LANDSLIDE SUSCEPTIBILITY MAPPING USING LOGISTIC REGRESSION MODEL: A CASE STUDIES IN THE VAN YEN, YEN BAI PROVINCE

Xuan Quang Truong[1], **Xuan Luan Truong**[3], **Thi Thanh Thuy Pham**[1],

Thi Thu Ha Le [1], **Ngoc Hoan Nguyen**[1], **Chi Cong Nguyen**[2], **Thi Hai Yen Phi**[1]

[1]Faculty of Information Technology, Hanoi University of Natural Resources and Environment
[2]Faculty Geology, Hanoi University of Natural Resources and Environment
41A, Phu Dien Road, Bac Tu Liem, Hanoi, Vietnam
Email: txquang@hunre.edu.vn
[3]Faculty of Information Technology, Hanoi University of Ming and Geology
18 Pho Vien, Duc Thang Ward, Tu Liem District, Hanoi , Vietnam
Email: truongxuanquang@gmail.com

ABSTRACT

Landslide is one of the most complex natural phenomena and is quite common in Vietnam. Recently, Machine learning models are applying in order to improve accuracy of creating landslide susceptibility map and combining into the system as the prediction model for early warning and forecasting. The logistic regression model is widely used for prediction analysis in a variety of applications. Landslide susceptibility mapping, prediction analysis is important to predict the areas which have high potential for landslide occurrence in the future. The study area is in Van-Yen in the northern province of Yen Bai. Landslide inventory maps (302 landslides) were compiled by reference to historical reports and aerial photographs. All landslides were randomly separated into two data sets: 70% were used to establish the models (training data sets) and the 30% for validation (validation data sets). 11 factors were considered as conditioning factors related to landslide and divided into six categories (topography, hydrology, soil, geology, forest and road. The accuracy of the results was evaluated by using ROC. The area under the curve (AUC) for the logistic regression model was 0.906.

1. INTRODUCTION

Landslides are very common, occurs in a variety of form in the mountain areas of Vietnam. Landside happens due to many reasons such as poor condition of soil together with long-term rainfall or mud and/or rocks downhill due to gravity. In Vietnam, northern areas and central region are the most affected by landslide. From 2001 to 2010, natural disasters such as floods, flash floods, landslides, droughts, salinity and other types of natural disasters cause 9.500 deaths and missing, causing losses of about 1.5% of annual GDP. In 2020 Natural disasters cost Vietnam 1.6 bln including landslide, Vietnam was hit by 14 storms that triggered heavy flooding and 130 people have died and missing cause of landslides based on Vietnam Disaster Management Authority report (VDMA 2021). The rainfall-triggered landslide is especially exacerbated in countries that are located in storm centers of the world, including Vietnam (Truong et al, 2018).

The landslide occurrence is a complex process which is a combination of multiple interacting factors. In order to minimize the damage due to landslides, machine learning methods are suitable when a direct mathematical relationship cannot be established between cause and effect (Shano et al, 2020). Recently, methods for landslide susceptibility assessment have relied on statistical-based approaches and data-driven approaches. Such as a logistic regression (Pourghasemi et al. 2013) and (Tsangaratos and Ilia 2016); artificial neural networks (Pradhan and Lee 2010) and (Bhardwaj and Venkatachalam 2014); support vector machines (Peng et al. 2014) and (Chen et al. 2018); Naive Bayes (Tsangaratos & Ilia 2016); decision trees models (Pradhan 2013); and random forest (Catani et al. 2013) and (Truong, X.Q et al. 2018) or hybrid machine learning approach (Truong, X.L et al. 2018).

The main purpose of this study is to use logistic regression analysis for landslide susceptibility mapping of Van Yen district, Yen Bai province. Remote sensing data along with data collection from the field in study area have been used to generate the landslide susceptibility areas for the Ven Yen district. The collection of landslide inventories was compiled from Geo-Web portal of the Vietnam Institute of Geosciences and Mineral Resources VIGMR (VIGMR, 2020), 302 landslides have been identified. The factors such as slope, aspect, geology, distance from fault, distance from river, distance from road, normalized difference vegetation index (NDVI) and stream power index (SPI), Profile Curvature, Plan Curvature have been used to facilitate the quantification of landslide.

