Science on Natural Resources and Environment 43 (2022) 104-113

Science on Natural Resources and Environment
Journal homepage: tapchikhtnmt.hunre.edu.vn

APPLICATION OF THE GIS AND R PROGRAM FOR LANDSLIDE SUSCEPTIBILITY MAPPING: A CASE STUDY IN VAN YEN, YEN BAI, VIETNAM

Pham Thi Thanh Thuy[1], Le Thi Thu Ha[1], Vu Ngoc Phan[1], Vu Ngoc Phuong[2]
[1]Hanoi University of Natural Resources and Environment, Vietnam
[2]University of Transport and Communications, Vietnam

Received 31 October 2022; Accepted 28 November 2022

Abstract

This study presents the r.landslide tool, an open source add-on to the open source Geographic Information System (GIS) GRASS software for landslide susceptibility mapping. The tool was written in Python language and works on the top of an Artificial Neural Network (ANN) fed with environmental parameters and landslide databases, such as: DTM, NDVI, Aspect, Geology, Faults, Plan Curvature, Profile Curvature, Rivers, Roads, Slope, No Landslide Zones (NLZ). In order to illustrate the application and effectiveness of the developed tool, a case study is presented for the Van Yen district, Yen Bai province, Vietnam. The resulted map with four landslide susceptibility classes: Low, moderate, high and very high susceptibility for landslide, which are derived based on the correspondence with landslide inventory. The map indicates that about 42 % of the area is very high and highly susceptible for landslide. The landslide susceptibility map can be useful for the decision - makers and planners in choosing suitable locations for the long - term development.

Keywords: GIS; R program/r.landslide; Landslide susceptibility zone.

Corresponding author. Email: pttthuy.tdbd@hunre.edu.vn

1. Introduction

Landslide is soil or rock mass movement, or a mixture of both, down and out of the slope. The natural properties of slope stability influence its susceptibility. Recently year, Vietnam is influenced by climate change and human activities such excavation of slopes for road cuts or such deforestation, which are one of the causes contributed to landslide happening [1]. Especially, the Northwest mountainous regions of Vietnam with various strong dissections by tectonics, the areas are heavily affected by landslide phenomenon [2]. Frequency and magnitude of landslides in this region have been increased, not only causing losses and damages to people, also damaging enormous properties in terms of both direct and indirect costs [3, 4]. Landslide susceptibility mapping is an urgent task for the government for the mountainous regions [1], including Yen Bai province, to find proper and effective strategies in land use planning and management, also forecasting and finding measures to mitigate subsequent losses to future landslides [3, 4].

Research on the assessment and prediction of landslide susceptibility uses a variety of methods depending on the size of the study area. For example: The heuristic method applies geomorphological mapping to large - scale areas based on experts' judgment of variables such as slopes, faults and geology [5]. The deterministic method applies to the small - scale area by analyzing the geotechnical stability condition of parameters. A statistical approach is a new approach to mapping landslide hazards by combining the possibility of landslides from statistical data and the physical parameters of landslides. This approach is appropriate for assessing landslides in a medium - scale area which helps inform the regional spatial planning [6, 7]. Research on landslides has been widely applied using a method or comparing them [8 - 12].

Open-source Geographic Information System (GIS) software can process statistical models [13]. One of them is R program, which has cutting - edge spatial packages to behave as a fully featured GIS [14]. Several advantages of the utilization of R language for spatial analysis such as its command line interfaces allow a rapid description of workflow and reproducibility, has sophisticated and customizable graphics and have an extensive range of functions through an additional package, integrated processing, analysis and modeling framework. R statistics has a wide range of functions and libraries that allow using all statistical tools with advanced visualization capabilities [15]. The recent updates of the libraries attached to R environment made the output and result very handy and without the need to change the working environment or data format, which will reduce the uncertainty of switching back and forth between different geospatial and statistical analysis platforms [12]. Some studies have analyzed land susceptibility using R Program [16 - 18]. This study uses R program to control landslides and generate a landslides susceptibility map in Van Yen district, Yen Bai province.

2. Study area

2.1. Geographical location

The study area is Mo Vang commune in Van Yen district (Figure 1) (Van Yen is a mountainous district in the north of Yen Bai province, Vietnam), between the latitude 21º50'30"N and 22º12'N and between longitude 104º23'E and 104º48'E. The region happens landslide phenomena, losing properties and damaging constructions each year.

Figure 1: The study area map

2.2. Topography, hydrology and climate

Van Yen's topography is relatively complex, with many hills and mountains. The terrain gradually rises from the Southeast to the Northwest. The difference in topography between regions in the district is very large, with the highest peak at 1.952 m, the lowest place being 20 m above sea level.

