LANDSLIDE SUSCEPTIBILITY MODELING WITH FREQUENCY RATIO, LOGISTIC REGRESSION, ARTIFICIAL NEURAL NETWORK AND THE COMBINATION METHOD

Van-Trung Chu, Shou-Hao Chiang* and Tang-Huang Lin

Center for space and remote sensing research, National Central University, Taoyuan Taiwan
Email: gilbert@csrsr.ncu.edu.tw, thlin@csrsr.ncu.edu.tw, chuvantrung@tuaf.edu.vn

KEY WORDS: Landslide modeling, Logistic regression, Frequency ratio, Artificial neural network, model combination, Viet Nam

ABSTRACT

Several landslide models have been proposed to produce landslide susceptibility map, but no particular model has been considered as an optimal one in global scale due to the dominated varieties of landslide are different from regions. In this study, we try to integrate the advantage of each model for a better approach in landslide susceptibility mapping. Three commonly used models, namely frequency ratio (FR), logistic regression (LR) and artificial neural network (ANN), are examined to generate a combined susceptibility map in Thu Lum basin located in the mountainous area of Lai Chau Province, Viet Nam. For training and testing the models, landslide samples were selected from a landslide inventory map which was prepared by applying the change detection method based on Normalized Difference Vegetation Index (NDVI) images derived from Sentinel-2. Landslide susceptibility maps were constructed with 13 environmental factors. The performance of proposed model was assessed by using area under the receiver operation characteristic curve (AUC) and kappa coefficient (Kappa). The combined model outperforms the best results (AUC=0.953 and Kappa=0.79) when compared to the single models (AUC: 0.944, 0.929 and 0.91, and Kappa: 0.73, 0.72 and 0.65 for ANN, LR and FR, respectively) equipped with the high potential in mapping landslide susceptibility.

1. INTRODUCTION

Landslide is one of the most dangerous natural hazards cause damages to both property and human life, and the active procedure contributes to erosion and landscape evolution every year. Thus, landslide susceptibility map is considered a significant information of further research in landslide hazard assessment and prevention (Chae, Park, Catani, Simoni, & Berti, 2017; Chen, Shahabi, et al., 2018; Guzzetti, Reichenbach, Cardinali, Galli, & Ardizzone, 2005; Nsengiyumva et al., 2019; Yilmaz, 2009). Landslides susceptibility can illustrate the spatial distribution of landslide occurred potentially. Therefore, landslide susceptibility is important spatial component of landslide hazard (Guzzetti, Reichenbach, Ardizzone, Cardinali, & Galli, 2006; Nsengiyumva et al., 2019). Recently, the number of researches increased strongly focused on analyzing and assessing the sensitivity of landslide at different scales with different approaches to problems the number of quantitative and statistical model were considered single applied or gathered using to produce landslides susceptibility maps (Guzzetti, Carrara, Cardinali, & Reichenbach, 1999; Reichenbach, Rossi, Malamud, Mihir, & Guzzetti, 2018). However, a standard procedure for the production of landslide susceptibility maps does not exist. For this reason, many researchers have used different techniques. In the newest literature article review (Reichenbach et al., 2018) has statistics that the models are commonly used for landslides to mention as logistic regression analysis, neural network, support vector machine, index-based analysis, weight of evidence, frequency ratio etc. Especially, in recent a couple of decades, there are not many available and relevant researched have been done to compare and combine different modelling approaches (Reichenbach et al., 2018). And each model analysis has its own advantage and disadvantage when computing the landslide susceptibility index. We assume that, the combination model can reduce uncertainties of single model application result. So, this manuscript summarizes our result of on implementation and comparison of three quantitative model as frequency ratio, logistic regression and artificial neural network. Subsequently, we optimate the result based on the model combination approaches. The case study in Thu Lum basin (a high mountainous terrain area) belongs to the Da river system which is located in Muong Te district - Lai Chau province. (Fig.1). In 2018, here is the most impacted by landslides during summer time which were induced by heavy rainfall.

