November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

Landslides Susceptibility Assessment and Risk Mapping using Logistic Regression and Geographical Information System

Faten Syaira Buslima[#1], Rohayu Che Omar[#2], Rasyikin Roslan[#3], Intan Nor Zuliana Baharuddin[#4], Badariah Solemon[#5], Warishah Abdul Wahab[#6], Vignesvaran Gunasagaran[#7]

[#1,2,3,4,5,6,7]*Institute of Energy Infrastructure, Universiti Tenaga Nasional,*
43000 Kajang, Malaysia

fatensyaira@gmail.com[1], Rohayu@uniten.edu.my[2],*Badariah*@uniten.edu.my[3], warishah.wahab@uniten.edu.my[4], Intan@uniten.edu.my[5],Rasyikin@uniten.edu.my[6]

Article Info
Volume 81
Page Number: 3308 - 3315
Publication Issue:
November-December 2019

Article History

Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *16 December 2019*

Abstract:

Rapid development in the agriculture sector, land clearing, and construction have a great impact on the surface and soils structure especially in the mountainous area, for example, Cameron Highlands. These activities coupled with natural triggering factors like aspect of slope, elevation, geology, angle of slope, curvature, and rainfall may lead to serious geological hazard such as landslides. Cameron Highlands is one of the regions that is known to be susceptible to landslides. A study was carried out to classify susceptible areas and guide to the risk management. In this study, Logistics Regression (LR) using Geographical Information System (GIS) was applied to assess the susceptibility of landslides at Cameron Highlands. Ten (10) landslide contributing factors are taking into consideration including elevation, aspect, geology, slope, curvature, land use, distance from the fault, distance from drainage and road as well as rainfall. Based on the result, the LR approach obtained 82.5% landslides prediction accuracy and considered as a good result for the prediction. With the right information and updates from the landslides susceptibility map, it will assist the local authority in mitigating, treating and controlling this natural hazard at an early stage before any landslide happen.

Keywords: *landslide, logistic regression, geographical information system, area under curve*

I. INTRODUCTION

Landslides is a gravitational motion of the debris down a slope or mass of rock and earth[1].Many factors can trigger the landslides such as rainfall, increases and decreases in the groundwater level, development activities that cause disturbance and change in the slope profile, earthquakes and volcanic activity. However, it has been reported that landslides related to man-made are more than 80% cases including poor slope management practices [2].Cameron Highlands is one of the areas that are known to be susceptible to landslides [3].Rapid developments in the agriculture sector, land clearing, and construction have a huge impact on the surface and soils structure especially in the mountainous area like Cameron Highlands. This geological hazard caused great damage to the buildings,

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

properties, and communication lines [4] as well as annually thousands of fatalities and injuries [5], [6]. Thus, it is important to identify the potential landslide location to minimize the damages caused by this hazard and manage the landslide risk[3].

The identification of prone areas and risk management is conducted through landslides susceptibility assessment[7] and mapping. This assessment and mapping assists the planners or authorized personnel to plan out suitable mitigation methods in areas that are prone to landslides [3].There are two categories of susceptibility mapping methods; qualitative (direct) and quantitative (indirect)[8].Qualitative methods (direct) are subjective and heuristic and it is depending on experts' opinions [9]such as landslide inventory, where it identifies sites that are susceptible to failure. The quantitative methods, on the other hand are based on formula (data) or equations for the relationship between controlling variables and landslides.

Logistic Regression (LR) is one of the quantitative techniques that offers many advantages over landslides susceptibility assessment for example the dependent variables only can have two datasets of values (occurring or not occurring) and the predicted probability value is determined which falls within the interval of 0 and 1 [10].This method also needed fewer theoretical assumptions than discriminatory assessment.LR has been implemented for susceptibility mapping for other researchers including [11],[12],[13] and [14]. The Geographical Information System (GIS) has widely used in the assessment of natural disasters such as landslide, earthquake, erosion, flood and many more. Nowadays, GIS has been almost an important tool in determining landslide hazard and risk assessment[15].This is due to its ability to input data, manipulate big quantities of spatial data, manage data, analyze and query the inferable meaning of data. GIS also has the ability to generate spatial prediction by combining data layers according to deliberately embodied rules[16].