The river system is dense with different terrain types: Craggy high mountains, rolling hills, alternating with valleys and narrow alluvial fields along the river.

Van Yen district is located in a hot and humid tropical climate, combined with divided terrain to form two climate sub - regions:

- Northern region (from Trai Hut to the North): Average elevation is 500 m above sea level. The average temperature is 21 - 23 ºC. Average rainfall is 1.800 mm/year. Humidity is often 80 - 85 %, this area is affected by Lao wind;

- Southern mountainous region (from Trai Thu to the South): Influenced by the northeast monsoon, with heavy rainfall, average 1.800 - 2.000 mm/year, average temperature 23 - 24 ºC, air humidity 81 to 86 %.

2.3. Population

The average population as of 2019 is 129.059 people. Of which, 61.981 men, accounting for 50,37 %; Female 61.075 people, accounting for 49,63 %. The population in urban areas accounts for 10,26 %; rural areas accounted for 89,76 %. The natural population growth rate is 15,12 %, the average population density is 88,5 people/km^2. The whole district has 12 ethnic groups: Kinh ethnic group (52,86 %), Tay ethnic group (15,58 %), Dao ethnic group (25,4 %), H'mong ethnic group (4,43 %), other ethnic groups (1.73 %).

3. Data and methodology

An artificial neural network (ANN) is a set of interconnected nodes useful for modeling problems with a complex relationship between analysis factors [23]. Its central processing unit is the neuron, which performs mathematical procedures to generate a result based on a set of input variables [24]. The application of an ANN in landslide susceptibility analysis is ideal because this phenomenon is dynamic and nonlinear [25]. The ANN architecture consists of a set of inputs (conditioning factors), a set of intermediate layers (hidden layers) that perform the processing and an output layer [24] with the prediction result (Figure 2).

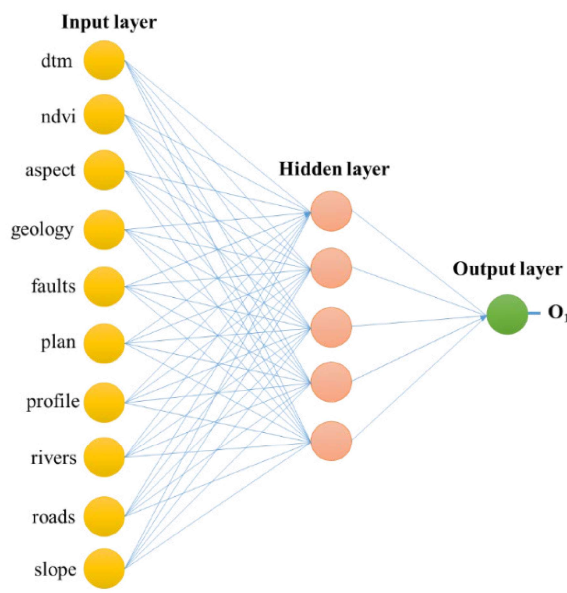


Figure 2: The structure of ANN

ANN implementation in this research was performed using the R program in QGIS which was written by Python language.

In QGIS, adding a script is simple. The easiest way is to open the *Processing toolbox* and choose *Create new R script* from the R menu (labelled with an *R* icon) at the top of the *Processing Toolbox*.