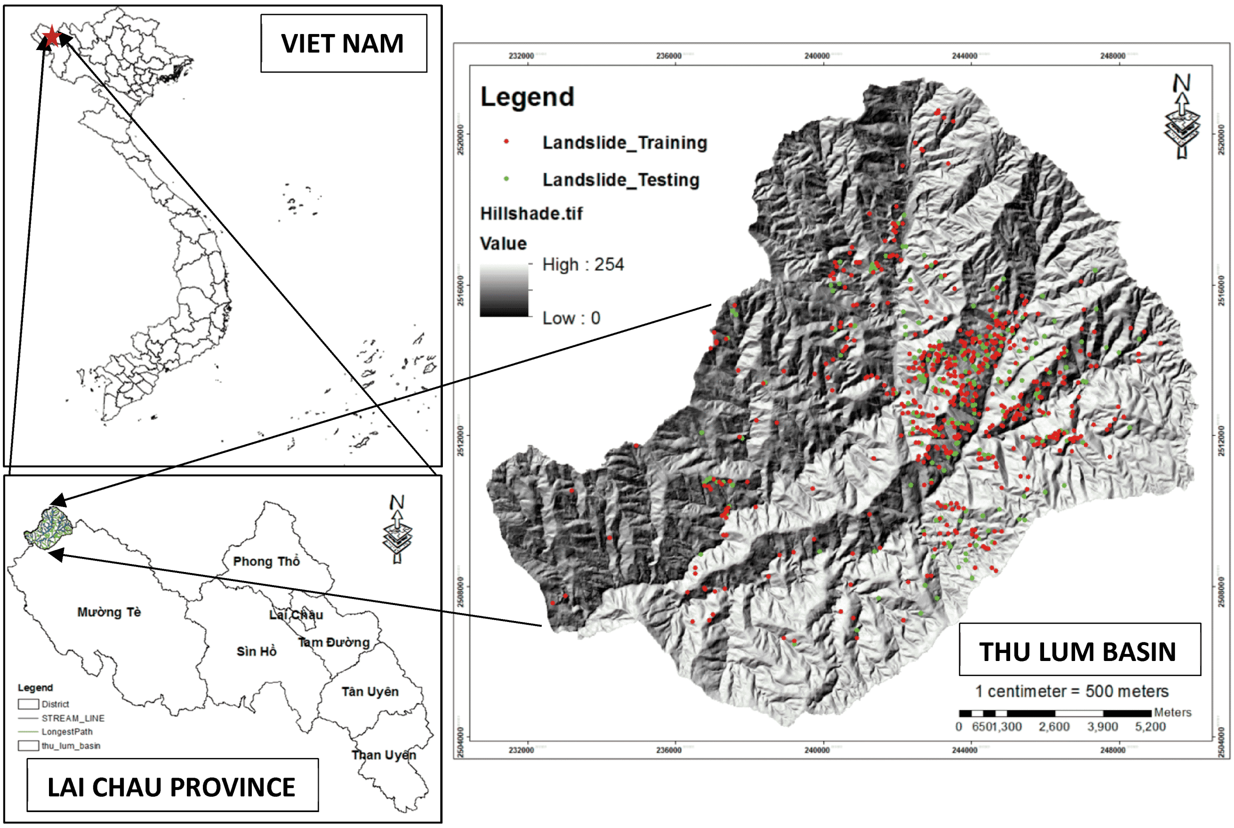

Figure 1. Location map of the study area

2. MATERIALS AND METHODS

2.1. Landslide causative factors preparation

This study used 13 landslide effected factors which are commonly used by previous literature review namely: elevation, slope angle, aspect, curvature, topographic wetness index (TWI), normalized difference vegetation index (NDVI), distance from the streams, land cover, distance from the faults, lithologic, soil types and dept of soil (Felicisimo, Cuartero, Remondo, & Quiros, 2013; Kadirhodjaev, Kadavi, Lee, & Lee, 2018; Mandal & Mandal, 2018; Park, Lee, Lee, & Lee, 2018; Pourghasemi, Gayen, Park, Lee, & Lee, 2018). Although the rainfall was trigger of the landslide occurred in this study. But the intensity of rainfall was excluded in this study because of it is almost the same value throughout the tire area.

Table 1. Data set collection and sources*

No.	Data	Type	Source	Annotation
1	County/City Administration borders.	Vector (polygon)	Department of Natural Resources and Environment	Provincial administrative map
2	Cadastral map	Vector (polygon)	Department of Natural Resources and Environment	Detailed surveying ratio 1/1000 (Ground surveying approach)
3	Topographical map	DEM (raster)	Department of Natural Resources and Environment	8 pieces DEM 10 m resolution
4	Satellite image (Sentinel 2) - Band 2- Blue - Band 3 – Green - Band 4 – Red - Band 8 – NIR	Images	USGS https://earthexplorer.usgs.gov/	Post event (2018.03.11): Flatform: Sentinel 2B, cloud cover: 1.42% Pre event (2018.02.21): Flatform: Sentinel 2A, cloud cover: 0%
5	Geology map: (Lithology, fault line)	Vector	Institute of Geological Sciences - Vietnam Academy of Science and Technology	Scale: 1:200.000 (Lai Chau province)

6	Soil map	Vector	Department of Natural Resources and Environment	Scale: 1:100.000 (Lai Chau province)

* All the data used UTM coordinate system (zone 48N)

2.2. Landslide inventory

According to local information provided, early August and early September 2018 took place with high-intensity rains and concentrated in the area where serious landslides occurred on a large scale. Considering the study area mostly is covered by forest so any significant changes of surface may affect vegetation index. Therefore, the NDVI (Normalized Difference Vegetation Index) change technique was employed for change detection between pre and post event (Tsai, Hwang, Chen, & Lin, 2010). Sentinel-2 images provided by USGS (images before the event observed on February 21, 2018 and the other after the event observed on November 3, 2018). There are 702 landslides detected. Because of LR model and ANN model requirement, it is necessary to do sampling the number of free landslide points, correspondingly. Whereas, to avoid uncertainty of none landslide point and still satisfied the nice distribution of none landslide sampling, we just avoided landslide area which are defined in this research and also avoided the unsafe area with high slope according to experience of local people (because, the rate of landslide located in that area consumed high value). Finally, there are 702 landslide point and 700 none landslide points selected for model training and model testing with randomly separate following rate of 70% and 30 %, respectively (fig.1).

2.3. Methods

**** Pre-processing***

The variance inflation factor (VIF) and tolerance (TOL) methods should be used to check for multicollinearity of the conditioning factors (Bui, Lofman, Revhaug, & Dick, 2011; Chen, Sun, & Han, 2019; Youssef, Pourghasemi, Pourtaghi, & Al-Katheeri, 2016; Zhang, Han, Chen, & Shahabi, 2018), VIF and tolerances used to check the multicollinearity magnitude. Based on experimental research figured out that, when the VIF > 10, tolerance < 0.1 (O'Brien, 2007) it means available multicollinearity with that data so we must exclude that factors.