2. STUDY AREA

The study area is the Van Yen district (Figure 1) of the Yen Bai province. Yen Bai province is located on the northern part of the Vietnam. Yen Bai province is between in the latitude 21^{o} 24'N and 22^{o}16' N and between 103^{o}56'E and 105^{o}03'E. The study area is covered with an 1.391,54 km^{2}. The lowest area is about 20m above sea level, the high mountain area has an elevation from 300 - 1.700m, mainly concentrated in the north west. 70% of study area is occupied by slope angles, those are higher than 15^{o}. The average temperature ranges between 18 and 20^{o}C and rainfall varies between 1500 mm and 2200 mm/yr, average annual humidity is between 83%-87%. There are two main seasons: a rainy season spanning from April to October, and a dry season starting from November to March. The study area is located in an active tectonic region with the relatively fast movement of the Red River fault zone that results in continuously landslide occurrences over the years (Truong, X.Q et al, 2018). The history of landslides that have caused damages to life, property, and infrastructures, they were investigated by the VIGMR in 2013 and 2017. These landslides mainly occur along national road and at various topographic types after heavy rain or/and human activities.

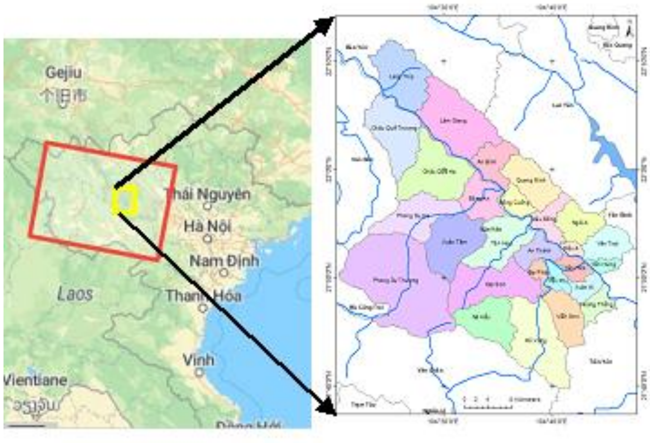

Figure 1: Study area

3. LANDSLIDE CONDITIONING FACTORS

In this paper, 11 factors were studied based on the data available and general geo-environment of the study area as shown in (Table 1). The condition factors are slope, aspect, altitude, landcover, lithology, soil type, distance to fault, distance to river, rainfall. A digital elevation model (DEM) from ALOS PALSAR data with 12.5 m resolution was used to process slope, aspect, elevation, topographic wetness index (TWI), stream power index (SPI), plan curvature and profile curvature maps using QGIS 2.16.3. Normalized Difference Vegetation Index (NDVI) was created from sentitnel-2 with 12.5m resolution. The geology and fault maps at scale 1:10.000 was provided by the Ministry of Natural Resource. The river and road maps were used from the Vietnam national atlas with scale 1:10.000. The categories of each layers were defined in the (Table 1).

Table 1. Landslide conditioning factors and their classes

Type	Classes	Number of class pixels	Landslide points	SI
distance to faults (m)	0-50	395353	90	1.076
	50-100	318816	54	1.186
	100-200	558240	73	0.927
	200-500	1546065	114	0.354
	>500	6115621	161	-0.675
Plan Curvature	-1	3869564	109	-0.182
	0	992414	52	0.438
	1	4072097	141	0.024
Profile Curvature	-1	4296900	147	0.011
	0	64641	33	0.412
	1	3990763	122	-0.100
NDVI	≤0	12207	25	5.0308
	0 – 0.2	1246475	160	9.8429
	0.2 – 0.4	638461	293	6.8225
	0.4 – 0.6	400772	413	9.0381
	0.6 – 0.8	184817	326	9.7929
	≥0.8	106719	429	5.0308
River	0-25	779239	42	0.466
	25-50	757665	48	0.628
	50-100	1411858	96	0.698
	100-200	2189294	55	-0.296
	200-300	1391479	32	-0.385
	≥300	2404560	29	-1.032
SPI	-13 to -5	3265726	164	0.395
	-5 - 0	281298	15	0.455
	0 - 5	264804	6	-0.400
	5 - 10	361359	13	0.062
	10 - 14	4760888	104	-0.436
Slope	0-5º	624239	54	0.939
	5-10º	885595	71	0.863
	10-15º	1143726	57	0.388
	15-20º	1538329	46	-0.122
	20-25º	1551737	36	-0.376
	25-30º	1384137	16	-1.073
	30-40º	1400540	21	-0.812
	40-50º	340851	1	-2.444
	≥50 º	64921	0	0.939
Aspect	-1	74552	0	0
	35-72 and 320-360	627791	20	-0.091
	72-113	1692255	72	0.197
	113-156	1117460	44	0.120
	156-198	1035447	43	0.173
	198-238	1131021	48	0.194