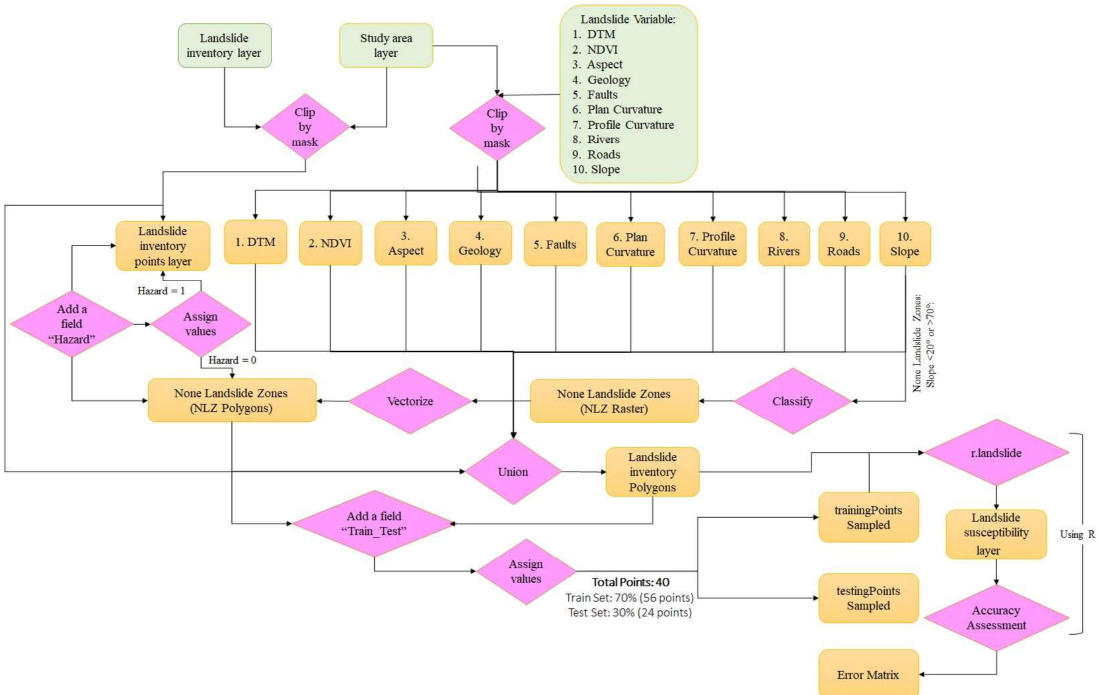


Figure 3: Flow chart of the research

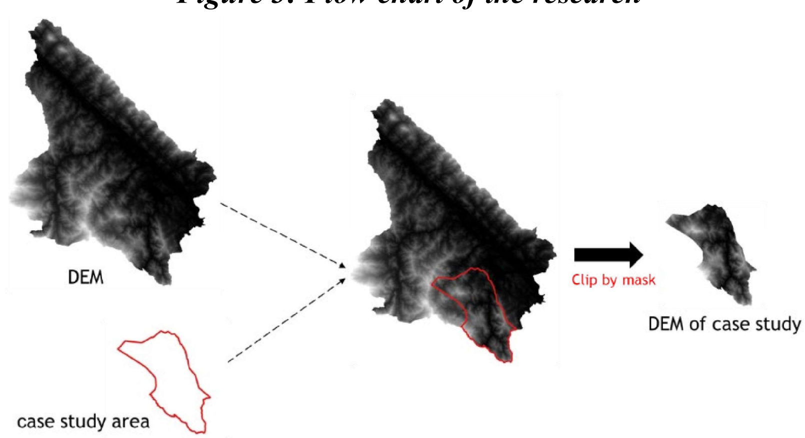


Figure 4: Clip DEM by case study area mask

Collected data in raster format (DTM, NDVI, Aspect, Geology, Faults, Plan Curvature, Profile Curvature, Rivers, Roads, Slope) and vector format (study area boundary layer, landslide inventory layer). Clip all raster layers using as a mask the vector layer of the group's sub - area (Figure 4 and Figure 5).

Causative factors for landslide susceptibility mapping in a certain study area should be selected carefully based on relevance, availability, and scale of mapping [19, 20]. Based on previous studies in the same area [21, 22], thereby determining the correlation and contribution of factors in the occurrence of landslides, therefore, 10 factors considered for landslide susceptibility mapping. Figure 3 describes causative factors which were selected: DTM, NDVI, Aspect, Geology, Faults, Plan Curvature, Profile Curvature, Rivers, Roads, Slope as input data. Rasters must have with equal resolution and extension of the clipped DTM.