Table 2. Multicollinearity analysis of landslide conditioning factor

Conditioning factor	Tolerance	VIF
NDVI	0.665	1.504
Aspect	0.894	1.118
Curvature	0.749	1.335
Depth of soil	0.550	1.817
Distance from roads	0.665	1.504
Distance from faults	0.800	1.249
Distance from streams	0.817	1.224
Elevation	0.495	2.019
Landcover	0.673	1.486
Lithology	0.554	1.807
Slope	0.751	1.331
Soil	0.587	1.703
TWI	0.633	1.580

The table 2 shows all of the conditioning factor are satisfied to input the model because of there aren't multicollinearity problem in that dataset.

**** Frequency ratio (FR) model***

The relationships between spatial landslides distribution and its factors are assumed to be significant and availability for predicting landslide occurred in the future, particularly, it can be strongly agreed that landslide can be occurred when the repeated rate of landslides in the past is exposed in the area with identical conditions. (Nsengiyumva et al., 2019; Pradhan & Lee, 2010; Regmi et al., 2014; Sahana & Sajjad, 2017; Yilmaz, 2009). The estimation of FR can be generated by the ratio of the area landslide occurred and the ratio of points with similar characteristics. The calculation will do for each class of each factor Eq.1 and landslide susceptibility index was computed by Eq.3. Al the computing functions use the raster calculator tool in ArcGIS to get the final susceptibility map.

$$w_{ij} = \frac{FL_{ij}}{FN_{ij}} \qquad (1)$$

Where:

- $\boldsymbol{w_{ij}}$ is the frequency ratio of class $\boldsymbol{i}$ of parameter $\boldsymbol{j}$;
- $\boldsymbol{FL_{ij}}$ is the rate of landslides points in class $\boldsymbol{i}$ of parameter $\boldsymbol{j}$;

- $\boldsymbol{FN_{ij}}$ is the rate of point of class $\boldsymbol{i}$ of parameter $\boldsymbol{j}$.

$$W'_{ij} = \sum_{j}^{n} w_{ij} \qquad (2)$$

$$LSI_{FR} = \frac{W'_{ij} - \min(W'_{ij})}{\max(W'_{ij}) - \min(W'_{ij})}(U - L) + L \quad (3)$$

Where:

- $\boldsymbol{LSI_{FR}}$= landslide susceptibility index (value range 0-1);
- $\boldsymbol{W'_{ij}}$ = The total value of frequency ratio each pixel
- $\boldsymbol{n}$ = number of parameters.
- ***U*** and ***L*** are the upper and the lower normalization boundaries, respectively (this study, the value must be range from 0 to 1 so ***U***=1 and ***L***=0).

**** Logistic Regression***

Logistic regression model (LR) is the most common statistical method. It is kind of multiple regression because of it was considered to evaluate relationships between dependent variable, and one or more categorical or numerical independent variables. LR range from 0 to 1 can be expressed as follows the Eq. 4 and Eq. 5.

$$LSI_{LR} = \frac{e^{z}}{1 + e^{z}} \qquad (4)$$

Where: LSI_{LR} is the probability of a landslide occurrence, z is the linear combination of a set of landslide related variables. Logistic regression includes fitting an equation following below:

$$z = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (5)$$

where b_0 is the constant or intercept of the model, the b_i (i = 1, 2, ..., n) is the slope coefficients of the logistic regression model, and xi (i = 1, 2, ..., n) are the independent variables. The linear model formed is then a logistic regression of presence or absence of landslides (present conditions) on the independent variables (pre-failure conditions).

In this study, there are 5 categorized factors namely: Aspect, curvature types, land cover types, soil types and lithology. Normally, logistic regression can analyze both numerical and categorical parameters. However, in this case, we have many categorized variables. So, to avoid the complex function and present of dummy variables, based on expert experiment, the categorical parameters can be converted to numerical data by using frequency ratio (Eq.1.). (Bai et al., 2010; Wang, Guo, Sawada, Lin, & Zhang, 2015). The correlations between landslides and each factor were calculated by using IBM SPSS Statistics 2.0 software.

Table 3. Statistics and accuracies of logistic regression models.

−2Log likelihood	Cox & Snell R^2	Nagelkerke R^2	Hosmer and Lemeshow test		
			Chi – square (X^2)	df	Sig
573.725	0.550	0.733	57.753	8	0.000

Table 4. Variables in the Equation of Logistic regression

factors	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I.for EXP(B)**	
							Lower	**Upper**
Dist_stream	0.018	0.014	1.606	1	0.205	1.018	0.990	1.046
NDVI	-2.362	1.113	4.501	1	0.034	0.094	0.011	0.835
Slope	0.037	0.011	10.307	1	0.001	1.037	1.014	1.061
Lithology	2.354	0.408	33.224	1	0.000	10.523	4.727	23.428
Landcover	0.926	0.235	15.503	1	0.000	2.523	1.592	4.000
Curvature	-0.789	0.526	2.244	1	0.134	0.454	0.162	1.275
Depth soil	-0.016	0.010	2.500	1	0.114	0.984	0.965	1.004
Dist_road	-0.069	0.014	23.006	1	0.000	0.933	0.907	0.960
Dist_fault	-0.001	0.000	35.879	1	0.000	0.999	0.999	0.999
Aspect	1.751	0.187	87.610	1	0.000	5.759	3.992	8.310
Elevation	-0.003	0.000	40.349	1	0.000	0.997	0.996	0.998
TWI	-0.786	0.099	62.994	1	0.000	0.456	0.375	0.553
Soil	2.057	0.192	114.801	1	0.000	7.819	5.368	11.391
Constant	3.949	1.766	5.003	1	0.025	51.898		

- B: Logistic coefficient; S.E.: Standard error estimate; Wald chi-square value $(B/S.E.)^2$; df: degree of freedom; Sig: Significance (P value); EXP(B): exponentiated coefficient (ODD ratio); and 95% C.I. for EXP(B): 95% confident index for EXP(B) range.