	238-279	1264409	28	-0.455
	279-320	1026233	28	-0.246
Geology	29 classes	values: -2.39 to 3.95		
Distance to Road	05 classes	values: -2.73 to 1.40		
TWI	04 classes	values: -0.26 to 0.19		

The statistical index (SI) method is applied for landslide susceptibility analyses in this study and was proposed by (Bui 2011) and Van Westen (1997). In this method, the weight for a parameter class, such as a slope class or NDVI class, is defined as the natural logarithm of the landslide density in each class divided by the landslide density in the all map. This method is based upon the formula given by as follows:

$$w_i = \ln\left(\frac{Dens\_class}{Dens\_map}\right) \tag{1}$$

Where w_i is the weight, *Dens_class* is the landslide density within the parameter class, and *Dens_map* is the landslide density within the entire map, the results of the calculation is showed in (Table 1) abd (Figure 2).

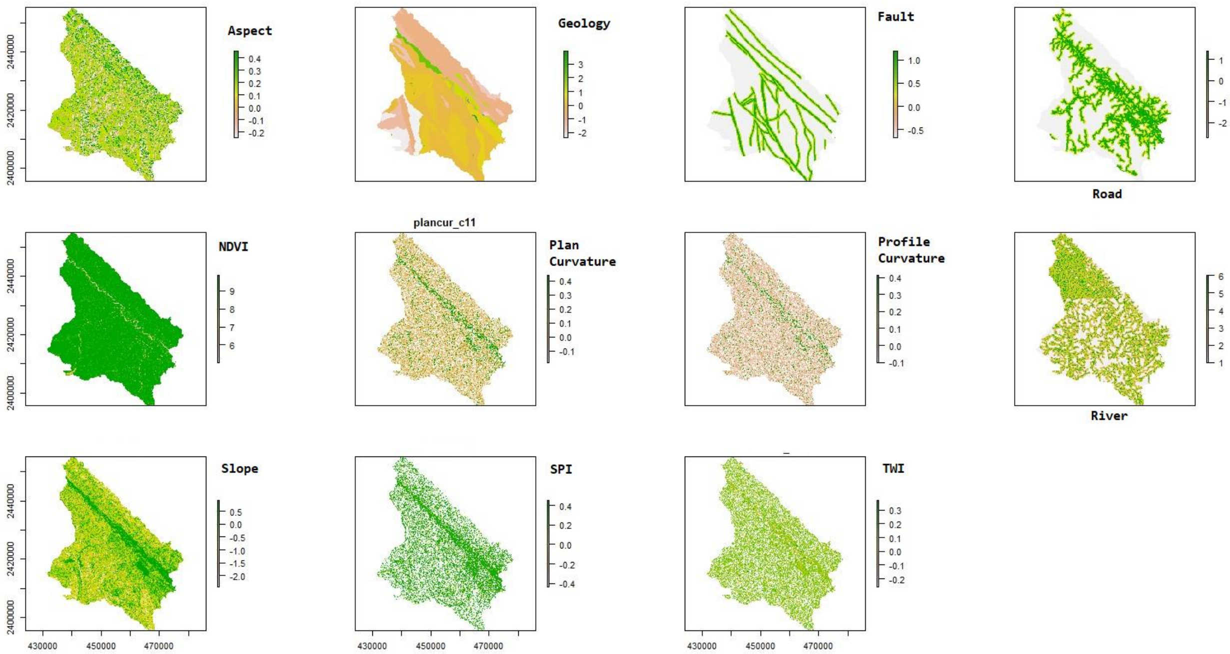

Figure 2: statistical index of factors based on Table 1

4. LANDSLIDE SUSCEPTIBILITY MAPPING USING LOGISTIC REGRESSION

Logistic regression is a mathematical modeling that can be applied for examination the presence or absence of outcome based on the values of a set of predictor variables (Lee 2005). The logistic regression function *f(z)*, which is defined as:

$$f(z) = \frac{1}{1+e^{-z}}; z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n \tag{2}$$

where z is a linear sum of a constant α and products of independent variables $x_i (i = 0,1,2,...,n)$ and their corresponding coefficients β_i $(i = 0,1,2,...,n)$ are the coefficients.

For building landslide model, the landslide inventory map was randomly divided into a training dataset of 70% (210 landslides with 2,781 pixels) for building the landslide models and a validation dataset 30% (92 landslides with 716 pixels). Using the logistic regression coefficients, landslide susceptibility index was calculated and the landslide susceptibility map was obtained (Figure 3). The linear sum of the constant and the product of the independent variables and their corresponding coefficients are given as in the following equation:

$$z = -1.81 - 1.19 * aspect + 0.69 * geology + 0.44 * fault + 0.96 * road + 0.2367 * NDVI + 0.98 * river - 1.28 * curvature\_plan + 1.49 * curvature\_profile + 0.29 * slope - 0.88 * SPI + 0.88 * TWI$$

5. RESULT AND DISCUSSION

In this paper, we have evaluated the logistic regression for landslide susceptibility assessment for Van Yen district of the Yen Bai province. Data processing for each factor was analyzed based on QGIS 2.16.3. The landslide inventory (302 landslide locations) and 11 landslide conditioning factors (slope angle, slope aspect, plan curvature, profile curvature, topographic wetness index, stream power index, NDVI, distance to roads, distance to rivers, distance to faults and geology) were used for building the logistic regression model.

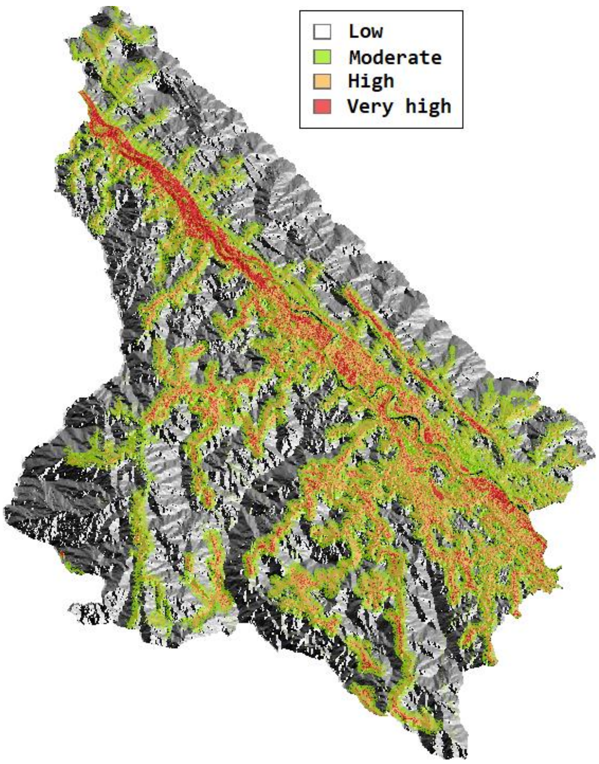

Figure 3: Landslide susceptibility mapping using Logistic Regression

The ROC curve can be used to provide predictions of the performance of the model (sensitivity vs. specificity). The ROC curve is a two-dimension graph showing true-positive rate (vertical axis) and false-positive (horizontal axis). The area under the ROC curve (AUC), which is the summarized information of the plot, can be used to estimate the validity of the model: accuracy or the overall quality of a model (Hosmer and Lemeshow 2000). Areas under the curves (AUC) were calculated (Figure 4). The result shows that, the logistic regression has AUC=0.906 (Figure 4).

As the result from regression model, the landslides happen in many and densely along the road through Dong An, Chau Que Ha, Dong Cuong communes. The results are given in a medium-scale map, this will clarify the details of high and very high susceptibility area related to landslide risk.