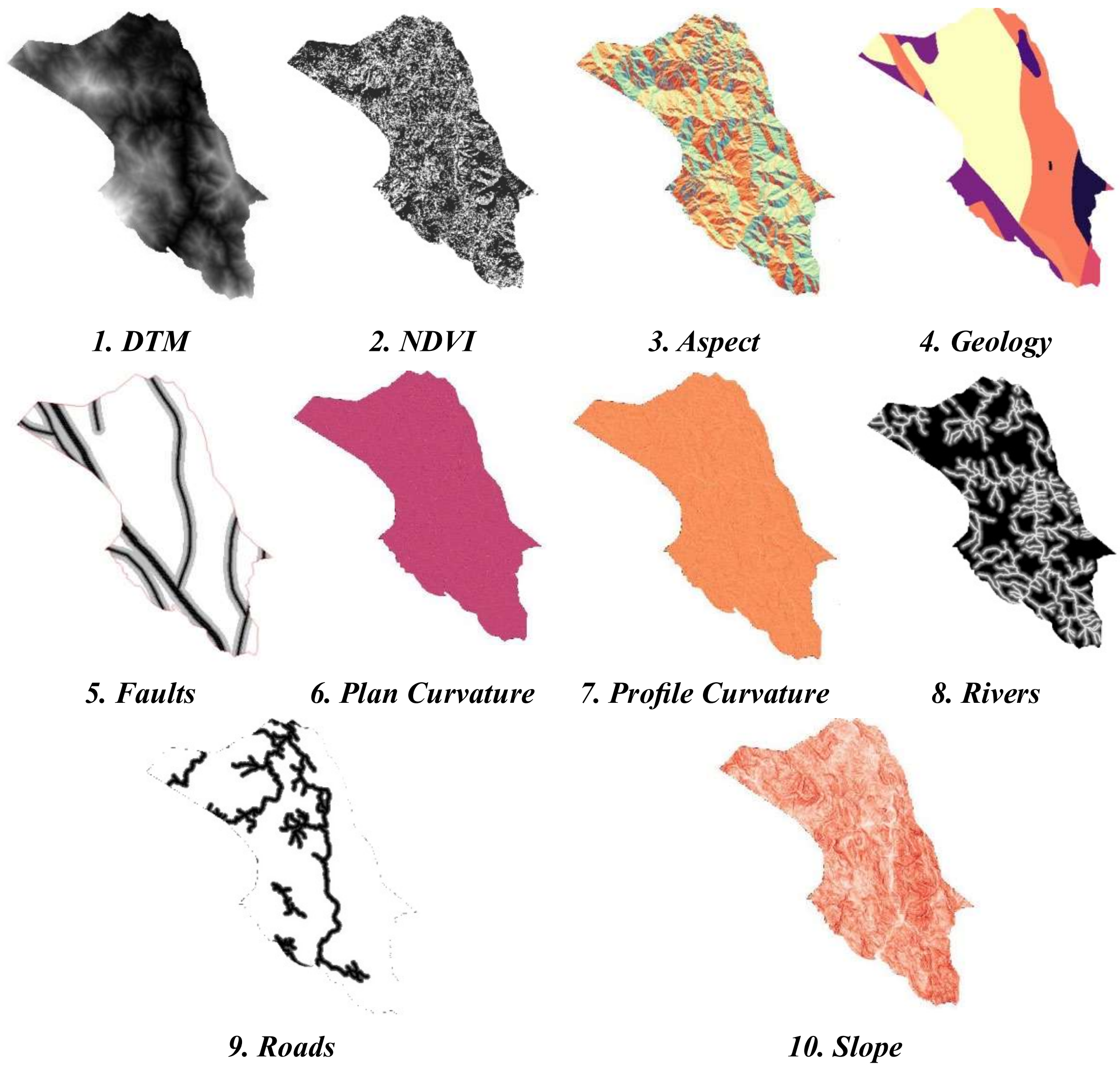


Figure 5: The causative factors for landslide susceptibility mapping

Clip the landslide inventory layer using as a mask the vector layer of the group's sub - area (Figure 6).

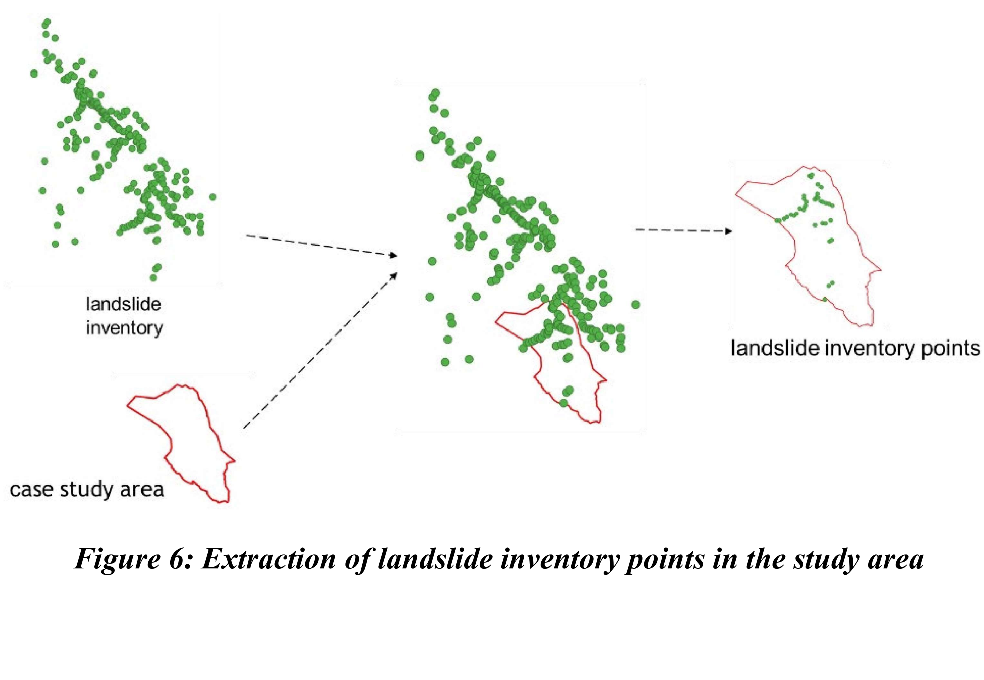
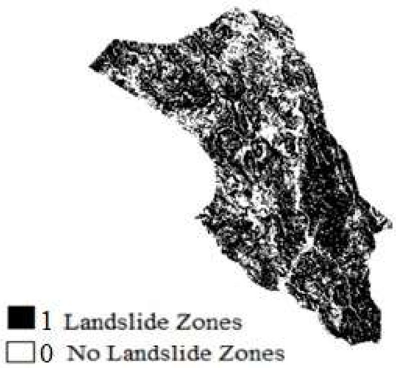


