TECHNOLOGY IN
NATURAL DISASTER PREVENTION AND RISK REDUCTION

INTERNATIONAL CONFERENCE

HANOI, AUGUST 31st, 2022

VENUE: HANOI UNIVERSITY OF NATURAL RESOURCES AND ENVIRONMENT

Organised by

HANOI UNIVERSITY OF NATURAL RESOURCES AND ENVIRONMENT

With the support and collaboration of:

NETHERLANDS UNIVERSITIES FOUNDATION FOR INTERNATIONAL COOPERATION (NUFFIC)

DELFT UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF TWENTE

PUBLISHING HOUSE FOR SCIENCE AND TECHNOLOGY

COMPARISON BETWEEN STATISTICAL INDEX (SI) AND BAYESIAN STATISTICS FOR LANDSLIDE SUSCEPTIBILITY MAPPING AT NGUYEN BINH COUNTY, CAO BANG PROVINCE

Nguyen Quoc Phi[1], Nguyen Thi Hoa[1], Pham Dinh Manh[1]
Nguyen Quang Minh[1], Phi Truong Thanh[2]
[1]University of Mining and Geology, Vietnam
[2]Hanoi University of Natural Resources and Environment, Vietnam
Corresponding author. Email: nguyenquocphi@gmail.com

Abstract

The fragile environment along the roads and mining sites provides advantageous conditions for the development of landslides in Nguyen Binh county of Cao Bang province. Landslide susceptibility mapping is of great importance for geo-environment control and restoration planning. In this paper, a total of 235 landslides in Nguyen Binh county, Cao Bang province were collected through historical landslide inventory and remote sensing identification. Eight lithological, geomorphic and hydrological predictive factors, mainly derived from geological map, remote sensing images and digital elevation model (DEM), were prepared initially for landslide susceptibility assessment. Predictive capability of these factors was evaluated by using the methods of Statistical Index (SI) and Bayesian statistics. The rank probability score was used to validate and compare the comprehensive predictive capabilities of the two models. Results showed that SI model achieved higher prediction capability, proving its advantage of solving nonlinear and complex problems. Comparing the estimated landslide susceptibility map with the ground-truth one, the high-prone area tends to be located in the middle area with multiple fault distributions and the steeply sloped hill.

Keywords: Statistical Index; Bayesian statistics; Landslide susceptibility mapping; Nguyen Binh; Cao Bang.

1. Introduction

Landslides have become one of the most critical natural hazards causing loss of lives, economic losses and environmental disruptions in Vietnam. Since 2000, landslide disasters have caused about 480 death or missing; the estimated economic loss is of 1.5b USD. Although earthquake and rainfall are the main triggering factors for landslide occurrences globally, terrain and geological conditions, and other human activities are factors that strongly influence to these occurrences (Nguyen et al., 2018, Nguyen, 2022). Consequently, landslides are complex processes that are still difficult to predict.

Various advanced methods and techniques for prediction landslides have been proposed to use for landslide assessment, however, prediction powers of resulting models have still a critical point. Though the technology and methods of landslide prediction have become more and more advanced and diversified, a few attempts have been made to predict landslides in Cao Bang province of Vietnam (Pham et al., 2020; Nguyen, 2022). For this reason, regional landslide susceptibility mapping is very important in predicting the landslide and preventing their damage in Cao Bang, especially in Nguyen Binh county.

The main purpose of this article was to analyse the landslide susceptibility using two models, namely Statistical Index (SI) and Bayesian statistics. The environmental factors were converted to 10 × 10 m grid. Through a detailed literature review (Corominas et al., 2014, Bui et al., 2016; Nguyen et al., 2018) and compared with the collected data of Nguyen Binh county, we chose the following as key factors: Elevation, slope angle, Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), lithology, distance to faults, NDVI and rainfall. The SI and Bayesian models were used to calculate the landslide susceptibility index (LSI); using each value, and then the landslide susceptibility map was acquired. At last, we used the Area Under the Curve (AUC) of Relative Operating Characteristics (ROC) to check the accuracy of the models. In the end, the comparison and evaluation of the two models were carried out to choose the better one. These landslide susceptibility maps were the first applied in the Nguyen Binh county, so it was very important to the government and local people.

