Vietnam Journal of Earth Sciences, 46(2), 202-220, https://doi.org/10.15625/2615-9783/20065

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Effect of time-variant rainfall on landslide susceptibility: A case study in Quang Ngai Province, Vietnam

Viet Long Doan[1*], Ba-Quang-Vinh Nguyen[2,3], Chi Cong Nguyen[1], Cuong Tien Nguyen[4,5]

[1]*The University of Danang, University of Science and Technology, Danang, Vietnam*
[2]*School of Civil Engineering and Management, International University, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam*
[3]*Vietnam National University, Ho Chi Minh City, Vietnam*
[4]*Faculty of Vehicle and Energy Engineering, Phenikaa University, Hanoi 12116, Vietnam, Phenikaa*
[5]*Research and Technology Institute (PRATI), No. 167 Hoang Ngan, Cau Giay, Hanoi 11313, Vietnam*

Received 06 December 2023; Received in revised form 19 January 2024; Accepted 01 February 2024

ABSTRACT

Rainfall is a triggering factor that causes landslides, especially in the regions where landslides often occur after consecutive days of heavy rainfall. Most previous studies only used a specific rainfall map for landslide susceptibility assessment. However, this approach was unreasonable because rainfall is a time-variant data. This study uses the time series data of 1-day, 3-day, 5-day, and 7-day maximum precipitation from 2016 to 2020 in the mountainous area of Quang Ngai province for landslide susceptibility assessment. These data and other influencing factors were used to develop landslide spatial prediction models using the Extreme Gradient Boosting method. The prediction model's performance was assessed using the statistical index and receiver operating characteristic curve methods. The testing results of 4 cases using consecutive days of maximum rainfall data demonstrated excellent performance. Of these, the model with a 3-day maximum rainfall with ACC = 0.813, kappa = 0.625, SST = 0.872, SPF = 0.754, and AUC = 0.895 had the best performance. In addition, these results were compared to the previous approach that used average annual rainfall. The validation result indicates that the cases using a time series of maximum precipitation (with AUC of approximately 0.9) outperform the cases with average annual rainfall (AUC=0.838). Finally, the model using 3-day maximum rainfall is then used for landslide spatial prediction mapping. These maps provide spatial prediction and assess landslide susceptibility corresponding to rainfall frequencies.

Keywords: Time-variant rainfall, landslide susceptibility, XGBoost, Boruta, ROC.

1. Introduction

Landslides are one of the most dangerous disasters in the world. The study (2018) indicates that landslides occur on all five continents, especially in Asia, causing severe damage to causalities, infrastructure, and properties. In Vietnam, landslides occur yearly in the Northern, Central, and Central Highland mountainous provinces. In particular, the major landslide events in October 2020 in the mountainous area of

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

central provinces caused significant damage (111 people dead and missing). To reduce damage, a landslide susceptibility map can help prevent this natural disaster (Liu et al., 2023). The approaches and methods for generating landslide susceptibility can be divided into geomorphological mapping, landslide inventory analysis, heuristic approaches, process-based methods, and statistical-based methods (Reichenbach et al., 2018). For statistical-based methods, the spatial prediction of landslides is calculated by analyzing the spatial relationship between past landslide occurrences and a set of influence factors (Guzzetti et al., 2005). The affecting factors can be classified into two groups: (i) geo-environment factors such as geomorphological factors, geological factors, hydrological factors, land use/land cover, and (ii) triggering factors such as earthquakes, rainfall, human activities, etc. Out of these factors, rainfall is one of the external factors that causes landslides because it increases the pore water pressure, leading to a decreased shear strength of soil and rocks in slope-forming materials (Varnes, 1984; Dao et al., 2023; Pham et al., 2023). Under the impact of climate change, extreme rainfall is rising, significantly affecting landslides (Tran & Neefjes, 2015; Hoang et al., 2022; Ngo-Duc, 2023).

Two types of rainfall data are widely utilized for landslide susceptibility assessment: the first one is the average annual rainfall data, and the second one is the cumulative rainfall data. Almost all studies used average annual rainfall data for landslide susceptibility assessment because this data is often available for all regions (Dahal et al., 2008; Reichenbach et al., 2018). However, this approach is less reasonable because this rainfall type cannot represent a triggering factor that causes landslide occurrences, especially in regions such as Vietnam, where slope failures occur in the rainy season. Therefore, rainfall does not accurately reflect its influence on landslide spatial prediction in previous studies (Adnan Ikram et al., 2023; Le et al., 2023; Moayedi & Dehrashid, 2023).

There are only a few studies using cumulative precipitation for landslide spatial prediction. This is mainly due to the limited availability of landslide inventory data, such as the time occurrence of landslides, as well as detailed rainfall data. However, there are some problems when using cumulative rainfall data in these previous studies. Bui et al. (2012) used the rainfall map constructed from an eight-day maximum of precipitation as a conditioning factor for landslide susceptibility assessment. However, this study did not explain why this rainfall data was used. Su et al. (2015) utilized two triggering factors of daily and cumulative rainfall data as the input data for the landslide susceptibility model. Nevertheless, this study did not mention the relationship between the landslide inventories, such as the date of landslide occurrence and rainfall data. Zhang et al. (2022) used 5 cumulative rainfall maps from 1-day to 5-day to assess the susceptibility of landslides in the Shenzhen region, China.

The results indicated that the model of 4-day cumulative rainfall has the best performance compared to other cases. However, it was unsuitable because this study only uses specific rainfall data related to a landslide event to represent a series of inventory data (from 2008 to 2018). Using detailed cumulative rainfall data of a landslide event, Dou et al. (2019) have provided a spatial prediction of the rainfall-induced

landslide. The 32-hour cumulative rainfall has been used for analysis, and the result indicated that rainfall was found to have the highest influence on slope failure occurrence. However, this approach only used a specific landslide event, so it is unsuitable for large scale and cannot make long-term predictions. There are many problems in assessing the impact of cumulative rainfall on landslide susceptibility, especially for areas with limited detailed data.

The mountainous region of Quang Ngai province is significantly affected by landslides, and cumulative rainfall is considered the leading cause of this disaster (Phuoc et al., 2019). To provide a suitable approach for rainfall-triggered landslide assessment in the regions with data limitations, this study uses time-variant rainfall data for landslide susceptibility mapping in the mountainous area of Quang Ngai Province, Vietnam. First, the time series of landslide events from 2016 to 2020 are collected based on project surveys and remote sensing techniques. Next, the daily rainfall data from 18 rain gauge stations in this area generates the maximum cumulative rainfall maps corresponding to 1 day, 3 days, 5 days, and 7 days. Then, these rainfall maps, along with the inventory data and other conditioning factors (slope, aspect, elevation, land use, topographic wetness index (TWI), curvature, soil type, land use, distance to stream, road distance), are utilized to build the landslide susceptibility models. The machine learning of the Extreme Gradient Boosting (XGBoost) model with good predictive ability (Can et al., 2021; Rabby et al., 2022; Sahin, 2020) is proposed in this study. After that, the statistical index method and the receiver operating characteristic (ROC) curve method are used to evaluate the performance of landslide susceptibility models. This study also compares the predictive capability of these models to the model using average annual rainfall data that was widely used in previous studies. Finally, the best model fit is used for landslide spatial prediction mapping.

2. Study area

The study area covers an area of about 3,237 km^2 in five mountainous districts of Quang Ngai province, central Vietnam, between 14°32′ to 15°25′ N and 108°06′ to 109°04′ E) (Fig. 1). The province's highest peak is at 1,598 m in the west of the province near the Central Highland provinces and the Truong Son Mountain range. The East of this area is adjacent to the coastal delta. Annually, the area is hit by an average of 3 to 17 tropical storms and heavy rainfall from September to December. More than 70% of the region's annual rainfall (about 2,500 mm) is accounted for by rainfall from tropical storms and tropical monsoons. The long rainfall duration is more likely to significantly impact landslide occurrences in this area (Phuoc et al., 2019). According to the report by The Vietnam Institutes of Geosciences and Mineral Resources (2020), the mountainous area of Quang Ngai is one of the regions with the highest landslide density in Vietnam (landslide density = 0.167 event/km^2). And most landslides occurred after consecutive days of heavy rainfall. That is the reason while this area is selected for evaluating the effects of rainfall on landslide susceptibility.