Figure 6: Extraction of landslide inventory points in the study area

Define areas with low possibility of landslides according to the Slope angle. We assume No Landslide Zones (NLZ) are where the Slope is < 20° or > 70° (Figure 7a). After that, vectorize the resulted raster (use the raster values as categories) to obtain the polygons of NLZ (Figure 7b). Thus, the landslide susceptibility areas will not appear in the NLZ.



a) NLZ: The Slope is <20° or >70°

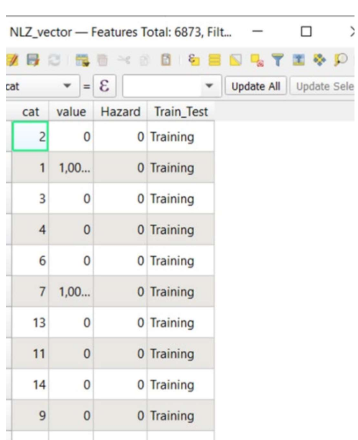
b) Vectorize the resulted raster to vector

Figure 7: No Landslide Zones (NLZ)

Create new field 'Hazard' in both of the attribute table (landslide inventory points and NLZ polygons). Where, 0 is assigned to the NLZ and 1 to the landslide inventory. Perform a Union operation on the Landslide Inventory polygons. Decide a training - testing ratio that was used for machine learning model.

After selecting the percentage of polygons for training/testing accordingly for both Landslide Inventory and NLZ. Create new text attribute 'Train_Test' and assign the value 'Training' or 'Testing' according to the select polygons. Create 40 random points inside the polygons since the landslide inventory is a point layer, we have to create the same number of points that represent the NLZ. That means we use 70/30 training/testing ration we will need to have 56 training points and 24 testing.

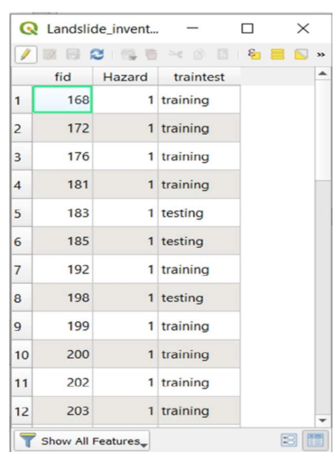
Using *Select Features by Value* and select according to the *'Hazard'* and '*Train_Test'* field to to assign the corresponding value (Figure 8).



NLZ_vector — Features Total: 6873, Filt...

cat	value	Hazard	Train_Test
2	0	0	Training
1	1,00...	0	Training
3	0	0	Training
4	0	0	Training
6	0	0	Training
7	1,00...	0	Training
13	0	0	Training
11	0	0	Training
14	0	0	Training
9	0	0	Training
17	0	0	Training

a) NLZ polygons



Landslide_invent...

	fid	Hazard	traintest
1	168	1	training
2	172	1	training
3	176	1	training
4	181	1	training
5	183	1	testing
6	185	1	testing
7	192	1	training
8	198	1	testing
9	199	1	training
10	200	1	training
11	202	1	training
12	203	1	training