2. Study area and data used

2.1. Study area

Nguyen Binh county is located in the southwestern part of the Cao Bang province in the northern region of Vietnam. It covers an area of about 840.6 km^2 between longitudes 105°42’15”E and 106°11’00”E, and between latitudes 22°29’15”N and 22°48’35”N. The elevation ranges from 140 to 1950 m above the sea level and gradually decreases from north-west to east direction.

The study area belongs to a humid subtropical monsoon region, with hot, rainy and dry seasons. The rainy season is normally from June to October with a high frequency of intense rainfall in July and August. The annual average precipitation of study area is about 1,500 - 2,000 mm and in the rainy season, the average rainfall is around 250 - 270 mm per month. According to the statistic from government report, the frequency and intensity of the rainfall are concentrated over a short period that triggers most of the landslides in the study area. On 14 and 15 July 2019, many rainfall induced landslides occurred in Cao Bang province, which caused the destruction of many houses and roads and the death of villagers, 918 houses and 1.000 ha of rice field were under the flood water. According to the Cao Bang Meteorological Bureau, between 14 and 16 July, Nguyen Binh county experienced three periods of heavy rain shortly before the destructive landslides and flows. The cumulative rainfall at this site reached 189.1 mm on 15 - 16 July and 98 mm on 17 - 18 July.

2.2. Data

A total of 235 landslides were identified and investigated by field surveys, remote sensing images and combining the landslides’ historical locations recorded by the Vietnam Geological Survey projects during 2012 - 2020. The landslide boundaries were mapped and the centroid of these landslides was chosen to map landslide susceptibility. For landslide modelling, these landslides were randomly split into two parts in a 70/30 ratio: Part 1 includes 165 landslides used for model training, whereas the remaining 70 landslides were used for model validation. Figure below shows the distribution of these landslide locations. In this study, the non-landslide locations were selected randomly using *iGeoHazards* software, the number of non-landslide points is equal to that of landslide points and randomly split into two parts (70/30) for verifying models.

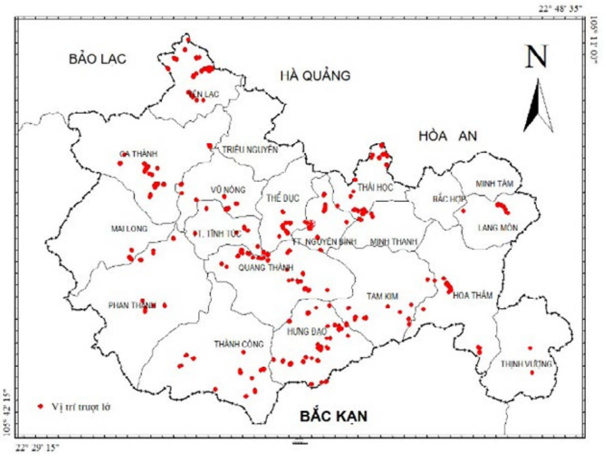
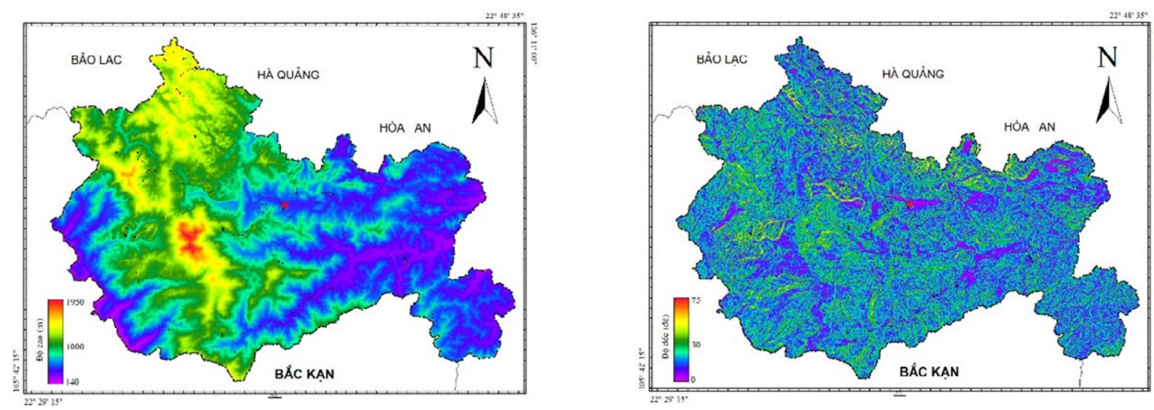