**** Artificial neural network***

ANN is computational information processing model. Multilayer perceptron (MLP) is the most popular ANN type which is included an input layer, an output layer, and one or more hidden layers between them (Polykretis & Chalkias, 2018). In this case study, we used three-layered feed-forward network: one input layer with 13 conditioning factors, hidden layers and one output layers. The number of hidden layers and the number of nodes in a hidden layer required for a particular classification problem are not easy to deduce but according to the experience of experts, it can be defined by: 2**Ni*+1 where *Ni* is the number of input factors. So, the structure is 13 x 27 x 1. It means the number of input factors, hidden layers, and output layers, respectively were generated by using MATLAB software.
As required by ANN model, all the range of input factors should be normalized to the range 0.1 to 0.9 following by the equation:

$$\mathrm{Normalized} = \frac{\mathrm{Pixel(i)} - \mathrm{Min}}{\mathrm{Max} - \mathrm{Min}}(0.8) + 0.1 \quad (6)$$

Where Normalized is the new value, Pixel(i) is the original value of the pixel i^{th}, Min is the minimum and Max is the maximum value of the original range.

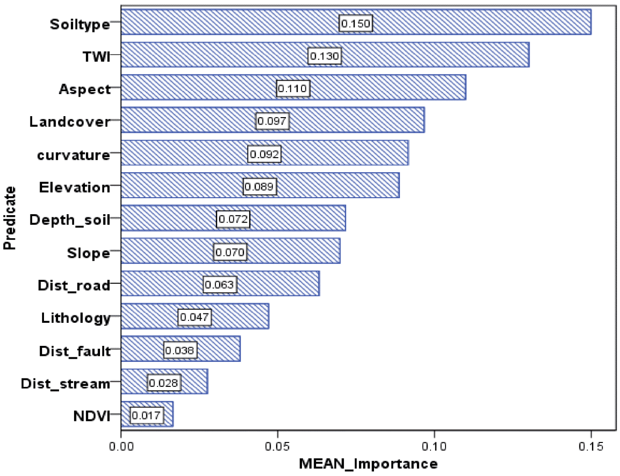

Figure 2. Importance of conditioning factors

**** Model combination***

To perform the model combination, landslide susceptibility maps obtained from the FR, LR, and ANN models were considered as input data. Kappa coefficient of model testing was used as weight of model combination.

$$\mathrm{LSI_{Combine}} = \frac{\mathrm{Kappa1} * \mathrm{LSI_{FR}} + \mathrm{Kapp2} * \mathrm{LSI_{LR}} + \mathrm{Kappa3} * \mathrm{LSI_{ANN}}}{\mathrm{Kappa1} + \mathrm{Kappa2} + \mathrm{Kappa3}} \quad (7)$$

Where: Kappa1, Kapp2, Kappa3 are Kappa coefficient of FR, LR and ANN respectively;

**** Model performance and validation***

In this study, confusion matrix was used to evaluate the performance of the trained landslide models. The reliability of the landslide models could be measured using 08 statistical criteria for evaluation as Sensitivity (producer's), Specificity (producer's), False Positive Rate (Commission Error), False Negative Rate (Omission Error), Positive Predictive power (User's), Negative Predictive power (User's) Eq, Overall accuracy, and Kappa index (Bui, Ho, et al., 2016; Bui, Tuan, Klempe, Pradhan, & Revhaug, 2016; Chen, Shahabi, et al., 2018; Tian, Xu, Hong, Zhou, & Wang, 2019). The strength of agreement given the Kappa magnitude is for 0.8–1.0 almost perfect, 0.6–0.8 substantial, 0.4–0.6 moderate, 0.2–0.4 fair, 0–0.2 slight, and ≤0 poor (Guzzetti et al., 2006; Landis & Koch, 1977).
The receiver operating characteristic curve is commonly used to generate the accuracy of landslide susceptibility model. Based on the result of analysis the AUC (area under the ROC curve) can be calculated by the curve performed as sensitivity on the y-axis and 1-specificity on the x-axis. The AUC was used for quantitative model analysis commonly. The value closes to zero means the model is non-informative, the value closes to 1 means a perfect model, while values in the range of (0.5–0.6), (0.6–0.7), (0.7–0.8), (0.8–0.9), and (0.9–1) can be categorized as poor, average, good, very good, and excellent, respectively. (Bui et al., 2011; Chen et al., 2017; Chen, Zhang, Li, & Shahabi, 2018; Costanzo, Chacon, Conoscenti, Irigaray, & Rotigliano, 2014; Wang et al., 2015; Yesilnacar & Topal, 2005).