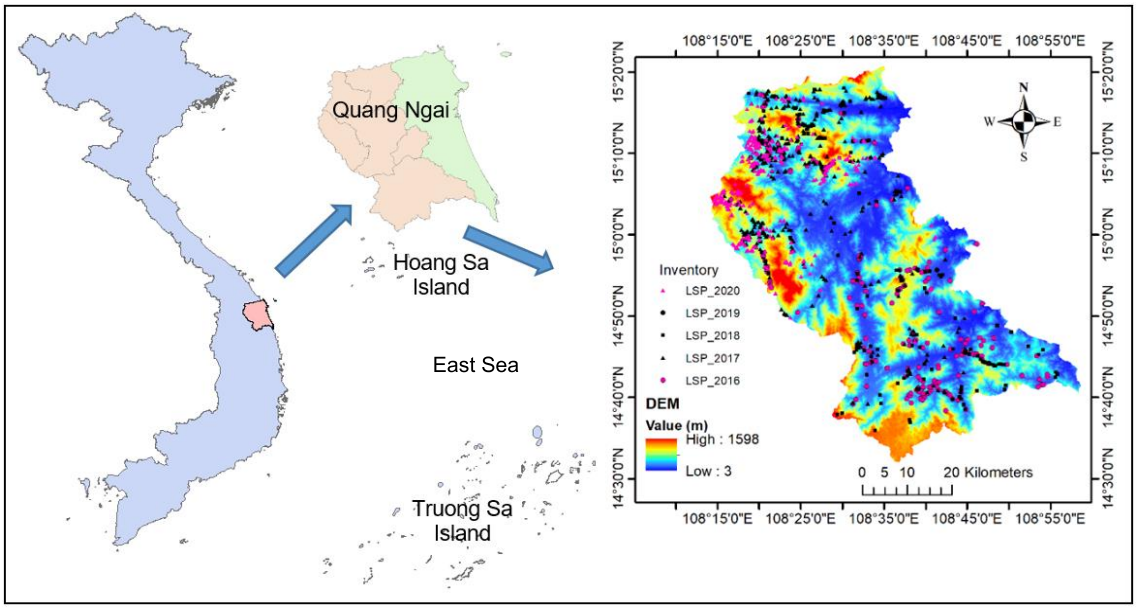

Figure 1. The study area of Quang Ngai mountainous region (Vietnam)

3. Data and Method

The methodology of this study is illustrated in Fig. 2, followed by the main steps (Pham et al., 2017):

Step 1: Develop time series of consecutive days of maximum rainfall data from 2016 to 2020 based on time series landslide inventories and daily rainfall data.

Step 2: Database preparation for landslide susceptibility model, including landslide inventory and landslide influencing factors of slope, aspect, elevation, land use, topographic wetness index, curvature, soil type, land use, distance to stream, distance to road, and the rainfall maps. There are 5 cases of the dataset, including:

Case 1: using 1-day maximum precipitation.

Case 2: using 3-day maximum precipitation.

Case 3: using 5-day maximum precipitation.

Case 4: using 7-day maximum precipitation.

Case 5: using average annual rainfall.

Step 3: Feature selection and importance evaluation using the Boruta method.

Step 4: Develop a landslide susceptibility model based on the XGBoost algorithm using 70% of the data.

Step 5: Validate and compare the models using the ROC curve, accuracy, sensitivity, and specificity values using 30% of the landslide data.

Step 6: Landslide spatial prediction mapping.

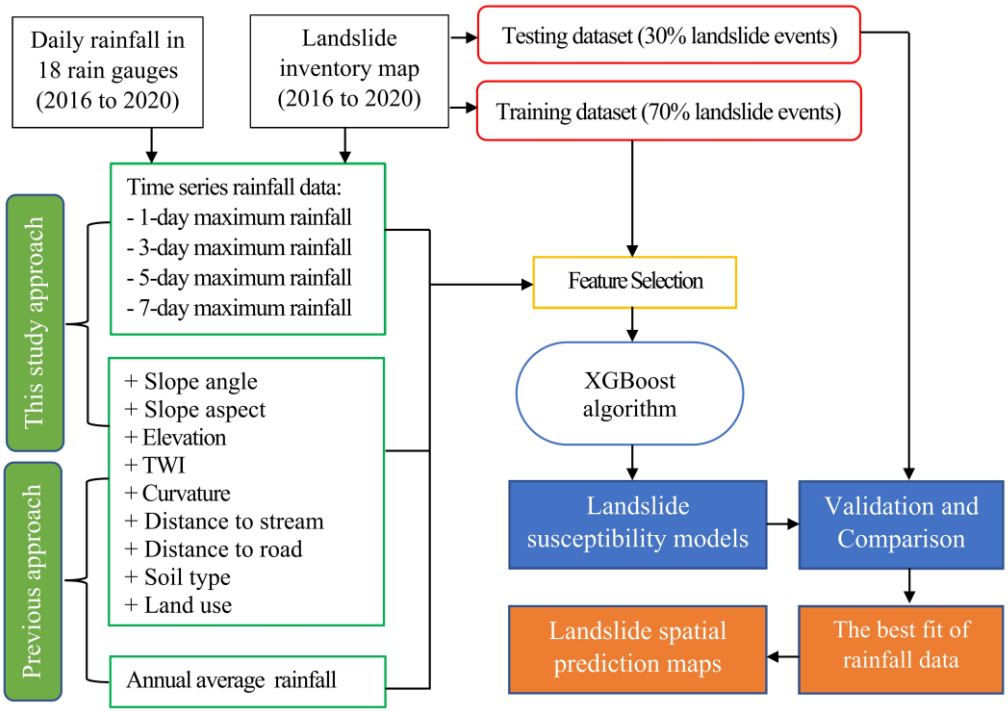

Figure 2. Flowchart of the methodology

3.1. Landslide inventory mapping

A landslide inventory map has been developed using a total of 1,279 landslide events that were identified by remote sensing and project survey. Among them, the project survey conducted by The Vietnam Institutes of Geosciences and Mineral Resources (2020) determined 549 landslide events in 5 mountainous districts of Quang Ngai province (Table 2). Additionally, the remote sensing technique using Google Earth images and Sentinel-2 satellite images (Table 1) has identified 730 landslide events (Table 2). The time series of landslide inventory data was created by change detection technique based on Sentinel-2 satellite images from 2016 to 2020 (Table 2 and Fig. 1). This data then has been divided into two groups: (i) training dataset (70% landslide inventory) and (ii) validation dataset (30% remaining landslide inventory) (Bui et al., 2014; Pham et al., 2019).

Table 1. Time collection of Sentinel-2 images for landslide detection

Year of landslide occurrences	Pre-event	Post-event
2016	10/09/2016	08/04/2017
2017	05/09/2017	03/04/2018
2018	05/09/2018	09/03/2019
2019	07/07/2019	08/03/2020
2020	06/07/2020	28/03/2021

Table 2. Time series landslide events from 2016 to 2020

Year	2016	2017	2018	2019	2020	Total
Number of landslides from Project survey	113	306	127	3	NA	549
Number of landslides from Remote sensing	117	204	60	2	347	730
Total	230	510	187	5	347	1,279

3.2. Landslide conditioning factors

3.2.1. Time series rainfall data

This study uses the daily rainfall of 18 rain gauge stations in Quang Ngai province and two adjacent Quang Nam and Binh Dinh provinces to map the consecutive days of maximum rainfall (Fig. 3). A total of 20 consecutive days of maximum rainfall maps are built by IDW interpolation in the period of 5 years from 2016 to 2020, corresponding to 1-day, 3-day, 5-day, and 7-day periods. These rainfall maps are then used to generate the time series rainfall data based on the time series of landslide inventories. In addition, the average annual rainfall map is also developed from the yearly rainfall of 18 rain gauge stations above.

