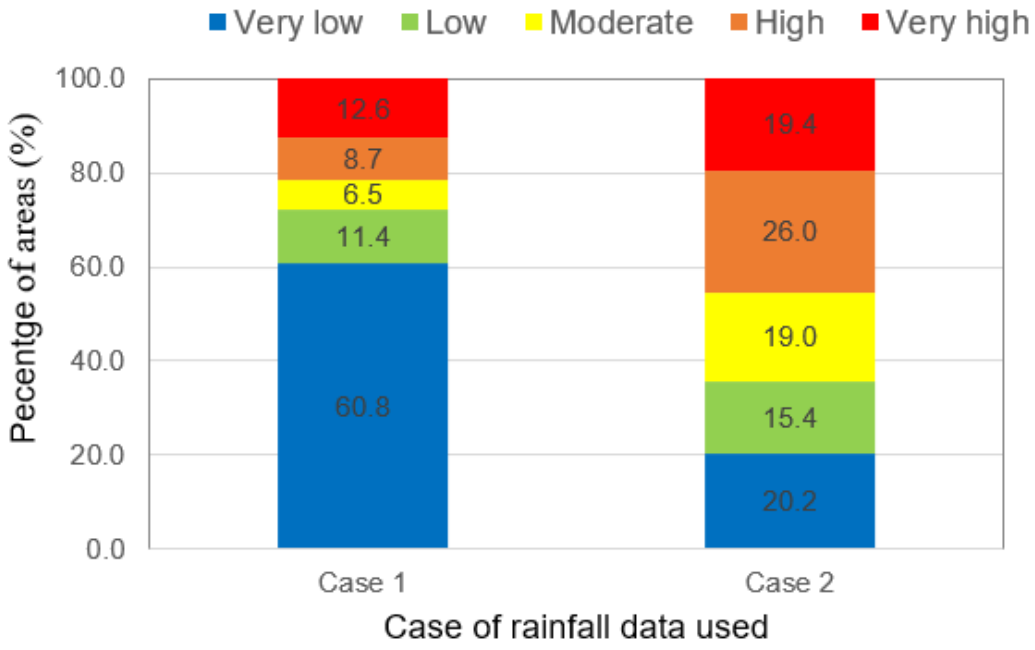


Fig. 13. The distribution proportions of landslide susceptible classes under two scenarios: (i) 12-hour cumulative rainfall and (ii) using a rainfall threshold of 320mm

5. Conclusion

This study proposes an event-based approach that utilizes a large landslide event that occurred on 28 October 2020 in the mountainous Phuoc Son area, Da Nang city, Vietnam, to assess landslide susceptibility. The principal conclusions drawn from this study are summarized as follows:

(i) Precipitation intensity and spatial distribution play a decisive role in landslide occurrence. Hourly rainfall analysis from five rain gauges revealed that extremely high rainfall concentrated within a short duration (12 hours) on 28 October 2020 was the primary trigger of landslides. In addition, the pronounced spatial

heterogeneity of rainfall contributed to the uneven spatial distribution of landslides across the study region. In particular, the southern part of the study area experienced the highest 12-hour accumulated rainfall, resulting in a concentration of landslide occurrences in this region.

(ii) The XGBoost model demonstrated excellent predictive performance for landslide spatial in the Phuoc Son mountainous area. The model consistently outperformed the Random Forest model in terms of statistical indices and ROC–AUC values.

(iii) Applying a rainfall threshold of 320 mm/12h provides valuable insights into regional-scale landslide susceptibility assessment. Based on the calibrated and validated XGBoost model, a landslide susceptibility map was produced using a uniform rainfall threshold of 320 mm/12h across the study area. Under this scenario, areas classified as high and very high susceptibility expanded northward, highlighting regions potentially exposed to landslide susceptibility during extreme rainfall events. These results offer important information for disaster risk reduction, land-use planning, and emergency management in mountainous areas.

(iv) The proposed approach is transferable to other regions with similar geomorphological and climatic conditions. However, this study still has several limitations: (i) the use of satellite-derived DEM data with relatively low spatial resolution (30 m × 30 m), from which multiple influencing factors were extracted; (ii) the low density of rainfall stations, which may introduce uncertainty in the interpolated rainfall maps; and (iii) the rainfall threshold for landslide occurrence was derived from a single event and has not been comprehensively validated. Therefore, future studies should focus on improving data accuracy and refining rainfall threshold estimation for landslide initiation, including the effects of prolonged and multi-day rainfall events, in order to better capture the complexity of rainfall–landslide relationships and enhance the robustness and generalizability of landslide susceptibility assessments.