3. RESULTS

3.1. Landslide susceptibility mapping

**** Frequency ratio***

The results show that each class of condition factors are calculated the frequency ratio based on the frequency of landslide overlapped with each class. The landslide frequency ratio value shows the elevation ranges from 1000 meter to 1500 meter have strong relationship with landslide where could be high potential of landslide. And other factors we also see that following: slope angle ranges from 30 to 40 degree have high potential, area with convex – convex and convex - concave of plan and profile type respectively could be have higher potential compare with others. The relationship between distance from the road and landslide occurrence seem to be closer to the road higher potential landslide and the same case with distance from the faults because the data shows that area located nearby faults less than 400 meter have higher potential. Oppositely, landslide occurrence frequently far from the stream area because the data show that area far from stream line over 40 meter have higher potential. The value of landslide frequency gains higher in residential/bare land and young forest/grass area and get lower in crop land and lowest in old forest area. The area gets TWI value smaller than 2.5 show higher frequency ration of landslide and get lower when the value of TWI increase. And other factor, we can see that aspect gets highest at south area, lithology unit gets highest at Phu Si Lung Complex-Dyke and vein phase, Soil gets highest at Hj (Yellow red humus soil on metamorphic rocks) and depth of soil reach thicker than 100 cm and the last but not least is NDVI get the highest value when it ranges from 0 to 0.35. Based on that value the landslide susceptibility index was calculated following Eq. (2) and Eq. (3). The result of landslide susceptibility map based on FR was produced (fig.3). Landslide susceptibility index was classified by using equal interval to 5 classes in ArcGIS: very low (0-0.2), low (0.2-0.4), medium (0.4-0.6), high (0.6-0.8) and remain very high (0.8-1) and statically result show 0.51%, 36.11%, 41.78%, 20.82% and 0.78% respectively. Intuitively for saying that, this result indicates the model's limitations on the distribution of class spaces that concentrate in the medium and the lower gradual values on both sides of the graph (Fig.3). This problem leads itself difficult to make decisions or warnings in real situation.

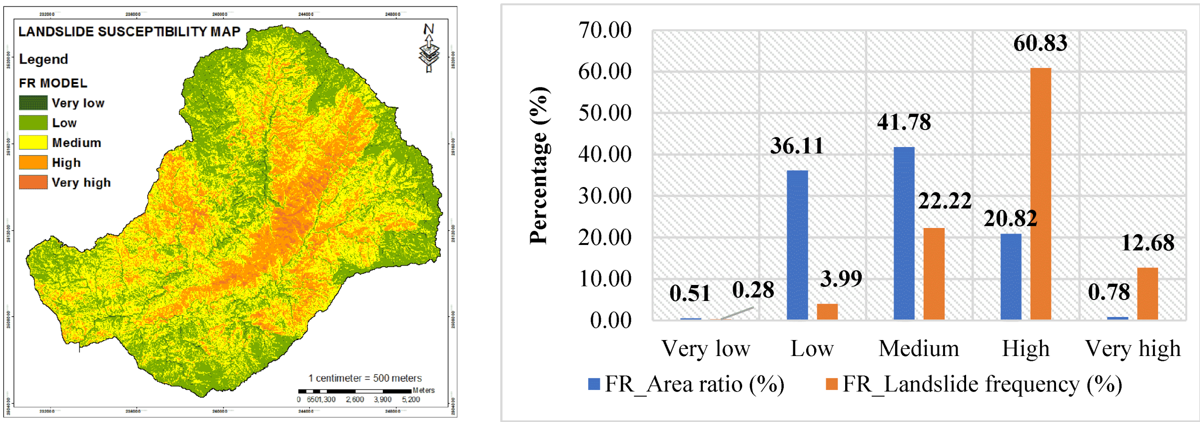

Figure 3. Frequency ratio model result: (left) Landslide susceptibility map, and (right) statistic chart of susceptibility categories and distribution of real landslides

**** Logistic regression model result***

This study using statistical package for social science software (SPSS) to calculate and evaluate the coefficient factor value and get the result show in table 3 and table 4. Based on the results, the achieved indicators are statistical significance. So, it is possible to apply the logistic regression equation (4) and (5) to generate the regression equation below:

Z = 3.949 + 0.037*SLOPE + 0.926*LANDCOVER + 2.057*SOIL + 1.751*ASPECT + (-0.003) *ELEVATION + (-0.001)*DIST_FAULT + (-0.786)*TWI + (-0.069)*DIST_ROAD + 2.354*LITHOLOGY + 0.018*DIST_STREAM + (-0.789)*CURVATURE + (-2.362)*NDVI + (-0.016)*DEPT_SOIL.

The value of coefficients of condition factor express the relative weight value of each one with landslide potential occurrence (positive and negative) so distance from the stream, NDVI, lithology, land cover type, aspect and soil types coefficients are positive and remain factor coefficients are negative. According to that result, the landslide susceptibility map obtained by LR model was produced (Fig.3.). And the value of landslide susceptibility index also classified using equal interval to 5 classes in ArcGIS: very low (0-0.2), low (0.2-0.4), medium (0.4-0.6), high (0.6-0.8) and remain very high (0.8-1) and statically result show 58.54%, 12.00%, 8.55%, 8.03%, and 12.87% respectively. The rate of real landslide points falling into very high level reached the highest, this is significant distribution because of normally landslide could be occurred in high and very high susceptibility (Fig.4.)