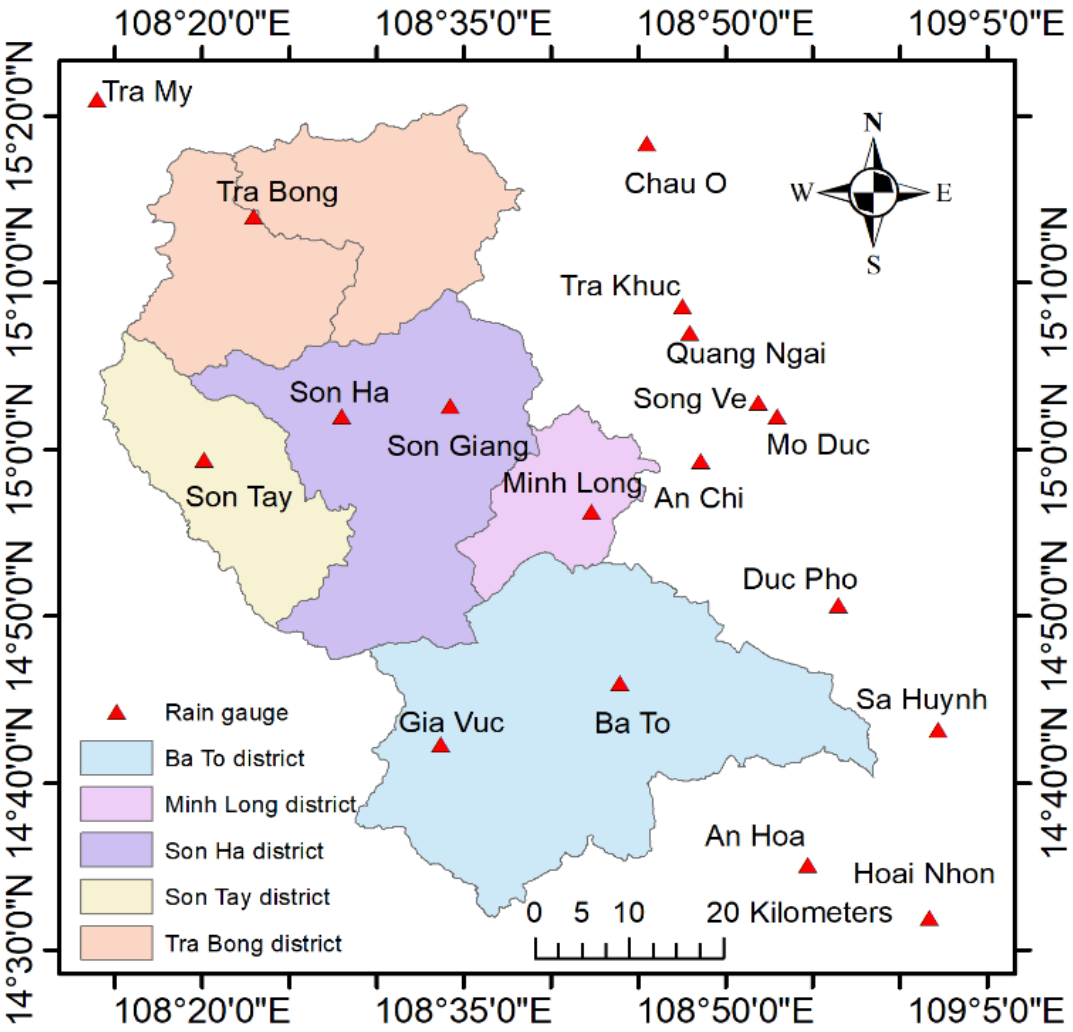

Figure 3. The location of rain gauge stations

3.2.2. Other conditioning factors

In addition to the rainfall factor, other conditioning factors are selected as the input dataset. They are slope, aspect, elevation, Topographic Wetness Index (TWI), curvature, distance to stream, distance to road, soil type, and land use. The detailed information of these influence factors is illustrated in Table 3.

Table 3. The information on landslide conditioning factors

No.	Landslide conditioning factors	Source	Scale/Resolution
1	Slope	NASA DEM	30×30 m
2	Aspect	NASA DEM	30×30 m
3	Elevation	NASA DEM	30×30 m
4	TWI	NASA DEM	30×30 m
5	Curvature	NASA DEM	30×30 m
6	Soil type	Departments of Natural Resource and Environment of Quang Ngai Province	1/100,000
7	Land use	https://landcovermapping.org	30×30 m
8	Distance to road (m)	Departments of Natural Resource and Environment of Quang Ngai Province	1/25,000
9	Distance to drainage (m)	Departments of Natural Resource and Environment of Quang Ngai Province	1/25,000

The effect of each factor on landslide occurrences is evaluated through the Frequency Ratio value (Binh Thai Pham et al., 2015). Specifically, landslides are more frequent in areas with slope angles from 20-30 degrees and less frequent in shallow (<10 degrees) or high slope angles (>50 degrees) (Fig. 4a). Regarding the aspect, landslides are more frequent on the southern-facing slope (Fig. 4b). For the elevation factor, most slope failures occurred at an elevation ranging from 400-600 m (Fig. 4c). According to the frequency analysis results of the land use factor, the highest FR value is found in the bush class. On the other hand, landslides did not occur in the classes of Water Body (WB), Urban and Built-up (UB), and wetland (Fig. 4d). In the case of curvature, Fig. 4e shows that slope failures are more frequent in the classes of concave and convex. Turning to the TWI factor, it can be observed that the landslides focus more on the lower TWI value (Fig. 4f). For soil type factor (Fig. 4g), the frequency analysis shows that most landslides occurred in the class of Epi Lithi Ferralic Acrisols (ELFA), followed by Epi Lithi Humic Acrisols (ELHA), Hapli Ferralic Acrisols (HFA), and Hapli Humic Acrisols (HHA). According to the distance to drainage factor, the FR value does not significantly change across classes (Fig. 4h). According to the frequency analysis of the road map, landslides are more frequent in the areas near the road, especially in the regions ranging from 0-50 m (Fig. 4i).

3.3. Feature selection

This study utilizes the Boruta method for evaluating the importance of conditioning factors and factor selection. The Boruta algorithm is a wrapper method that uses the random forest classification algorithm. This method has been implemented in the R package "Boruta" (Kursa & Rudnicki, 2010). The Boruta algorithm includes the following steps: (i) First, extend the information system by adding copies of all variables and shuffle the added attributes to remove their correlations with the response; (ii) Second, the random forest classifier calculates the Z score values on the extended information system and find out the maximum Z score among shadow attributes (MZSA). Lastly, the

influence factors with a Z score value better than MZSA were selected as the critical factors in the landslide susceptibility model (Prasad et al., 2022).