Figure 1: Location of landslides in Nguyen Binh county

The number of conditioning factors to be used for landslide susceptibility assessments is also debated and varies from few factors (Pradhan et al., 2010; Akgun, 2012) to a larger number of factors (Catani et al., 2013; Meinhardt et al., 2015). However, a landslide model with many factors does not guarantee a high prediction result (Pradhan and Lee, 2010). In some cases, the quality of the resulting models is reduced with the inclusion of noise factors (Bui et al., 2015).

Based on our analysis of these landslide inventories and study area, a total of 8 landslide-related factors were selected: Elevation, slope angle, Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), lithology, distance to faults, NDVI and rainfall. Since slope stability is directly related by terrain types and geomorphologic processes, a DEM for the study area was generated first based on topographic maps at scale of 1:10,000. Then slope angle, TWI, TRI were extracted. TWI and TRI were also selected as factors that influencing instability of slopes because these factors are correlated to hydrogeological conditions such as the surface runoff and infiltration (Lanni et al., 2012).



a) b)

Figure 2: Maps of a) Elevation and b) Slope angle

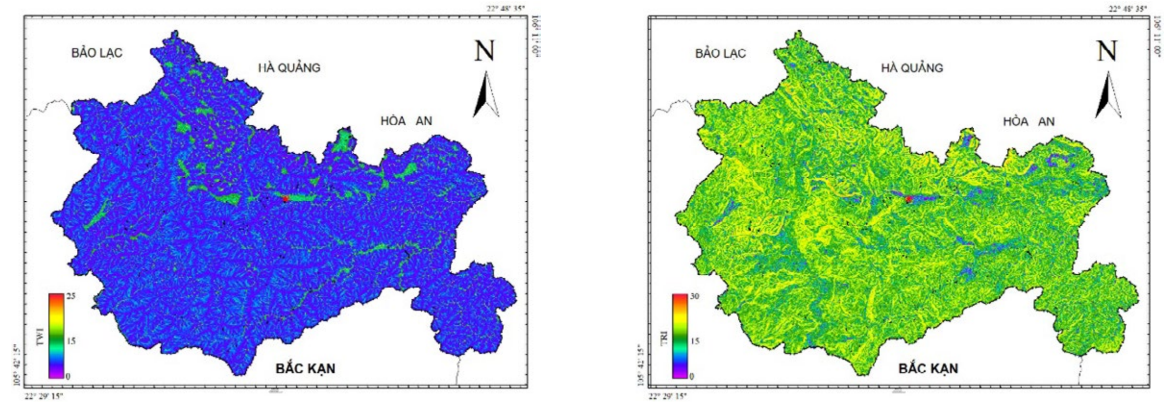
TWI (Moore and Burch, 1986) is defined as:

$$TWI = \ln(A_s / \tan\beta) \tag{1}$$

where A_S represents the catchment area (m^2/m), β is the slope of each grid, respectively.