Show All Features

b) Landslide inventory

Figure 8: Attribute tables of NLZ polygons and Landslide inventory

Merge separately the training and testing layers into two point layers training points and testing points. Sample the environmental factors with the training and testing point layers. At the end, we have two layers trainingPointsSampled and testingPointsSampled with following attribute tables:

trainingPointsSampled.gpkg — Features Total: 56, Filtered: 56, Selected: 0

	fid	Hazard	geology	roads	rivers	ndvi	faults	profile	aspect	slope	plan	dtm
1	1	0	7	501	3	4	500	0,00214	243,43495	19,68573	0,0644	369
2	2	0	36	500	3	1	501	-0,01134	306,8699	11,30993	0,02816	504
3	3	0	36	200	2	6	500	0,00049	348,69006	22,19161	-0,00151	678
4	4	0	15	501	2	1	50	-0,01095	7,12502	17,87401	-0,02015	307
5	5	0	27	200	1	6	501	0,01341	71,56505	14,19695	0,04301	1005
6	6	0	27	501	4	1	501	0,00383	51,34019	14,36606	0,05851	672
7	7	0	15	501	3	1	501	-0,0066	101,30993	11,52797	-0,02836	544
8	8	0	23	501	5	1	501	0,00795	263,65982	19,91117	0,00894	617
9	9	0	36	200	2	1	501	-0,00507	8,1301	15,79317	0,02014	519
10	10	0	15	501	4	6	200	-0,00571	270	15,64225	0	388
11	11	0	15	200	3	1	50	0,00525	321,34018	14,36606	0,05241	733
12	12	0	27	500	5	6	501	-0,00428	45	15,79317	-0,02828	407
13	13	0	15	501	2	4	500	-0,00189	32,47119	27,54367	0,0052	329

Show All Features

a) TrainingPointSampled

testingPointsSampled.gpkg — Features Total: 24, Filtered: 24, Selected: 0

	fid	Hazard	geology	roads	rivers	ndvi	faults	profile	aspect	slope	plan	dtm
1	1	0	27	50	3	6	501	-0,00498	191,30994	11,52797	0,02595	677
2	2	0	15	501	5	1	500	-0,00016	350,53769	13,67495	-0,02559	551
3	3	0	27	501	3	6	501	-0,00022	339,44397	18,86838	0,0195	289
4	4	0	27	501	1	1	501	0,00609	254,05461	16,23579	0,0423	212
5	5	0	27	501	4	1	501	-0,01517	281,30994	11,52797	-0,10922	152
6	6	0	15	501	3	6	50	0,01206	279,46231	13,67495	0,02488	709
7	7	0	27	50	5	4	501	0,00575	108,43495	20,78042	-0,01855	41
8	8	0	36	500	1	6	200	-0,00253	284,03625	9,36516	0,05478	310
9	9	0	15	501	3	6	500	-0,0043	203,19859	16,94236	-0,06792	1094
10	10	0	27	501	1	1	501	0,02883	296,56506	15,02026	0	424
11	11	0	36	500	4	1	501	-0,00566	130,60129	20,24306	0,00123	593
12	12	0	15	500	1	1	50	0,00609	15,9454	16,23579	0,0423	585
13	13	1	36	100	1	4	501	-0,00964	196,69925	22,66601	-0,01659	530

Show All Features

b) TestingPointSampled

Figure 9: Attribute tables of two layers trainingPointsSampled and testingPointsSampled

4. Result and discussions

After running the r.landslide tool, the result is a landslide susceptibility layer (Figure 10).

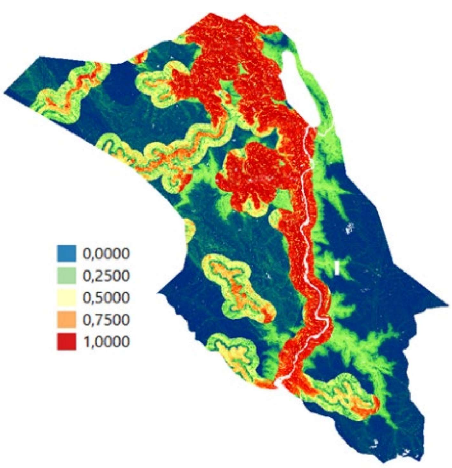


Figure 10: The first landslide susceptibility layer

Reclassify the susceptibility raster map using 4 classes such as: [0, 0.25) = low; [0.25, 0.5) = moderate; [0.5, 0.75) = high; [0.75, 1] = very high (Figure 12b).