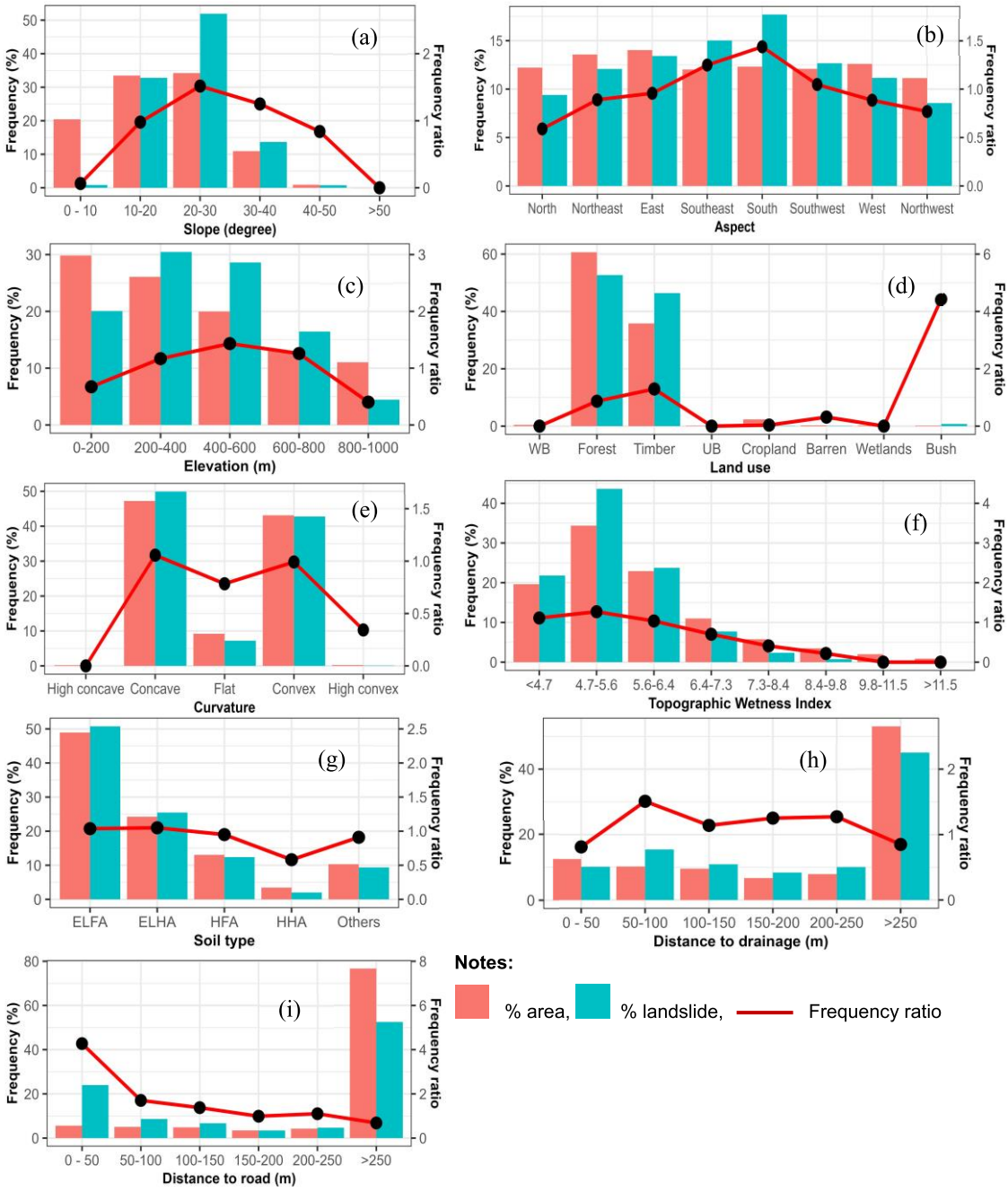

Figure 4. The frequency assessment of landslide conditioning factors

3.4. XGBoost method

XGBoost is a scalable machine-learning system for tree boosting (Chen & Guestrin, 2016). XGBoost creates many classification and regression trees (CART) and integrates them using the gradient boosting algorithm. XGBoost has three crucial aspects: regularized objective function for better generalization, gradient tree boosting for additive training, and shrinkage and column subsampling for preventing overfitting. The goal of the XGBoost algorithm is to minimize the following regularized objective function (Chen & Guestrin, 2016):

$$L(\emptyset) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k);\ \Omega(f) = \gamma T + \frac{1}{2}\lambda\left|\left|w\right|\right|^2 \quad (1)$$

where $\hat{y}_i$ and y_i are the predicted value and observed value, respectively; $\Omega(f)$ is the penalty term that helps to smooth the final learnt weights to avoid overfitting; γ, λ are the regularization degrees, respectively; w is the score of each leaf.

The main parameters of the XGBoost algorithm applied in landslide susceptibility assessment are grounds, max_depth, beta, gamma, colsample_bytree, min_child_weight, and subsample. The detailed information on these hyperparameters is shown in Table 4.

Table 4. The main parameters of the XGBoost algorithm

Parameter	Description
Grounds	Number of rounds
Max_depth	Maximum depth of a tree. Default: 6
Eta	Learning rate, 0<eta<1, default: 0.3. Low eta value means the model is more robust to overfitting but slower to compute.
Gamma	Regularization degree. The larger the gamma, the more conservative the algorithm.
Colsample_bytree	subsample ratio of columns when constructing each tree, default: 1
Min_child_weight	minimum sum of instance weight (hessian) needed in a child, default: 1.
Subsample	subsample ratio of the training instance, default: 1

3.5. Evaluation and comparison methods

This study uses several popular methods, such as statistical indexing and ROC, to evaluate landslide susceptibility (Frattini et al., 2010).

3.5.1. Statistical indexed methods

Four statistical indexes are chosen to evaluate the performance of the landslide susceptibility model: accuracy, sensitivity, specificity, and Kappa (Table 5).

Where TP is the value that indicates the number of pixels that have been predicted correctly as landslide; FP is the value that indicates the number of pixels that have been mispredicted as landslide; TN is the value that illustrates the number of pixels that have been predicted correctly as non-landslide; FN is a value indicating the number of pixels that have been mispredicted as non-landslide; Pobs is the proportion of number of pixels that have been classified correctly as landslide or non-landslide pixels; P_{exp} means the expected agreements.

Table 5. Some of the statistical indexes for landslide susceptibility assessment (Frattini et al., 2010)

Statistical indexes	Equation	Definition
Accuracy (ACC)	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	The proportion of landslide and non-landslide pixels that the resulting models correctly classified.
Sensitivity (SST)	$SST = \frac{TP}{TP + FN}$	The proportion of landslide pixels that are classified correctly as "landslide."
Specificity (SPF)	$SPF = \frac{TN}{TN + FP}$	The proportion of non-landslide pixels classified correctly as "no landslide."
Kappa (k)	$k = \frac{P_{obs} - P_{exp}}{1 - P_{exp}}$	The reliability of the landslide models.

3.5.2. ROC method

ROC method is usually used to assess the quality of landslide susceptibility models (Jones & Athanasiou, 2005). ROC curve represents the sensitivity value on the y-axis and (1-specificity) value on the x-axis. The Area Under the ROC Curve (AUC) can be used as an index to evaluate the overall performance of a model. The larger the area, the better the performance of the model. The AUC value can be divided into many intervals with the model quality respectively, including 0.6-0.7 (poor), 0.7-0.8 (fair), 0.8-0.9 (good), and 0.9-1.0 (very good) (Kantardzic, 2011)

4. Results

4.1. Effect of time series rainfall on landslide occurrences

This study uses four types of maximum rainfall, namely, 1 day, 3 days, 5 days, and 7 days, to evaluate the effect of rainfall on time series landslides from 2016 to 2020. Figure 5 shows the relationship between the consecutive days of maximum rainfall and the number of yearly landslide occurrences. These graphs illustrate the significant effect of consecutive days of maximum rainfall on the time series of landslide events in 5 years. For the case of 1-day maximum rainfall (Fig. 5a), a strong relationship between the rainfall and the landslides has been recognized. Landslide events are generally more frequent in the year with higher rainfall. Specifically, the highest number of landslide events (540 landslides) was found in 2017, along with the highest 1-day maximum rainfall this year (up to 600 mm).

On the other hand, the lowest number of landslide events (5 landslides) was in 2019, corresponding to the lowest 1-day maximum rainfall. Similar results appear in the cases of 3-day, 5-day, and 7-day maximum rainfall (Figs. 5b, c, d). The consecutive days of maximum rainfall can be seen as the leading cause of landslides in this region.

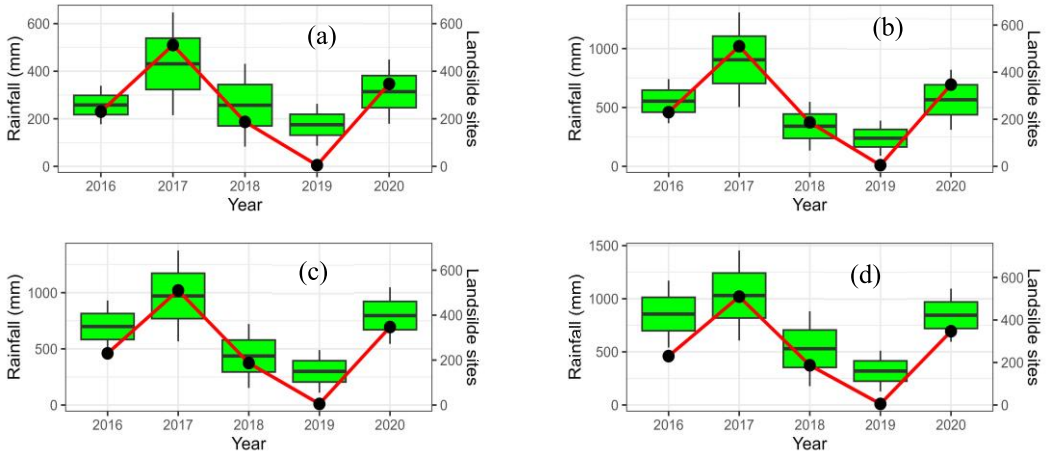

Figure 5. The relationship between the consecutive days of maximum rainfall and landslide events: a. 1-day maximum rainfall, b. 3-day maximum rainfall, c. 5-day maximum rainfall, and d. 7-day maximum rainfall

4.2. Important landslide variables

The Boruta method uses an importance index to evaluate the importance of the landslide conditioning factor. The higher the importance index, the more critical factor. Fig. 7 shows the results of Boruta assessment in five cases. Generally, no attributes are deemed unimportant, so all landslide conditioning factors are selected for landslide susceptibility modeling. Regarding the rainfall factor, the first position of 1-day, 3-day, 5-day, and 7-day maximum rainfall was found in s 6a with an importance value greater than 60. In contrast, the average annual rainfall was the third most influential variable (Fig. 7e), with an importance value under 30.