II. STUDY AREA

Fig. 1 represents the location of Cameron Highlands. Cameron Highlands which located in Pahang State was selected for the case study since it is considered as a landslide prone area based on the record of landslide occurrences from year 1967 to 2007 compiled by JabatanKerja Raya (JKR).

Fig. 1 Location of Cameron Highlands (Google Satellite)

III. METHODOLOGY

The workflow of this study is divided intothree phases as shown as Fig. 2.

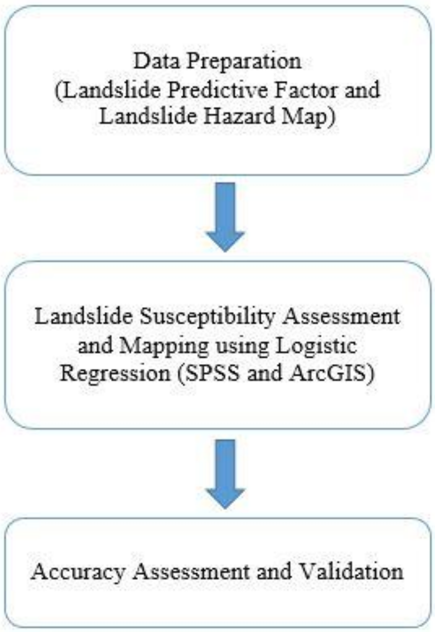

Fig.2 Workflow of the Study

A.Data Preparation

1) Landslide Predictive Factor

Many factors can trigger the landslides. In this research, ten factors were considered for the landslide susceptibility assessment. This includes elevation, aspect of slope, angle of slope, curvature, distance from fault, distance from drainage, geology, land use, distance from the road and rainfall data as shown in Fig. 3 (a) to (j).This dataset was acquired from Department of Survey and Mapping Malaysia, Department of Mineral and Geoscience Malaysia, Malaysian Meteorological Department and Federal Department of Town and Country Planning Malaysia.