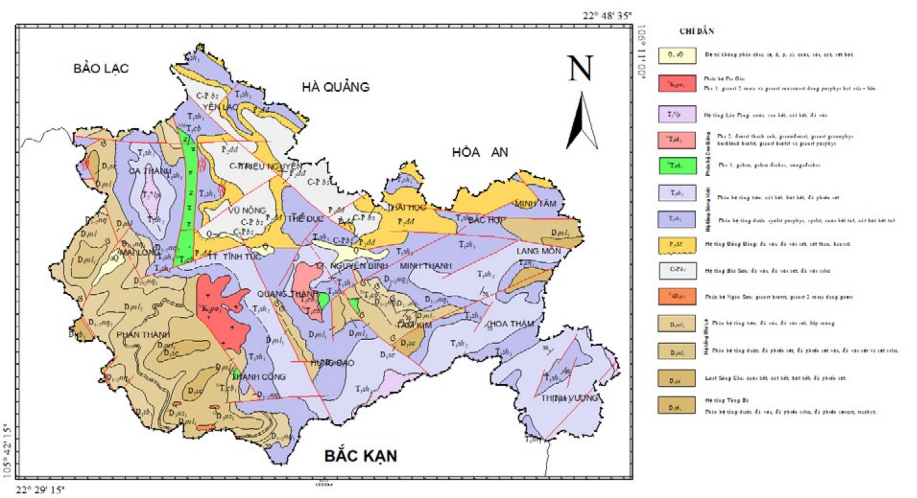
The TRI is computed for each grid cell of a DEM by calculating the sum change in elevation between the central grid cell and the mean of an 8-cell neighborhood of surrounding cells (Riley et al., 1999). The equation is:

$$TRI = \sqrt{\left|(\max DEM)^2 - (\min DEM)^2\right|} \quad (2)$$

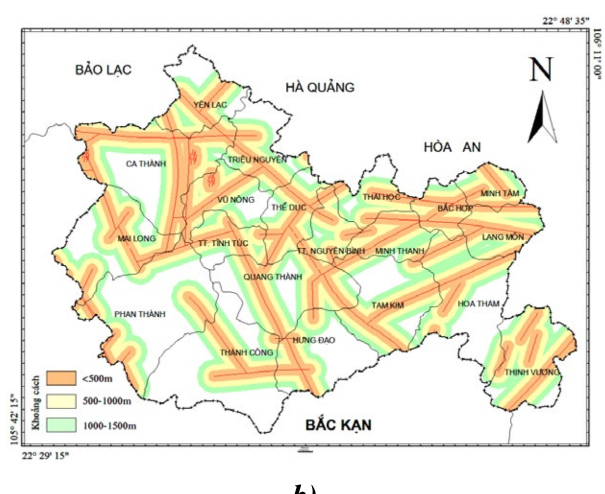


a) ***b)***

Figure 3: Maps of a) TWI and b) TRI



a)



b)

Figure 4: Maps of a) Geology and b) Distance to faults

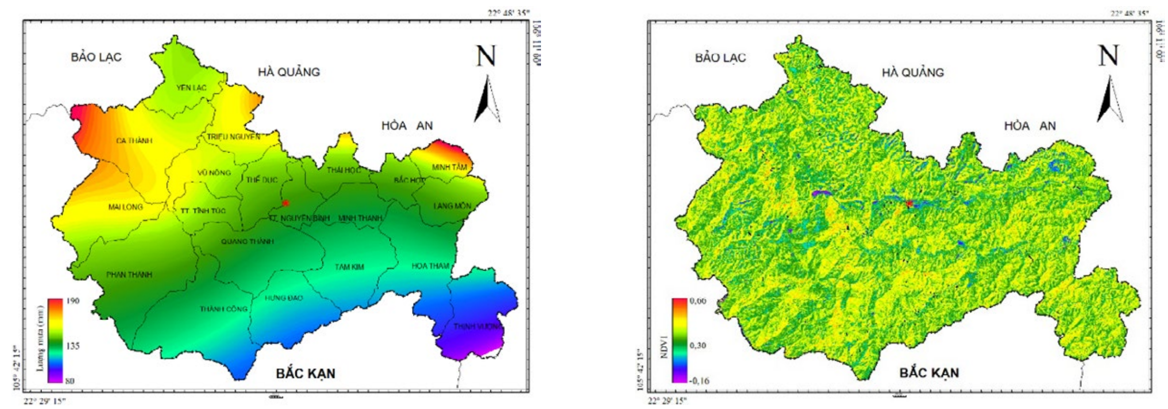
Lithology relates to geomechanical properties that influences weathering crust in underlying bedrocks (Dai and Lee, 2002); Therefore, lithology is considered to be one of the most important factors for landslide susceptibility assessment. The lithologic map of the study area was constructed based on the geological maps at the scale of 1:50,000 (Nguyen Binh - Phu Thong and Cao Bang - Dong Khe sheets) and partly at the scale of 1:200,000 (Bao Lac and Chinh Si - Long Tan sheets) collected from the General Department of Geology and Minerals of Vietnam. Seven groups were defined based on the similarity of lithologies. Fault is also a factor that influences strength parameters of bedrocks and therefore closely relates to slope instability (Dai and Lee, 2002). The distance to faults map in this study was constructed by buffering fault lines.