Use the QGIS tool *Raster->Raster Calculato*r along with this expression:

("firstlandslide susceptibility layer" < 0.25) 1 + (("first landslide susceptibility layer" >= 0.25) AND ("first landslide susceptibility layer" < 0.5)) *2 + (("first landslide susceptibility layer">= 0.5) AND ("first landslide susceptibility layer" < 0.75))*3 + ("first landslide susceptibility layer" >=0.75)* 4*

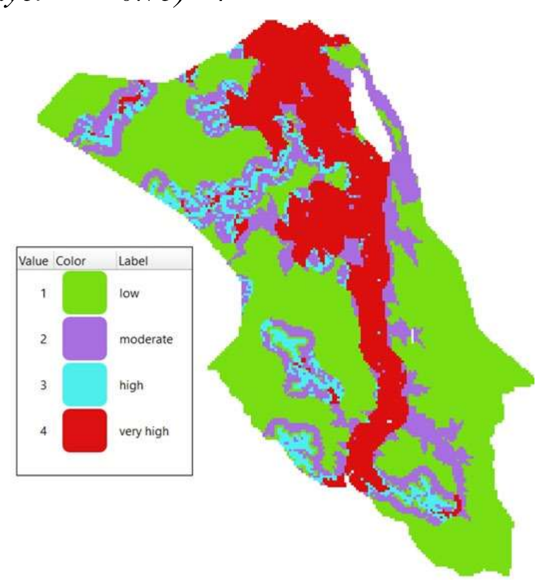


Figure 11: The second landslide susceptibility layer after reclassification with resolution: 12.5 m

Validation of the efficiency of the GIS and R program on producing landslide susceptibility maps was done using Accuracy Assessment tool (which is also written in Python language).

Reclassify the first landslide susceptibility layer into two classes (0 and 1): [0,0.5) = 0 and [0.5,1) = 1. The reclassified raster is used only for the validation purpose. In QGIS, use *Processing => Scripts => Accuracy Assessment* and *Sampling and first landslide susceptibility* layer and *testingPointsSampled.gpkg*, where reference data column is *Hazard*.

Table 1 shows that landslide sensitivity classification accuracy reaches 75 %. The accuracy of the classification results is average, which can be useful to decision makers and planners in choosing the right site for long - term development.

Table 1. Error matrix

1	Thematic...	0.0	1.0	UA	PA	OA
2	1.0	4	10	0.71	0.83	
3	0.0	8	2	0.8	0.67	0.75

Use the QGIS Processing tool *Processing => Raster Analysis => Raster layer zonal statistics* to compute the population counts in each susceptibility class. The percentage of population per each susceptibility class was showed by a pie chart (Table 2).

Table 2. Landslide susceptibility Zonal Statistic

fid	zone	m²	sum	count	min	max	mean
1	1	90795249,3568692	3141,64875661	10811	0,16516007	0,66012973	0,29059742
2	3	11883755,23448062	461,915346	1415	0,17642273	0,80610842	0,32644194
3	2	33803614,71292189	1394,14942141	4025	0,17619582	0,71580124	0,34637253
4	4	40354377,31567446	2127,63867755	4805	0,20355086	0,84888726	0,44279681

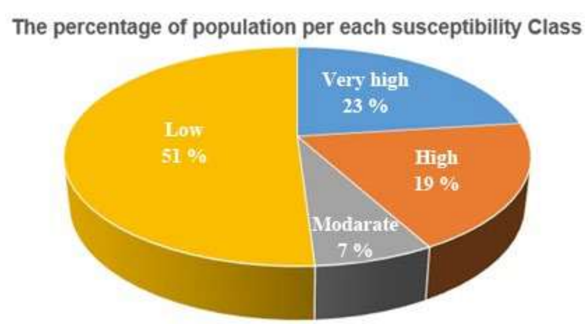


Figure 12: Landslide susceptibility statistic chart

The study applied GIS and the R program of QGIS to process input information layers, created necessary data layers for the purpose of calculating and statistic the extent of landslides in the Mo Vang area, Van Yen district, Yen Bai province, Vietnam. Research results show that landslide with very high risk is 23 %, high is 19 %, medium is 7 % and low is 51 %.

R is an open - source program widely used because it can integrate data, analysis, and graphs in a single narrative. We use this program to model landslide susceptibility algorithm using the ANN method and apply it to a region. The result of this model is no different from using ArcGIS software.

However, creating a landslides susceptibility algorithm in R model has an advantage in that other researchers can reinterpret and reevaluate the program by modifying its syntax and codes to get a more comprehensive and appropriate model applying in a specific region.

Acknowledgments: The author would like to thank Politecnico di Milano, University and Hanoi University of Natural Resources and Environment for providing valuable data and an advanced GIS course (the course is a product of international cooperation between the two universities within the framework of the Protocol, code number NDT/IT/21/14 led by Dr. Truong Xuan Quang).

REFERENCES

[1]. Q. H. Le (2014). *Landslide inventory and susceptibility assessment for mountainous provinces in Vietnam*. The government project 2012 - 2017.

[2]. T. Trinh, D. M. Wu & J. Z. Huang (2016). *Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study in Yen Bai province, Vietnam*. Conference: ICETI, Volume: 1, ISBN 978-1-138-02996-5.