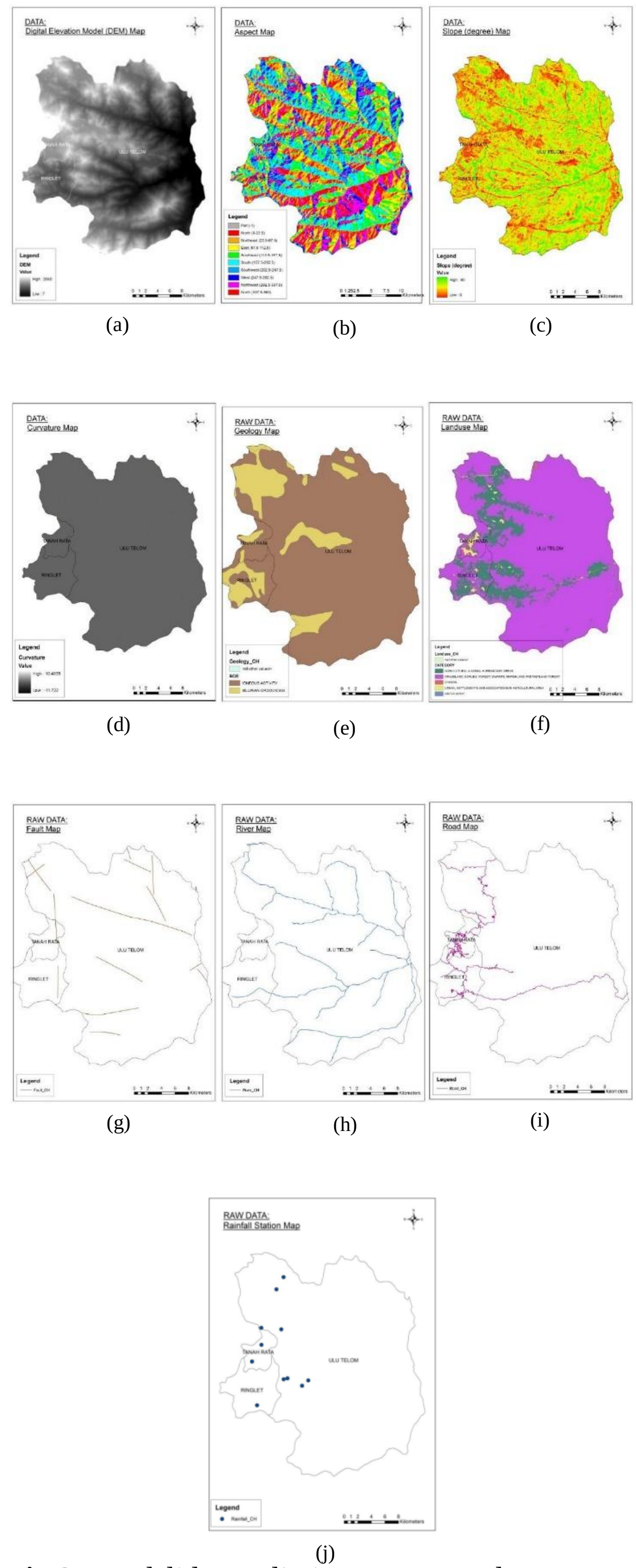

Fig.3. Landslide Predictive Factors and Classification: (a)DEM; (b) Aspect; (c) Slope; (d)

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

Curvature; (e) Geology; (f) Landuse; (g)Fault; (h) River; (i) Road and (j) Rainfall Station

A. Landslide Susceptibility Assessment using Logistic Regression, SPSS and ArcGIS

In this study, the LR will discover the best and reasonable model for describing the correlation between the existence of landslides (dependent variable) and a set of independent parameters such as angle of slope, aspect of slope and lithology. The more independent variable included, the more complete a model will be, but this will only occur when they play a significant role in determining the dependent variable. The probability of landslide occurred once can be determinedthroughEq.1;

$$P = \frac{1}{1 + e^{-z}} \qquad \text{Eq. 1}$$

where P = The probability of landslide occurrences (varies from 0 to 1 as z varies from -∞ to +∞)

$z = B_0 + B_1X_1 + B_2X_2 + ... + B_nX_n$

where $B_{i(i=1\ \text{to}n)}$ = Coefficient estimated from sample data

n = Number of independent variables / parameter

$X_{i(i=1\ \text{to}\ n)}$ = The independent variables

SPSS (a statistical software package) was used in this study to aid in producing a landslide susceptibility map based on LR method. By using backward stepwise process in SPSS, the less significant factors are eliminated from the algorithm and then, transferred to GIS software (ArcGIS) to develop the landslide susceptibility map of the study area with probability ranging from 0 to 1. In order to produce an unbiased sample representation, both presence (indicate as value 1) and absence of landslide (indicate as value 0) were used to fit the LR analysis by choosing similar number of points from non-landslide areas as sample representing the absence of landslides. These data were then used as input in the LR algorithm within the SPSS to obtain coefficients (B_i) and intercept (B_0) of the LR.

AnLR model of the study area was constructed based on ten independent variables/parameters by using the backward stepwise method. At each step, a parameter that did not perform or contribute significantly to the strength of regression was eliminated. The final regression analyses retained are only those which contribute significantly to the landslide. Both B_0 and $B_i$value from the final step of LR analysis were then transferred into ArcGIS and a landslide susceptibly map was created using Raster Calculator.

C. Accuracy Assessment and Validation

Once the landslide susceptibility map has been generated LR method, the susceptibility maps are compared with landslide hazard map generated from exact landslide location at Cameron Highland. Each factor has been applied, and the frequency ratio has been compared. The rate curves were developed and their area under curve were calculatedfor all cases. The rate describes how well the model and factors predict landslides. Therefore, the area under curve line can evaluate the pattern of prediction, the calculated index values were sorted in descending order for all cells in the study area. Then the ordered cell values were divided into 100 classes and 1.0 percent intervalswere accumulated. The

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

rates verificationresults appear as a line. The area under curve was recalculated to compare the resulting quantities as the total area is 1,which implies ideal forecast precision.