NDVI is an index for vegetation density (Peduzzi, 2010) that indirectly influences to landslides, so NDVI was also used in this analysis. NDVI values were estimated from the Landsat 8 OLI image using the common equation as follows:

$$NDVI = (R - R)(R + R) \tag{3}$$

where *NIR* is the reflectance of the Earth's surface in the near infrared channel (0.725 - 1.1 µm) and *R* is the red channel (0.5 - 0.68 µm).

Regarding rainfall, spatial distribution of rainfall could influence saturation of soils, so rainfall was used in this analysis. Maximum daily rainfall data were extracted and the inverse distance weighted method was used to construct the rainfall map of the study area.



a) ***b)***

Figure 5: Maps of a) Maximum daily rainfall and b) NDVI

3. Landslide susceptibility mapping

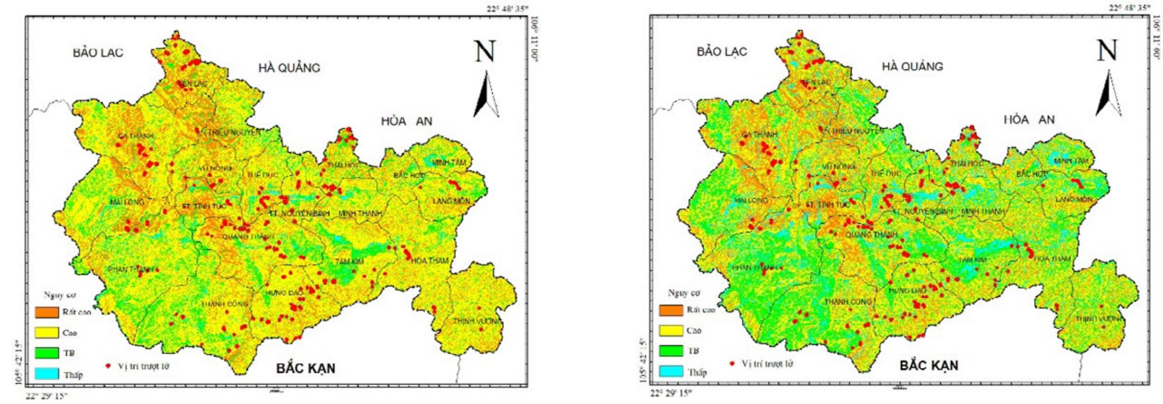
3.1. Application of SI and Bayesian models

The *SI* model is relatively a simple and understandable probabilistic model. It is defined as the ratio of the area where landslides occurred in the total study area and is also the ratio of the probabilities of a landslide occurrence to a non-occurrence for a given attribute (Solaimani et al., 2012). The *LSI* was calculated by a summation of each factor value, defined as:

$$LSI = \sum_{i=1}^{n} I_i \tag{4}$$

where *I* is the statistical index (*SI*) or contrast value (*C*) from Bayesian model of each factor's class and *LSI* is the landslide susceptibility index.

The study area contains 8,386,334 pixels, converted into point type and mapped by *PG-Steamer* platform. The *SI* and *C* value from Bayesian model of all the thematic layers used in this study was calculated in *iGeoHazards* and MS Excel. A *SI* or *C* value of 1 is an average value for the area landslides occurring in the total area, a value higher than 1 means a higher correlation which indicates a high probability of landslide occurrence, and a value lower than 1 means lower correlation which indicates a low probability of landslide occurrence (Ozdemir and Altural, 2013). Landslide susceptibility maps were constructed using the *LSI* values. The calculated *LSI* values for *SI* model of the study area range from 2.24 to 16.75 with the average of 12.33 and standard deviation of 2.02. The calculated *LSI* values for Bayesian model of the study area range from -11.61 to 8.66 with the average of 3.55 and standard deviation of 2.75. Obviously, larger *LSI* values indicate a higher susceptibility for landslides. The *LSI* values were classified into four classes for low, moderate, high and very high, respectively.