[3]. Bui, T. D., B. Pradhan, O. Lofman, I. Revhaug & O. B. Dick (2012). *Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS*. Comput. Geosci. 45, 199 - 211.

[4]. Duc, D. M. (2013). *Rainfall - triggered large landslides on December 15th 2005 in Van Canh district, Binh Dinh province, Vietnam*. Landslides 10(2), 219 - 230.

[5]. 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.

[6]. Rossi, M., Reichenbach, P. (2016). *LAND-SE: A software for statistically based landslide*. Geoscientific Model Development, p. 9533 - 9543.

[7]. Fausto, G., Alberto C., Mauro C., Paola R. (1999). *Landslide hazard evaluation: A review of current techniques and their application in a multi - scale study, Central Italy*. Geomorphology, p. 181 - 216.

[8]. Anis, Z., Wissem, G., Vali, V., Smida, H., Essghaie, G. M. (2019). *GIS - based landslide susceptibility mapping using bivariate statistical methods in North - western Tunisia*. Open Geosciences, 11:708-726.

[9]. Pamela, Sadisun, I. A., Arifianti, Y. (2018). *Weights of evidence method for Landslide susceptibility mapping in Takengon, Central Aceh, Indonesia*. IOP Conference Series: Earth and Environmental Science, 118, 012037.

[10]. Silalahi, F.E.S., Pamela, Arifianti, Y., Hidayat, F. (2019). *Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia*. Geoscience Letters, 6(1).

[11]. Singh Pradhan, A. M., Dawadi, A., Kim, Y. T. (2012). *Use of different bivariate statistical landslide susceptibility methods: A case study of Khulekhani watershed, Nepal*. Journal of Nepal Geological Society, 44, 1 - 12.

[12]. Sumaryono, Muslim D., Sulaksana N., Triana Y. D. (2015). *Weights of evidence method for Landslide susceptibility mapping in Tandikek and Damar Bancah, West Sumatra, Indonesia*. International Journal of Science and Research (IJSR), Vol. 4, Issue 10, October 2015, 1283 - 1290.

[13]. Christos P., Christos C. (2018). *Comparison and evaluation of landslide susceptibility maps obtained from the weight of evidence, logistic regression, and artificial neural network models*. Natural Hazards, vol. 93, no. August 2018, p. 249 - 274.

[14]. Roger, S. B., Edzer, P. and Virgilio G. R. (2013). *Applied spatial data analysis with R*. New York: Springer.

[15]. Hadley, W. (2014). *Tidy data*. Journal of Statistical Software, vol. 59, no. 10, p. 1 - 23.

[16]. Althuwaynee, O. F., Musakwa, W., Gumbo, T., Reis, S. (2017). *Applicability of R statistics in analyzing landslides spatial patterns in Northern Turkey*. 2nd International Conference on Knowledge Engineering and Applications (ICKEA).

[17]. Thinnukool O., Kongchouy, N., Choonpradub C. (2014). *Detection of land use change using R program (A case study of Phuket island, Thailand)*. Research Journal of Applied Sciences, 9:228-237.

[18]. Tonini, Marj Abellan, Antonio (2013). *Rockfall detection from terrestrial LiDAR point clouds: A clustering approach using R*. Journal of Spatial Information Science. 8. 10.5311/JOSIS.2014.8.123.

[19]. Cascini, L. (2008). *Applicability of landslide susceptibility and hazard zoning at different scales*. Eng. Geol. 102(3-4), 164 - 177.

[20]. Soeters, R. & C. J. V. Van Westen (1996). *Slope instability recognition analysis and zonation*.

[21]. Nguyen, T. V. (2009). *Building methodology for estimating geo-hazard risk in the northwest mountainous cities of Vietnam using RS&GIS*: *Case study in Yen Bai city*. Ministry of Sciences and Technology.

[22]. Nguyen, X. K. (2012). *Assessment on present situation of geo-hazards in provinces of Ha Giang, Cao Bang, Tuyen Quang and Bac Kan - Causes, forecast zoning and recommendation for risk prevention and reduction*.

[23]. Tien Bui, D.; 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 2016, 13, 361 - 378.

[24]. Ciaburro, G.; Venkateswaran, B. (2017). *Neural Network with R: Smart models using CNN, RNN, Deep Learning and Artificial Intelligence Principles*. Packt Publishing Ltd: Birmingham, UK, 2017; Volume 91.

[25]. Chen, H.; Zeng, Z.; Tang, H. (2015). *Landslide deformation prediction based on Recurrent Neural Network*. Neural Process. Lett. 2015, 41,169 - 178.