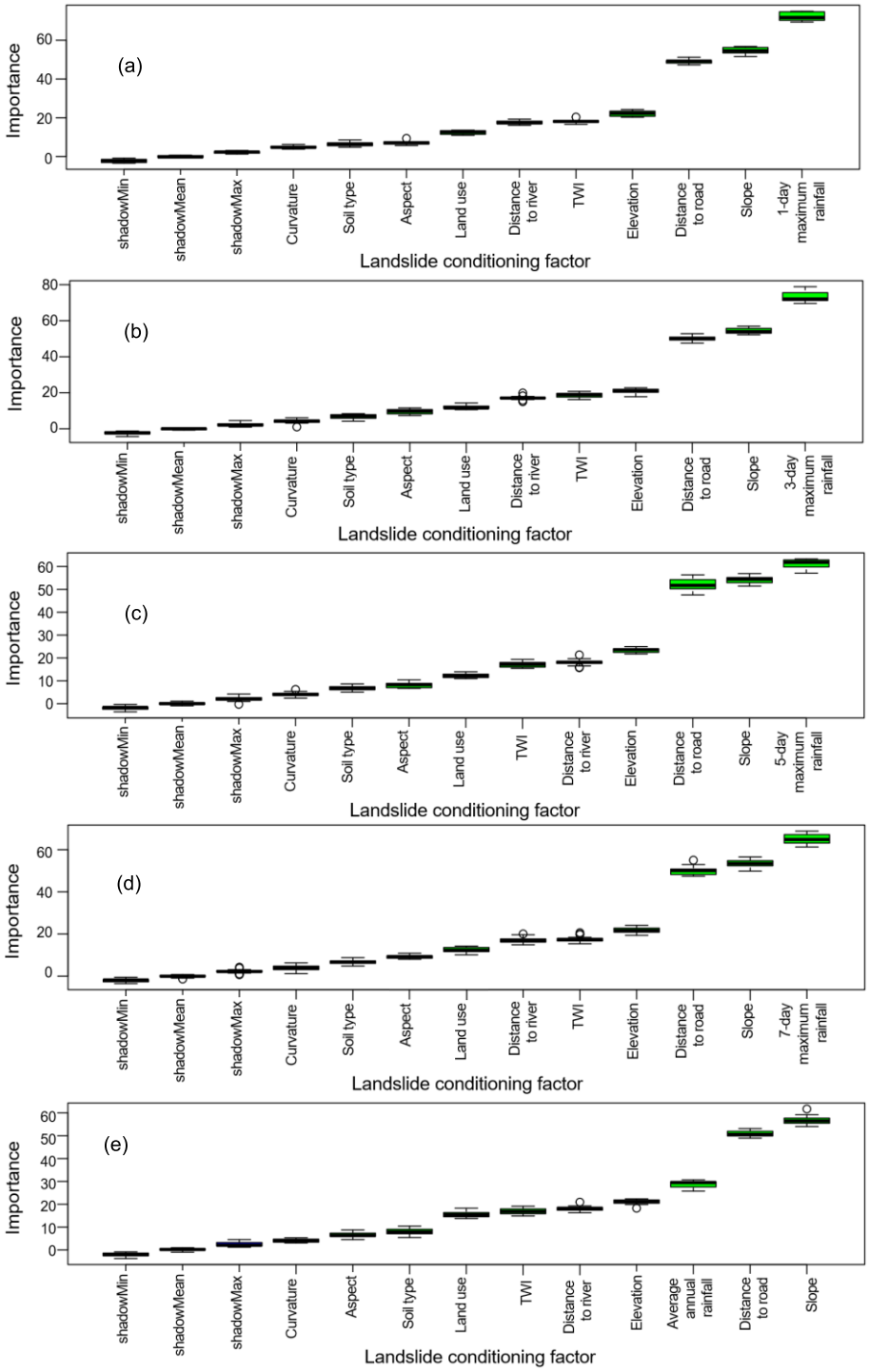

Figure 6. The importance of conditioning factors using the Boruta method in 5 cases of rainfall data: a. Case 1, b. Case 2, c. Case 3, d. Case 4, and e. Case 5

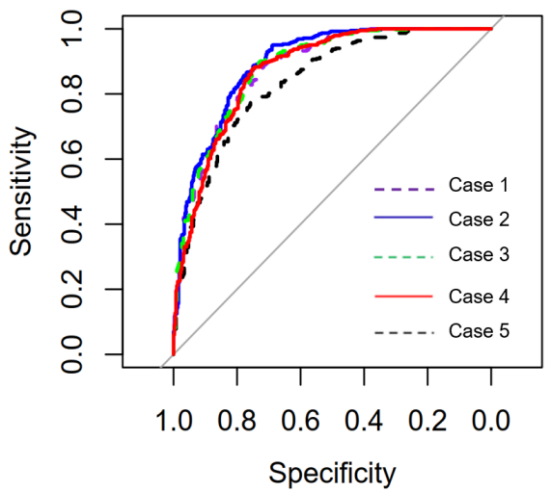

Figure 7. Analysis of ROC curves of the prediction model in 5 cases

4.3. Landslide susceptibility models

This study used 70% of landslide inventories for training models. The fine-tuning technique was utilized to find the optimized hyperparameters of the XGBoost model (Table 6). Additionally, 30% of landslide inventories were used for model validation. The result of statistical indexes and AUC value utilized for evaluating the predictive capability of five cases are shown in Table 7 and Fig. 8. Overall, the outcomes from ACC, k, SST, SPF, and AUC indicate that the cases using consecutive days of maximum rainfall data have moderately higher values than those utilizing average annual rainfall data. Specifically, results in Table 7 show that Case 2 with 3-day maximum precipitation has the highest value in all of ACC, k, SST, and SPF (0.813, 0.625, 0.872, and 0.754, respectively). The lowest ACC, k, SST, and SPF values belong to Case 5 with average annual rainfall (0.748, 0.497, 0.791, and 0.705, respectively).

Table 6. The optimized hyperparameters of the XGBoost model

Parameter	Values	The best value
Grounds	100/250/500	100
Max_depth	2/4/6	6
Eta	0.01/0.025/0.05/0.3	0.05
Gamma	0/0.1/0.2	0
Colsample_bytree	0/0.5/1	0.5
Min_child_weight	0/0.5/1	0
Subsample	0.8/1	1

Table 7. The testing result of landslide susceptibility models

Statistical Index	Case 1	Case 2	Case 3	Case 4	Case 5
Accuracy (ACC)	0.806	0.813	0.806	0.806	0.748
Kappa (k)	0.612	0.625	0.612	0.612	0.497
Sensitivity (SST)	0.869	0.872	0.854	0.869	0.791
Specificity (SPF)	0.742	0.754	0.754	0.742	0.705
Area Under the ROC Curve (AUC)	0.882	0.895	0.881	0.873	0.838

Regarding the ROC curves in Fig. 7, the highest AUC value is achieved by Case 2 (AUC = 0.895), and the lowest AUC value belongs to Case 5 (AUC = 0.838). The model results in Case 2 show an exceptionally reliable performance on landslide spatial prediction because the AUC values are approximately 0.9. Compared with Case 2, the model in Case 5 performs well as the AUC value ranges from 0.8 to 0.9.

4.4. Landslide spatial prediction mapping

Landslide spatial prediction maps have been developed in two main steps: (i) generating the landslide susceptible indexes (LSI) and (ii) reclassifying the landslide-susceptible indexes.