a) ***b)***

Figure 6: Landslide susceptibility maps of a) SI and b) Bayesian models

3.2. Accuracy evaluation

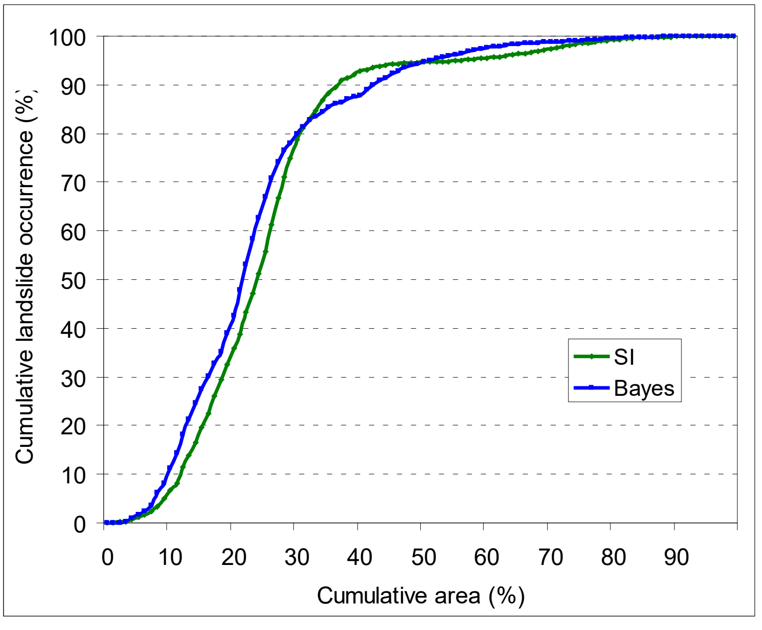


Figure 7: AUC representing success rate of the models

The landslide susceptibility maps derived by two models were tested using the training data sets that were used for model building process as well as from those landslide data (validating data sets) that were not used in these models building process. Spatial effectiveness of these susceptibility maps was checked by AUC. The rate curves were created, and the rate explains how well the model and controlling factors predict the landslide. The model with the highest

AUC is considered to be the best. To obtain the success rate curve and predictive rate curve for landslide susceptibility maps, the calculated index values of all cells in the maps were sorted in descending order. Then the cumulative area percentages of ordered index values were categorized into 100 classes with 1 % cumulative intervals as the horizontal axis. The cumulative percentage of landslide numbers corresponding to *LSI* values range as the longitudinal axis, and classified LSI maps were prepared by *iGeoHazards* and *MapInfo* packages.

The success rate results were obtained by comparing the landslide training data with the susceptibility maps AUC plot assessment results showed that the AUC values were 0.8034 and 0.7912 for *SI* and Bayesian models, meaning the overall accuracy was 80.34 % and 79.12 %, respectively. From the results of the AUC evaluation, it is seen that the map produced by *SI* model exhibited the best result for landslide susceptibility mapping in the study area.

4. Discussion and conclusion

The study indicated that the two models are suitable in landslide susceptibility mapping. These models were validated and compared using known landslide locations and the receiver operating characteristics curve. The result shows that all the models perform well on both the training and validation data. The area under the curve showed that the goodness-of-fit with the training data is 80.34 % and 79.12 % for *SI* and Bayesian model, respectively. The map produced by *SI* model exhibited the best result for landslide susceptibility mapping in Nguyen Binh county of Cao Bang province. The superior performance of the *SI* model may be due to the fact that it does not need any assumptions about the distribution of variables, nor assume a linear model, and it can select the most conditioning factors on landslide occurrence.

Landslide susceptibility maps from this study provide fundamental and essential information of the causes and effective factors on landslide occurrence of the area. These maps can be effective in hazard management and its mitigation measures. Also, it is worth mentioning that the similar method can be used elsewhere where the same geological and topographical feature prevails.