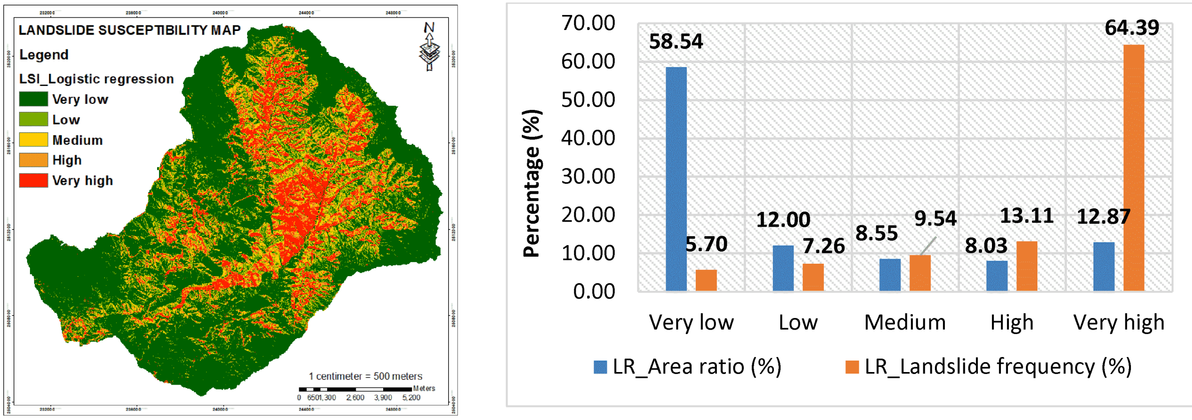

Figure 4. Logistic regression model result: (left) Landslide susceptibility map, and (right) statistic chart of susceptibility categories and distribution of real landslides

**** ANN model result***

ANN has the ability to perform parametric analysis through manipulation of their connection weights. In this study, the method of partitioning weights was applied to figure out the hidden-output connection weights into components associated with each input neuron. The importance values of the factors are showed in figure 2 mean the contribution of each factor to compute the ANN model. Subsequently, the landslide susceptibility index was generated in whole area and categorized by using equal interval in ArcGIS software. The real landslide points frequency falling in very high susceptibility area rates 79.08% and few points of landslide distributed in low and very low level (figure 5). This is kind of very good performance of model prediction ability.

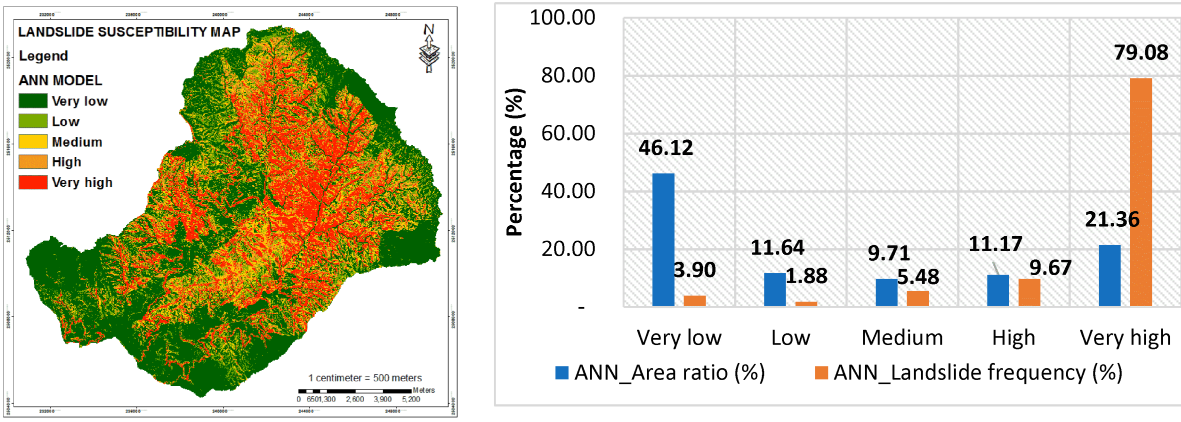

Figure 5. ANN model result: (left) Landslide susceptibility map, and (right) statistic chart of susceptibility categories and distribution of real landslides

****Combination model result***

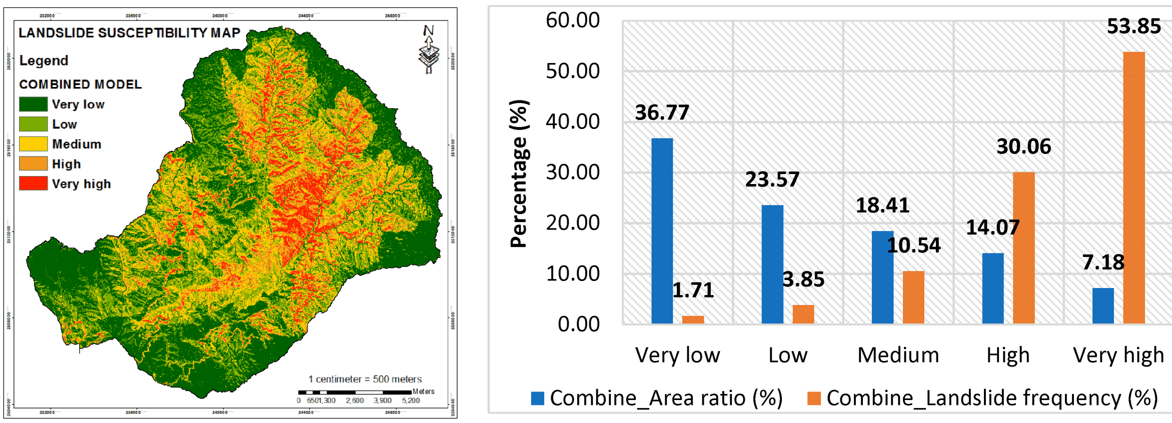

Figure 6. Combination model result: (left) Landslide susceptibility map, and (right) statistic chart of susceptibility categories and distribution of real landslides

According to the accuracy analysis of individual model results, the testing data was used to compute the Kappa coefficients then applied the equation 7 to generate the LSI_combined model. The landslide susceptibility map was produced and classified to 5 categories using equal interval function in ArcGIS software (fig.6).