4.4.1. Generating the landslide susceptible indexes (LSI)

In the first step, the LSI of all pixels in the total region is generated based on the landslide spatial prediction model of case 2 (using 3-day maximum rainfall data) and a set of all sampling pixels of 10 conditioning factors. For rainfall maps, this study

recommends using annual maximum precipitation maps created by the Regional Frequency Analysis (RFA) method (Cong et al., 2019). Four 3-day annual maximum rainfall maps corresponding to the frequency of 50%, 20%, 10%, and 5% are recommended for assessment.

The frequency analysis results of LSI (Fig. 9) indicate the distribution of LSI along with the decrease in rainfall frequency. It can be observed that the high region of LSI rises gradually when the rainfall frequency is down from 50% to 10%. These rainfall maps have indicated a positive influence of landslides in this frequency period. However, when the rainfall frequency decreases from 10% to 5%, the high region of LSI shows a slight decrease in this period. This means that these rainfall maps have shown a negative effect on landslides. Therefore, the LSI value of the cases using 3-day annual maximum rainfall corresponding to the frequency of 50%, 20%, and 10% are used to generate landslide spatial prediction maps.

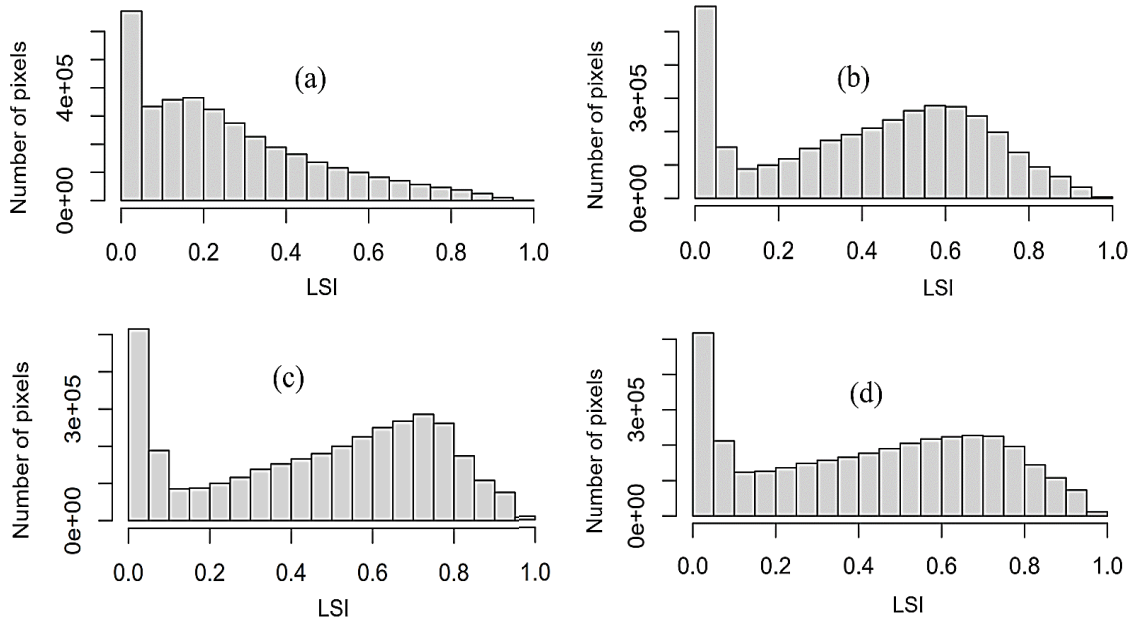

Figure 8. The histogram of landslide spatial probability uses 3-day annual maximum precipitation maps corresponding to the frequencies of 50%, 20%, 10%, and 5%

4.4.2. Generating the landslide spatial prediction maps

The reclassifying step generates landslide susceptibility maps based on the LSI results. In this step, five susceptible classes corresponding to LSI intervals have been created by the Natural Breaks method: very low, low, moderate, high, and very high (Basofi et al., 2015). As a result, three landslide spatial prediction maps are developed for the case of 3-day annual maximum rainfall corresponding to the frequency of 50%, 20%, and 10%, respectively (Fig. 10 to Fig. 12).

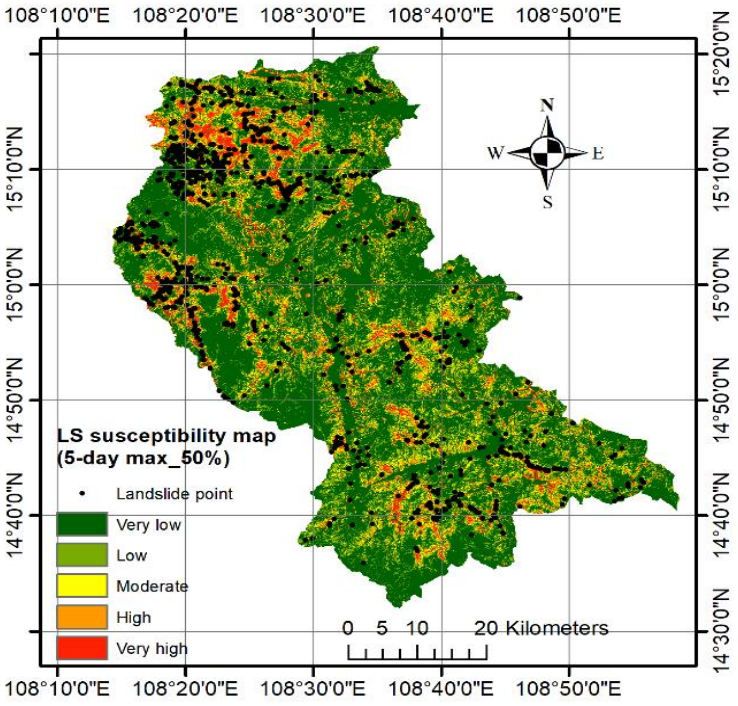

Figure 9. Landslide spatial prediction maps for three scenarios of 3-day annual maximum rainfall - frequency of 50%

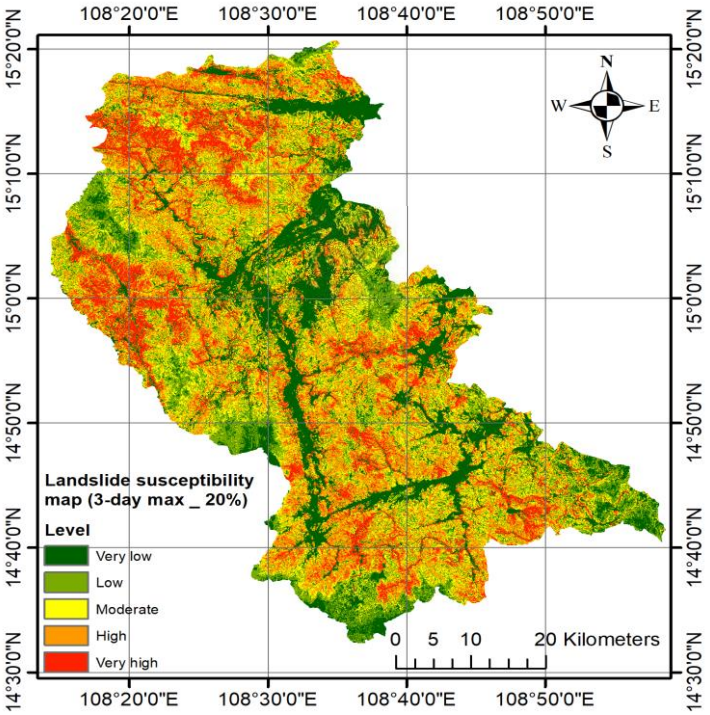

Figure 10. Landslide spatial prediction maps for three scenarios of 3-day annual maximum rainfall - frequency of 20%