Acknowledgments

This work was supported by The University of Danang - University of Science and Technology, code number of Project: T2025-02-15.

References

[1] P. Reichenbach, M. Rossi, B.D. Malamud, M. Mihir, and F. Guzzetti. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60–91. DOI:https://doi.org/10.1016/j.earscirev.2018.03.001

[2] S. Liu, L. Wang, W. Zhang, Y. He, and S. Pijush. (2023). A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geological Journal*, 58(6), 2283-2301. DOI: https://doi.org/10.1002/gj.4666

[3] F. Guzzetti, A.C. Mondini, M. Cardinali, F. Fiorucci, M. Santangelo, and K.-T. Chang. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1–2), 42–66. DOI: https://doi.org/10.1016/j.earscirev.2012.02.001

[4] T.A. Tuan, T.T. Tam, P.V. Hong, N.T.A. Nguyet, and Q.-V. Doan. (2025). Landslide susceptibility mapping in tropical cyclone-affected areas of Central Vietnam using bivariate statistical models. *IOP Conference Series: Earth and Environmental Science, IOP Publishing*, 1501, 012008. DOI: https://doi.org/10.1088/1755-1315/1501/1/012008

[5] N.T.T. Hien and D.T. Hue. (2024). Landslide susceptibility assessment in Quang Nam province using Statistical Index and Analytical Hierarchical process. *Journal of Science Natural Science*, 69(1), 144–160. DOI: https://doi.org/10.18173/2354-1059.2024-0014

[6] B.-Q.-V. Nguyen and V.-L. Doan. (2026). The role of conditioning factors in machine learning-based landslide spatial probability. *Environmental Geotechnics*, pp. 1–20. DOI: https://doi.org/10.1680/jenge.25.00077

[7] D.T. Bui, B. Pradhan, O. Lofman, I. Revhaug, and Ø.B. Dick. (2013). Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Natural Hazards*, 66, 707–730. DOI: https://doi.org/10.1007/s11069-012-0510-0

[8] B.T. Pham, D.T. Bui, H.V. Pham, H.Q. Le, I. Prakash, and M.B. Dholakia. (2017). Landslide hazard assessment using random subspace fuzzy rules based classifier ensemble and probability analysis of rainfall data: a case study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *Journal of the Indian Society of Remote Sensing*, 45, 673–683. DOI: https://doi.org/10.1007/s12524-016-0620-3

[9] D.T. Bui, B. Pradhan, O. Lofman, I. Revhaug, and O.B. Dick. (2012). Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. *International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany.*

[10] C. Su, L. Wang, X. Wang, Z. Huang, and X. Zhang. (2015). Mapping of rainfall-induced landslide susceptibility in Wencheng, China, using support vector machine. *Natural Hazards*, 76, 1759–1779. DOI: https://doi.org/10.1007/s11069-014-1562-0

[11] W. Zhang, H. Li, L. Han, L. Chen, and L. Wang. (2022). Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(4), 1089–1099. DOI: https://doi.org/10.1016/j.jrmge.2021.12.011

[12] V.L. Doan, B.-Q.-V. Nguyen, C.C. Nguyen, and C.T. Nguyen. Effect of time-variant rainfall on landslide susceptibility: A case study in Quang Ngai Province, Vietnam. *Vietnam Journal of Earth Sciences*, 46(2), 202–220, 2024. DOI: https://doi.org/10.15625/2615-9783/20065

[13] R. Can, S. Kocaman, and C. Gokceoglu. (2021). A comprehensive assessment of XGBoost algorithm for landslide susceptibility mapping in the upper basin of Ataturk dam, Turkey. *Applied Sciences*, 11(11), 4993. DOI: https://doi.org/10.3390/app11114993

[14] E.K. Sahin. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2, 1308. DOI: https://doi.org/10.1007/s42452-020-3060-1

[15] V.-H. Dang, N.-D. Hoang, L.-M.-D. Nguyen, D.T. Bui, and P. Samui. (2020). A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. *Forests*, 11(1), 118. DOI: https://doi.org/10.3390/f11010118

[16] V.L. Doan, B.-Q.-V. Nguyen, H.T. Pham, C.C. Nguyen, and C.T. Nguyen. (2023). Effect of time-variant NDVI on landslide susceptibility: A case study in Quang Ngai province, Vietnam. *Open Geosciences*, 15(1), 20220550. DOI: https://doi.org/10.1515/geo-2022-0550

[17] C.C. Nguyen, P. Vo, V.L. Doan, Q.B. Nguyen, T.C. Nguyen, and Q.D. Nguyen. (2023). Assessment of the Effects of Rainfall Frequency on Landslide Susceptibility Mapping Using AHP Method: A Case Study for a Mountainous Region in Central Vietnam. *Progress in Landslide Research and Technology, 1(2), 2022, Springer, pp. 87–98.* DOI: https://doi.org/10.1007/978-3-031-18471-0_7

[18] N.D. Vo and P. Gourbesville. (2016).