3.2. Model evaluation and comparison of the landslide susceptibility mapping

This study used confusion matrix, Kappa coefficient, AUC value, standard error and 95% confidence index of AUC evaluation to represent accuracy of each model and compare the result. Both in training and testing result, the highest accuracy is combination model and following by ANN, LR and the lowest is FR model (Table 5). Therefore, the ROC curve shows the high value of AUC with good predicable (Fig.7.).

Table 5. Accurate index for the model training and model testing

Index	Model training			
	LR	**FR**	**ANN**	**Combined**
Sensitivity (Producer's)	0.83	0.85	0.96	0.91
Specificity (Producer's)	0.95	0.84	0.84	0.94
False Positive Rate (Commission Error)	0.05	0.16	0.16	0.06
False Negative Rate (Omission Error)	0.17	0.15	0.04	0.09
Positive Predictive power (User's)	0.95	0.84	0.86	0.94
Negative Predictive power (User's)	0.85	0.85	0.95	0.92
Overall accuracy	0.89	0.84	0.90	0.93
Kapa	0.78	0.69	0.8	0.85
AUC	0.949	0.913	0.954	0.97
S.E.	0.007	0.009	0.007	0.005
95% CI (lower – Upper)	0.935-0.963	0.894-0.931	0.941-0.967	0.960-0.980
Index	**Model testing**			
	LR	**FR**	**ANN**	**Combined**
Sensitivity (Producer's)	0.82	0.85	0.88	0.88
Specificity (Producer's)	0.90	0.80	0.85	0.91
False Positive Rate (Commission Error)	0.10	0.20	0.15	0.09
False Negative Rate (Omission Error)	0.18	0.15	0.12	0.12
Positive Predictive power (User's)	0.89	0.81	0.85	0.90
Negative Predictive power (User's)	0.84	0.84	0.88	0.88
Overall accuracy	0.86	0.83	0.86	0.89
Kapa	0.72	0.65	0.73	0.79
AUC	0.929	0.91	0.944	0.953
S.E.	0.013	0.014	0.011	0.01
95% CI (lower – Upper)	0.905-0.954	0.882-0.938	0.922-0.966	0.933-0.973

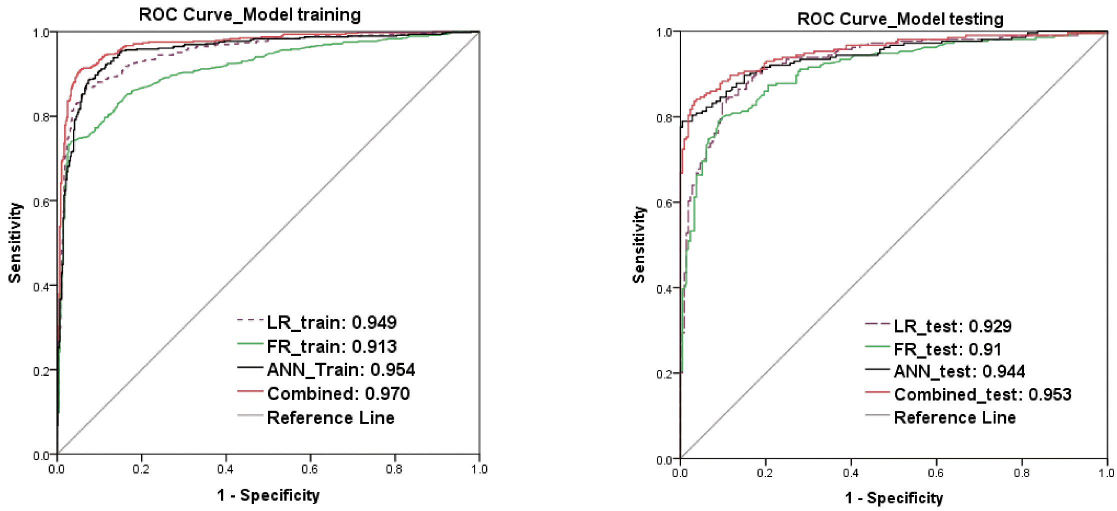

Figure 7. Receiver operating characteristics (ROC) curves using the training dataset (left) and testing dataset (right)

CONCLUSION AND DISCUSSION

The result of Frequency ratio, Logistic regression, Artificial neural network and combination models shows reasonable performance in landslide susceptibility mapping for Thu Lum basin (very local area) with high prediction accuracies. All three models FR, LR and ANN are commonly used previously but the combined one is rarely used especially this study considered Kappa coefficient of individual model result to generate the model combination. The assumption gave at the beginning of this study was demonstrated preliminarily based on result performances. The combination model is completely applicable to combining multiple independent models together and based on this research, it can be better than others. Conclusively, the results of this study for generating landslide detection and susceptibility analysis may significant contribution in providing basic information which could be useful for decision making such as land use planning, slope management, hazard prevention and mitigation activities.

REFERENCES

Bai, S. B., Wang, J., Lu, G. N., Zhou, P. G., Hou, S. S., & Xu, S. N. (2010). GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology, 115*(1-2), 23-31.