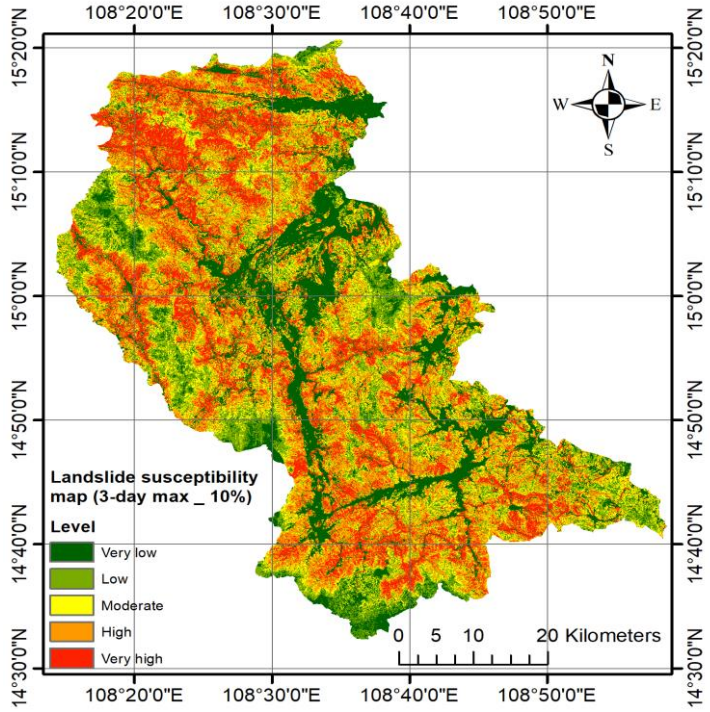

Figure 11. Landslide spatial prediction maps for three scenarios of 3-day annual maximum rainfall - frequency of 10%

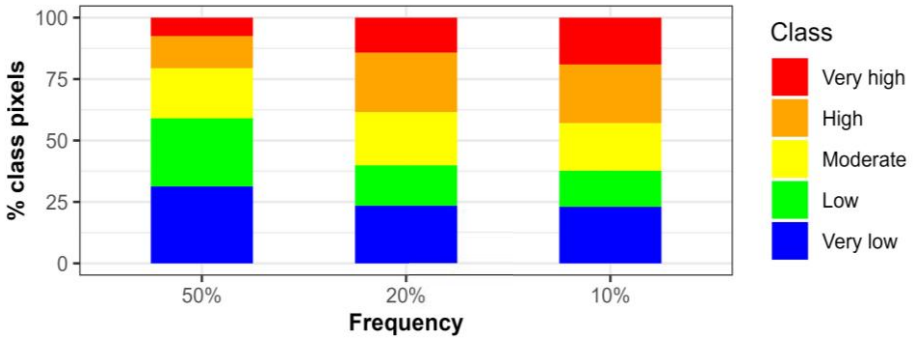

Figure 12. The percentage of class pixels of landslide spatial prediction maps corresponding to rainfall frequency scenarios

4.3. Evaluating the performance of landslide spatial prediction maps

The landslide susceptibility maps in this study provide the landslide spatial prediction corresponding to the different rainfall frequencies. Generally, the areas of susceptible level classes change along with decreased rainfall frequencies (Fig. 13). It can be observed that the area percent of the "very high" class gradually expands from 7.5% to 23.0% as the rainfall frequency declines from 50% to 10%. In the period of frequency falling from 50% to 20%, the area of "high" and "moderate" classes moderately rises, whereas the area of "low" and "very low" classes quickly drops. Turning to the successive reduction of frequency (from 20% to 10%), the area of "moderate" and "low" classes slightly decreases, whereas the area of other classes shows little change.

This study also evaluates the prediction abilities of landslide spatial prediction maps by assessing landslide density (LD) value. LD value is the ratio of the landslide pixel percentage and the class pixel percentage (Pham et al., 2017). In general, the result in Fig. 13 shows that the highest LD has been observed in the "very high" class, followed by the "high" class, "moderate" class, "low" class, and "very low" class, respectively.

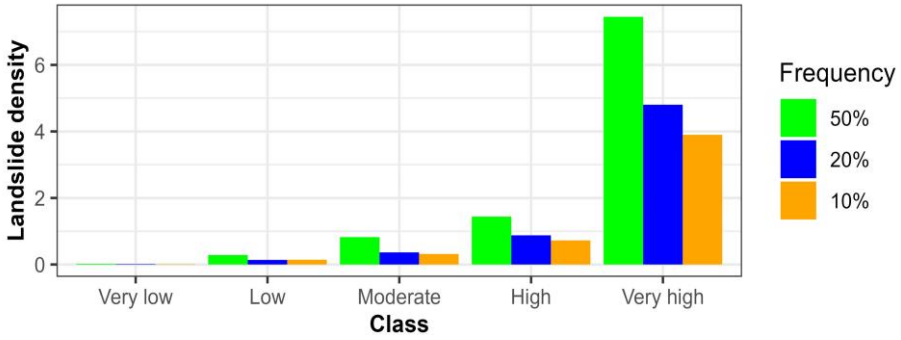

Figure 13. The landslide density of landslide spatial prediction maps corresponds to rainfall frequency scenarios

5. Discussions

Rainfall is the triggering factor that causes landslides in the mountainous area of Quang Ngai province. Applying a time series of consecutive days of precipitation for landslide assessment has shown a significant effect of rainfall on landslides (Fig. 5). The higher the consecutive days of maximum rainfall, the more occurrence of landslides. This approach is more reasonable than previous studies when only using an average annual rainfall map (Adnan Ikram et al., 2023; Le et al., 2023; Moayedi & Dehrashid, 2023) or a specific cumulative rainfall map (Bui et al., 2012; Zhang et al., 2022) for landslide assessment.

In evaluating the importance of landslide conditioning factors, Fig. 6a to 6d results indicate that consecutive days of maximum rainfall data influence landslides most. Regarding the case using average annual rainfall (Fig. 7e), this rainfall type only ranks in the 3rd position in terms of importance index. The consecutive days of maximum rainfall factor are more important than the average annual rainfall factor and the remaining factors. Previous studies also showed similar results when using average annual rainfall data on landslide susceptibility assessment. Le et al. (2023) have used thirteen conditioning factors for landslide spatial prediction mapping in Trung Khanh district, Cao Bang province, Vietnam. The critical assessment results indicated that average annual rainfall was the eighth most influential variable. Another study in Mu Cang Chai District, Yen Bai Province, Vietnam, showed that average annual rainfall was among the least important variables (Pham et al., 2019).

In this study, landslide susceptibility assessment using consecutive days of maximum rainfall data has improved the prediction model's performance. By using a powerful model of XGBoost and these rainfall data, landslide spatial prediction models show an exceptionally reliable performance as AUC is approximately 0.9 (Table 7 and Fig. 8). These results outperform the case using average annual precipitation (AUC = 0.838). This means that the approach of this study has improved the performance of the landslide susceptibility model when compared with the wide use of average annual rainfall in the previous approach. Additionally, the case using 3-day maximum rainfall has better

performance than others. This means that 3-day maximum rainfall significantly influences landslides in the mountainous area of Quang Ngai Province. This result is similar to the study of Phuoc et al. (2019) when assuming that 3-day maximum rainfall strongly correlates to landslides in this region.

By approaching the time series of 3-day maximum rainfall data in the model-building step, this study has generated a set of landslide spatial maps corresponding to different rainfall frequencies in the step of landslide susceptibility mapping. These maps illustrate the "where" slope failure will occur in this region and predict "how frequently" it appears. This is a significant improvement of this study compared to previous studies that use a specific rainfall map for landslide susceptibility mapping (Adnan Ikram et al., 2023; Le et al., 2023; Moayedi & Dehrashid, 2023). However, the 5 years of time series rainfall data (2016-2020) applied in this paper is not long enough to make a long-term prediction. Therefore, this study only provides the landslide predictions corresponding to the rainfall frequency from 50% to 10%. The landslide spatial prediction maps corresponding to the frequency of 50%, 20%, and 10% have highlighted the influence of rainfall on landslide susceptibility in this region (Fig. 10 to Fig. 12). The higher the 3-day maximum precipitation, the larger the "very high" class in the landslide susceptibility map (Fig. 13). In addition, most landslides are concentrated in the "very high" class, while very few landslides are found in the "low" and "shallow" classes (Fig. 14). This result indicates a reliable performance of these landslide spatial prediction maps. According to the very high class, the case with the frequency of 50% is found to have the highest LD; then, the LD has an upward trend along with the decrease in rainfall frequency. This can be explained by the fact that as the rainfall increases, the number of landslides and the "very high" class area increase. However, the "very high" class percentage increase is higher than the rise in the landslides' percentage, leading to a decrease in LD value.