REFERENCES

[1]. Akgun A (2012). *A comparison of landslide susceptibility maps produced by logistic regression, multi - criteria decision and likelihood ratio methods: A case study at İzmir, Turkey*. Landslides 9, 93 - 106.

[2]. Bui T. D., Pradhan B., Revhaug I., Nguyen D. B., Pham H. V., Bui Q. N (2015). *A novel hybrid evidential belief function - based fuzzy logic model in spatial prediction of rainfall - induced shallow landslides in the Lang Son city area (Vietnam).* Geomatics. Nat Hazards Risk 6, 243 - 271.

[3]. Bui T. D., Nguyen Q.P., Hoang N. D., Klempe H (2016). *A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS*. Landslides. Doi: 10.1007/s10346-016-0708-4.

[4]. Catani F., Lagomarsino D., Segoni S., Tofani V (2013). *Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues*. Nat Hazards Earth Syst Sci. 13, 2815 - 2831.

[5]. Corominas J., van Westen C., Frattini P., Cascini L., Malet J. P., Fotopoulou S., Catani F., Van Den Eeckhaut M., Mavrouli O., Agliardi F., Pitilakis K., Winter M.G., Pastor M., Ferlisi S., Tofani V., Hervás J., Smith J. T (2014). *Recommendations for the quantitative analysis of landslide risk*. Bulletin of Engineering Geology and the Environment. 73, 209 - 263.

[6]. Dai F. C., Lee C. F (2002). *Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong*. Geomorphology 42, 213 - 228.

[7]. Lanni C., Borga M., Rigon R., Tarolli P (2012). *Modelling shallow landslide susceptibility by means of a subsurface flow path connectivity index and estimates of soil depth spatial distribution*. Hydrol Earth Syst Sci. 16, 3959 - 3971.

[8]. Meinhardt M., Fink M., Tünschel H (2015). *Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: comparison of a new method to calculate weighting factors by means of bivariate statistics*. Geomorphology. 234, 80 - 97.

[9]. Quoc Phi Nguyen, Du Duong Bui, Sang Gi Hwang, Khac Uan Do, Thi Hoa Nguyen (2018). *Rainfall - triggered landslide and debris flow hazard assessment using data mining techniques: A comparison of decision trees, artificial neural network and support vector machines*. Proceedings of the 2018 Vietnam Water Cooperation Initiative (VACI 2018) - Highlights. Science and Technics Publishing House, Hanoi, Vietnam, p.138 - 141.

[10]. Nguyen Quoc Phi (2022). *Assessment of landscape disturbance in Trung Khanh area, Cao Bang province using different decision tree (C4.5, CART and LMT) models*. International Conference on Technology in Natural Disaster Prevention and Risk Reduction. Hanoi University of Natural Resources and Environment, Vietnam.

[11]. Ozdemir A., Altural T (2013). *A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey*. J Asian Earth Sci. 64, 180 - 197.

[12]. Pham Van Tien, Le Hong Luong, Tran Thanh Nhan, Do Minh Duc, Dinh Thi Quynh, Nguyen Chau Lan, Nguyen Quoc Phi, Do Canh Hao, Nguyen Huu Ha, Dang Thi Thuy and Vu Ba Thao (2020). *Secondary processes associated with landslides in Vietnam*. Proceedings of the International Conference on Innovations for Sustainable and Responsible Mining (ISRM 2020). Lecture Notes in Civil Engineering 108, pp. 192 - 209.

[13]. Pradhan B., Sezer E. A., Gokceoglu C., Buchroithner M. F (2010). *Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia)*. IEEE Trans Geosci Remote Sens. 48, 4164 - 4177.

[14]. Pradhan B and Lee S (2010). *Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling*. Environ Model Softw. 25, 747 - 759.

[15]. Riley S. J., De Gloria S. D., Elliot R (1999). *A terrain ruggedness index that quantifies topographic heterogeneity*. Intermt. J. Sci. 5, 23 - 27.

[16]. Solaimani K., Mousavi S.Z., Kavian A (2012). *Landslide susceptibility mapping based on frequency ratio and logistic regression models*. Arab J Geosci. 6, 2557 - 2569.