Bui, D. T., Ho, T. C., Pradhan, B., Pham, B. T., Nhu, V. H., & Revhaug, I. (2016). GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences, 75*(14). doi:ARTN 1101
10.1007/s12665-016-5919-4

Bui, D. T., Lofman, O., Revhaug, I., & Dick, O. (2011). Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Natural Hazards, 59*(3), 1413-1444.

Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides, 13*(2), 361-378. doi:10.1007/s10346-015-0557-6

Chae, B. G., Park, H. J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal, 21*(6), 1033-1070.

Chen, W., Shahabi, H., Zhang, S., Khosravi, K., Shirzadi, A., Chapi, K., . . . Bin Ahmad, B. (2018). Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Applied Sciences-Basel, 8*(12).

Chen, W., Sun, Z. H., & Han, J. C. (2019). Landslide Susceptibility Modeling Using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. *Applied Sciences-Basel, 9*(1).

Chen, W., Xie, X. S., Peng, J. B., Wang, J. L., Duan, Z., & Hong, H. Y. (2017). GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naive-Bayes tree, and alternating decision tree models. *Geomatics Natural Hazards & Risk, 8*(2), 950-973. doi:10.1080/19475705.2017.1289250

Chen, W., Zhang, S., Li, R. W., & Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naive Bayes tree for landslide susceptibility modeling. *Science of the Total Environment, 644*, 1006-1018. doi:10.1016/j.scitotenv.2018.06.389

Costanzo, D., Chacon, J., Conoscenti, C., Irigaray, C., & Rotigliano, E. (2014). Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy). *Landslides, 11*(4), 639-653. doi:10.1007/s10346-013-0415-3

Felicisimo, A., Cuartero, A., Remondo, J., & Quiros, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides, 10*(2), 175-189. doi:10.1007/s10346-012-0320-1

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology, 31*(1-4), 181-216. doi:Doi 10.1016/S0169-555x(99)00078-1

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M. (2006). Estimating the quality of landslide susceptibility models. *Geomorphology, 81*(1-2), 166-184. doi:10.1016/j.geomorph.2006.04.007

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. doi:10.1016/j.geomorph.2005.06.002

Kadirhodjaev, A., Kadavi, P. R., Lee, C. W., & Lee, S. (2018). Analysis of the relationships between topographic factors and landslide occurrence and their application to landslide susceptibility mapping: a case study of Mingchukur, Uzbekistan. *Geosciences Journal, 22*(6), 1053-1067.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 16. doi:10.2307/2529310

Mandal, B., & Mandal, S. (2018). Analytical hierarchy process (AHP) based landslide susceptibility mapping of Lish river basin of eastern Darjeeling Himalaya, India. *Advances in Space Research, 62*(11), 3114-3132.

Nsengiyumva, J. B., Luo, G. P., Amanambu, A. C., Mind'je, R., Habiyaremye, G., Karamage, F., . . . Mupenzi, C. (2019). Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. *Science of the Total Environment, 659*, 1457-1472. doi:10.1016/j.scitotenv.2018.12.248

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690. doi:10.1007/s11135-006-9018-6

Park, S. J., Lee, C. W., Lee, S., & Lee, M. J. (2018). Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea. *Remote Sensing, 10*(10).

Polykretis, C., & Chalkias, C. (2018). Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Natural Hazards, 93*(1), 249-274. doi:10.1007/s11069-018-3299-7

Pourghasemi, H. R., Gayen, A., Park, S., Lee, C. W., & Lee, S. (2018). Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and NaiveBayes Machine-Learning Algorithms. *Sustainability, 10*(10).

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. *Environmental Modelling & Software, 25*(6), 747-759. doi:10.1016/j.envsoft.2009.10.016

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T., & Akgun, A. (2014). Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences, 7*(2), 725-742. doi:10.1007/s12517-012-0807-z

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. doi:10.1016/j.earscirev.2018.03.001

Sahana, M., & Sajjad, H. (2017). Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: a case from Rudraprayag district, India. *Journal of Mountain Science, 14*(11), 2150-2167. doi:10.1007/s11629-017-4404-1

Tian, Y. Y., Xu, C., Hong, H. Y., Zhou, Q., & Wang, D. (2019). Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: an example of the 2013 Minxian (China) Mw 5.9 event. *Geomatics Natural Hazards & Risk, 10*(1), 1-25. doi:10.1080/19475705.2018.1487471

Tsai, F., Hwang, J. H., Chen, L. C., & Lin, T. H. (2010). Post-disaster assessment of landslides in southern Taiwan after 2009 Typhoon Morakot using remote sensing and spatial analysis. *Natural Hazards and Earth System Sciences, 10*(10), 2179-2190. doi:10.5194/nhess-10-2179-2010

Wang, L. J., Guo, M., Sawada, K., Lin, J., & Zhang, J. C. (2015). Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. *Catena, 135*, 271-282. doi:10.1016/j.catena.2015.08.007

Yesilnacar, E., & Topal, T. (2005). Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology, 79*(3-4), 251-266. doi:10.1016/j.enggeo.2005.02.002

Yilmaz, I. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Computers & Geosciences, 35*(6), 1125-1138. doi:10.1016/j.cageo.2008.08.007

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides, 13*(5), 839-856. doi:10.1007/s10346-015-0614-1

Zhang, T. Y., Han, L., Chen, W., & Shahabi, H. (2018). Hybrid Integration Approach of Entropy with Logistic Regression and Support Vector Machine for Landslide Susceptibility Modeling. *Entropy, 20*(11). doi:ARTN 884

10.3390/e20110884