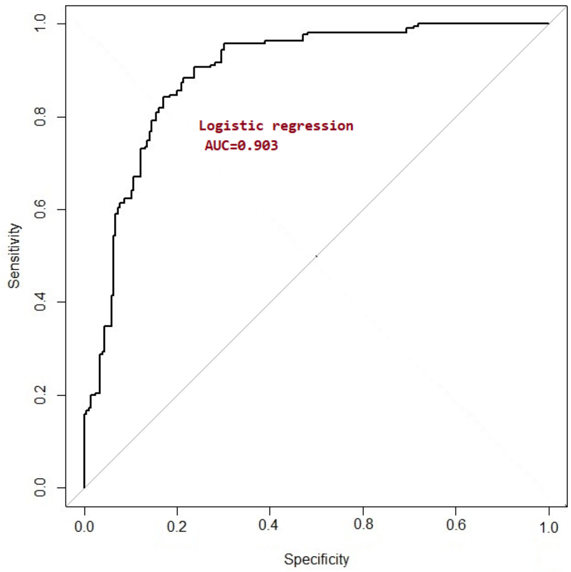

Figure 4: The receiver operating characteristic (ROC) curve

The landslide susceptibility maps of this study may be support to local administrator in order to manipulate in decision makers, engineers in slope management and land use planning. The accuracy of the logistic regression model can be improved if additional landslides or other factors such as land use land cover, soil maps are included in the analysis. This work needs a long-term period and require stakeholders as well as human resources.

6. ACKNOWLEDGMENTS

This work is supported by the Bilateral Scientific Research project between Vietnam and Italy, code: NĐT/IT/21/14 funded by Ministry of Science and Technology of Vietnam (MOST) and Funded by a grant from the Italian Ministry of Foreign Affairs and International Cooperation

REFERENCES

Bhardwaj A, Venkatachalam G. 2014. Landslide hazard evaluation using artificial neural networks and GIS. In Landslide science for a safer geoenvironment. Cham: Springer; p. 397–403.

Bui, D.T., Lofman, O., Revhaug, I. et al. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat Hazards 59, 1413 (2011). https://doi.org/10.1007/s11069-011-9844-2

Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull. Eng. Geol. Environ. 2018, 77, 647–664.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. Natural Hazards and Earth System Sciences, 13(11), 2815–2831

Lee S (2005) Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data journals. Int J Remote Sens 26(7):1477–1491. doi:10.1080/01431160412331331012.

Hosmer DW, Lemeshow S (2000) Applied logistic regression, 2nd edn. Wiley, New York.

Pradhan B, 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.

Pradhan B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine andneuro-fuzzy models in landslide susceptibility mapping using GIS. Compu Geosci. 51:350–365.

Pourghasemi, H.; Moradi, H.; Aghda, S.F. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat. Hazards 2013, 69, 749–779.

Shano, L., Raghuvanshi, T.K. & Meten, M. Landslide susceptibility evaluation and hazard zonation techniques – a review. Geoenviron Disasters 7, 18 (2020). https://doi.org/10.1186/s40677-020-00152-0

Tsangaratos P, Ilia I. (2016). Comparison of a logistic regression and Navıve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *CATENA*. 145:164–179.

Truong, X.L; Mitamura, M; Kono, Y; Raghava, V; Truong, X.Q; Do, T.H; Bui, T.D; Lee, S. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree, Appl. Sci. 2018, 8(7), 1046; https://doi.org/10.3390/app8071046.

Truong, X.Q., Truong, X.L., Dang, T.K.L Nguyen., T.D; Nguyen., D.A. Landslide susceptibility mapping using random forest model in Lao Cai Province, Vietnam, GIS-IDEAS 2018, Geoinformatics for Rural-Urban Synergy (2018)

VIGMR, Vietnam insititute of geosciences and mineral resources (2020) http://www.canhbaotruotlo.vn/ (accessed on Dec, 2020)

Van Westen CJ (1997) Statistical landslide hazard analysis. ILWIS 2.1 for Windows Application guide. ITC Publication, Enschede

VDMA, Vietnam Disaster Management Authority (2021) http://phongchongthientai.mard.gov.vn/ (accessed on 12, Aug 2021)