Application of deterministic distributed hydrological model for large catchment: a case study at Vu Gia Thu Bon catchment, Vietnam. *Journal of Hydroinformatics*, 18(5), 885–904. DOI: https://doi.org/10.2166/hydro.2016.138

[19] A. Basofi, A. Fariza, A.S. Ahsan, and I.M. Kamal. (2015). A comparison between natural and Head/tail breaks in LSI (Landslide Susceptibility Index) classification for landslide susceptibility mapping: A case study in Ponorogo, East Java, Indonesia. *2015 International Conference on Science in Information Technology (ICSITech), IEEE, 2015, pp. 337–342*. DOI: https://doi.org/10.1109/ICSITech.2015.7407828

[20] B. Pradhan and M.I. Sameen. (2017). Landslide susceptibility modeling: optimization and factor effect analysis. *Laser Scanning Applications in Landslide Assessment, pp. 115–132*. DOI: https://doi.org/10.1007/978-3-319-55342-9_6

[21] B.T. Pham, D.T. Bui, P. Indra, and M.B. Dholakia. (2015). Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS–based statistical approach of frequency ratio method. *International Journal of Engineering Research & Technology (IJERT)*, 4(11), 338–344. DOI: https://doi.org/10.17577/IJERTV4IS110285

[22] V. Vakhshoori and M. Zare. (2016). Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomatics, Natural Hazards and Risk*, 7(5), 1731–1752. DOI: https://doi.org/10.1080/19475705.2016.1144655

[23] D.T. Bui, B. Pradhan, O. Lofman, I. Revhaug, and O.B. Dick. (2012). Landslide Susceptibility Assessment at Hoa Binh Province of Vietnam Using Frequency Ratio Model. *2012 Asia Pacific Conference on Environmental Science and Technology. Advances in Biomedical Engineering, vol. 6, pp. 476–484*.

[24] L. Breiman. (2001). Random forests. *Machine Learning*, 45, 5–32. DOI: https://doi.org/10.1023/A:1010933404324

[25] K. Taalab, T. Cheng, and Y. Zhang. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, 2(2), 159–178. DOI: https://doi.org/10.1080/20964471.2018.1472392

[26] K. Zhang, X. Wu, R. Niu, K. Yang, and L. Zhao. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. *Environmental Earth Sciences*, 76, 405. DOI: https://doi.org/10.1007/s12665-017-6731-5

[27] T. Chen and C. Guestrin. (2016). XGBoost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785

[28] Y.W. Rabby, M.B. Hossain, and J. Abedin. (2022). Landslide susceptibility mapping in three Upazilas of Rangamati hill district Bangladesh: application and comparison of GIS-based machine learning methods. *Geocarto International*, 37(12), 3371–3396. DOI: https://doi.org/10.1080/10106049.2020.1864026

[29] P. Frattini, G. Crosta, and A. Carrara. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, 111(1–4), 62–72. DOI: https://doi.org/10.1016/j.enggeo.2009.12.004

[30] D.T. Bui et al. (2018). Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, Malaysia. *Remote Sensing*, 10(10), 1527. DOI:

https://doi.org/10.3390/rs10101527

[31] B.T. Pham et al. (2019). A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto International*, 34(13), 1385–1407. DOI: https://doi.org/10.1080/10106049.2018.1489422

[32] M.A. Hussain et al. (2022). Landslide susceptibility mapping using machine learning algorithm. *Civil Engineering Journal*, 8(2), 209–224. DOI: http://dx.doi.org/10.28991/CEJ-2022-08-02-02

[33] M. Kantardzic. (2011). Data mining: concepts, models, methods, and algorithms. *John Wiley & Sons*.

[34] T.A. Tuan et al. (2025). Landslide susceptibility in Phuoc Son, Quang Nam: A deep learning approach. *Vietnam Journal of Earth Sciences*, 47(1), 39–57. DOI: https://doi.org/10.15625/2615-9783/21658