6. Conclusions

This study has provided an approach to using rainfall data on landslide susceptibility assessment in the mountainous region of Quang Ngai province, Vietnam. The time series of consecutive days of maximum rainfall data is proposed to replace the average annual rainfall data widely used in previous studies. Overall, the findings of this study provide important insights into the effect of rainfall on landslide susceptibility. The feature selection results by Boruta method indicate that the consecutive days of maximum precipitation significantly influence landslide occurrences, whereas the average annual rainfall shows less importance. In addition, the validation results from XGBoost model show that the cases using consecutive days of maximum rainfall have excellent performance. The case with a 3-day maximum rainfall provides the best prediction ability among them.

On the contrary, the lowest performance belongs to the case using average annual rainfall data. Therefore, this study has demonstrated that the time series of consecutive days of maximum rainfall data has improved the landslide susceptibility model's prediction ability in the Quang Ngai province case study. By applying the annual maximum precipitation maps in calculating landslide susceptible indexes, this study can generate landslide spatial prediction maps corresponding to different rainfall frequencies. These maps indicate the strong influence of rainfall on landslide susceptibility. The higher the rainfall, the more the area of very high susceptible class expands. More importantly, these maps are handy for disaster prevention and land use management. This paper also recommends paying more attention to data collection, especially for detailed landslide inventories, to improve the prediction efficiency.

References

Adnan Ikram R.M., Khan I., Moayedi H., Ahmadi Dehrashid A., Elkhrachy I., Nguyen Le B., 2023. Novel evolutionary-optimized neural network for predicting landslide susceptibility. Environment, Development and Sustainability, 1–33. https://doi.org/10.1007/s10668-023-03356-0.

Basofi A., Fariza A., Ahsan A.S., Kamal I.M., 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), 337342. https://doi.org/10.1109/icsitech.2015.7407828.

Bui D.T., Pradhan B., Lofman O., Revhaug I., Dick O.B., 2012. Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. Proceedings of the sixth biannial meeting of the International Environmental Modelling and Software Society. Germany July 1-5, 2012, 382–089. Doi: 10.13140/RG.2.1.3041.4889.

Bui D.T., Pradhan B., Revhaug I., Tran C.T., 2014. A comparative assessment between the application of fuzzy unordered rules induction algorithm and J48 decision tree models in spatial prediction of shallow landslides at Lang Son City, Vietnam. In Remote sensing applications in environmental research. Springer, 87–111. https://doi.org/10.1007/978-3-319-05906-8_6.

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

Chen T., Guestrin C., 2016. Xgboost: A scalable tree boosting system. Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785.

Cong N.C., Binh N.Q., Phuoc V.N.D., 2019. Landslide Susceptibility Mapping by Combining the Analytical Hierarchy Process and Regional Frequency Analysis Methods: A Case Study for Quangngai Province (Vietnam). International Conference on Asian and Pacific Coasts, 1327–1334. https://doi.org/10.1007/978-981-15-0291-0_180.

Dahal R.K., Hasegawa S., Nonomura A., Yamanaka M., Masuda T., Nishino K., 2008. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. Environmental Geology, 54(2), 311–324. https://doi.org/10.1007/s00254-007-0818-3.

Dao M.D., Vu C.M., Hoang H.Y., Nguyen T.L., Do M.D., 2023. Analysis of landslide kinematics integrating weather and geotechnical monitoring data at Tan Son slow moving landslide in Ha Giang province. Vietnam J. Earth Sci., 45(2), 131–146.. https://doi.org/10.15625/2615-9783/18204.

Dou J., Yunus A.P., Bui D.T., Merghadi A., Sahana M., Zhu Z., Chen C.-W., Khosravi K., Yang Y., Pham B.T., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Science of the Total Environment, 662, 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221.

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

Guzzetti F., Reichenbach P., Cardinali M., Galli M., Ardizzone F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology, 72(1–4), 272–299. https://doi.org/10.1016/j.geomorph.2005.06.002.

Hoang-Cong H., Ngo-Duc T., Nguyen-Thi T., Trinh-Tuan L., Xiang C.J., Tangang F., Jerasorn S., Phan-Van T., 2022. A high-resolution climate experiment over part of Vietnam and the Lower Mekong Basin: performance evaluation and projection for rainfall. Vietnam J. Earth Sci., 44(1), 92–108. https://doi.org/10.15625/2615-9783/16942.

Jones C.M., Athanasiou T., 2005. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. The Annals of Thoracic Surgery, 79(1), 16–20. https://doi.org/10.1016/j.athoracsur.2004.09.040.

Kantardzic M., 2011. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 126–132. https://doi.org/10.1002/9781119516057.ch1.

Kursa M.B., Rudnicki W.R., 2010. Feature selection with the Boruta package. Journal of Statistical Software, 36, 1–13. https://doi.org/10.18637/jss.v036.i11.

Le Minh N., Truyen P.T., Van Phong T., Jaafari A., Amiri M., Van Duong N., Van Bien N., Duc D.M., Prakash I., Pham B.T., 2023. Ensemble models based on radial basis function network for landslide susceptibility mapping. Environmental Science and Pollution Research, 30(44), 99380–99398.

https://doi.org/10.1007/s11356-023-29378-9.

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

Moayedi H., Dehrashid A.A., 2023. A new combined approach of neural-metaheuristic algorithms for predicting and appraisal of landslide susceptibility mapping. Environmental Science and Pollution Research, 1–26. https://doi.org/10.1007/s11356-023-28133-4.

Ngo-Duc T., 2023. Rainfall extremes in Northern Vietnam: a comprehensive analysis of patterns and trends. Vietnam J. Earth Sci., 45(2), 183–198. https://doi.org/10.15625/2615-9783/18284.

Pham Binh T., Prakash I., Khosravi K., Chapi K., Trinh P.T., Ngo T.Q., Hosseini S.V, Bui D.T., 2019. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. Geocarto International, 34(13), 1385–1407. https://doi.org/10.1080/10106049.2018.1489422.

Pham Binh Thai, Bui D.T., Prakash I., Dholakia M.B., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena, 149, 52–63. https://doi.org/10.1016/j.catena.2016.09.007.

Pham Binh Thai, Khosravi K., Prakash I., 2017. Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India. Environmental Processes, 4(3), 711–730. https://doi.org/10.1007/s40710-017-0248-5.

Pham Binh Thai, Prakash I., Chen W., Ly H.-B., Ho L.S., Omidvar E., Tran V.P., Bui D.T., 2019. A novel intelligence approach of a sequential minimal optimization-based support vector machine for landslide susceptibility mapping. Sustainability, 11(22), 6323. https://doi.org/10.3390/su11226323.

Pham Binh Thai, Tien Bui D., Indra P., Dholakia M., 2015. Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS-based statistical approach of frequency ratio method. Int J Eng Res Technol, 4(11), 338–344. https://doi.org/10.17577/ijertv4is110285.

Pham V.T., Le H.L., Tran T.N., Nguyen Q.P., Phan T.T., Dinh T.Q., Dao M.D., Nguyen C.L., Nguyen H.C., 2023. Mechanism and numerical simulation of a rapid deep-seated landslide in Van Hoi reservoir, Vietnam. Vietnam J. Earth Sci., 45(3), 357–373. https://doi.org/10.15625/2615-9783/18539.

Phuoc V.N.D., Binh N.Q., Hung P.D., Long D.V., Cong N.C., 2019. Studies on the causes of landslides for mountainous regions in central region of Vietnam. The University of Danang, Journal of Science and Technology, 17, 29–32. https://doi.org/10.31130/jst-ud2019-170e.

Prasad P., Loveson V.J., Das B., Kotha M., 2022. Novel ensemble machine learning models in flood susceptibility mapping. Geocarto International, 37(16), 4571–4593. https://doi.org/10.1080/10106049.2021.1892209.

Rabby Y.W., Hossain M.B., Abedin J., 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. https://doi.org/10.1080/10106049.2020.1864026.

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

Sahin E.K., 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(7), 1308. https://doi.org/10.1007/s42452-020-3060-1.

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

The Vietnam Institutes of Geosciences and Mineral Resources, 2020. Investigate, evaluate and landslide zonation in the mountainous region of Vietnam.

Tran T., Neefjes K., 2015. Viet Nam Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX Vietnam).

Varnes D.J., 1984. Landslide hazard zonation: a review of principles and practice. Commission on landslides of the IAEG. Natural Hazards, 3, 61p.

Zhang W., Li H., Han L., Chen L., Wang L., 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. https://doi.org/10.1016/j.jrmge.2021.12.011.