IV. RESULTS AND ANALYSIS

The SPSS Statistical Software was used to process data and estimate the regression coefficients for the whole study area. The standardized regression coefficient or t-values are expressed in unit of standard deviation where it indicates the relative contribution of each causative factor. The coefficient can be considered with 95.0 percent confidence if the absolute t-value is larger than 1.96[10]. TableI summarizes the result of LR.

TABLE 1

COEFFICIENT OF LOGISTIC REGRESSION FOR EACH FACTOR

Variables	**ß-coefficient**	**Std. Dev.**	**t-value**
Elevation	0.000429	0.000044	9.75
Slope Aspect	0.000422	0.000042	9.67
Slope Angle	0.018000	0.001000	18.00
Curvature	0.000306	0.000030	10.20
Distance to drainage	-0.141400	0.057300	-2.47
Distance to road	-6.881700	0.366300	-18.79
Landuse	-0.133800	0.060600	-2.21
Geology	-0.001870	0.000027	-69.26
Distance from fault	0.000079	0.000033	2.39
Rainfall	0.774000	0.020500	37.76

From Table I, it can be seen that there are four variables gives the negative ß-coefficient value including distance to drainage, distance to road, land use and geology. The negative value of ß-coefficient indicates that the parameter plays a negative role for land sliding and it should not be used in the next calculation stage. The final coefficient should be a positive value which significantly contributes to landslide.

A. Accuracy Assessment and Validation of the Landslide Susceptibility Map

The ß-coefficient value from the final step of LR are transferred into ArcGIS and a landslide susceptibility map of the study area is created using Raster Calculator tools. Hence, Eq.2 shows the result of LR which will be transferred to ArcGIS software.

Eq. 2

$$P = \frac{1}{1+e^{-z}}$$

where $z = B0 + B1X1 + B2X2 + \dots\dots\dots + BnXn$

$$= -193 + 0.000429X_{elevation} + 0.000422X_{aspect} + 0.018000X_{curvature} + 0.000079X_{fault} + 0.774000X_{rainfall}$$

Landslide susceptibility map using LR method has been overlaid with landslide inventory map where it shows the existing landslide locations as shown in Fig.4.

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

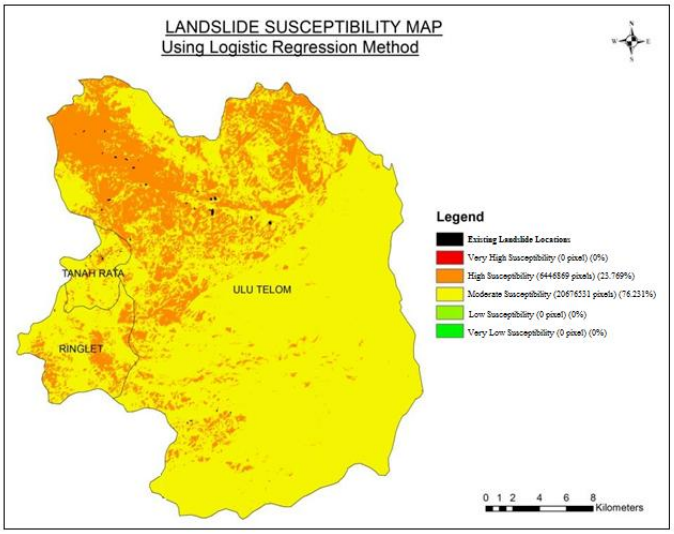

Fig. 4 The final landslide susceptibility map using LR

From Fig. 4, it is shown the area is only covered by orange area (high susceptibility class) and yellow area(moderate susceptibility class). Most of the black area (existing landslide) is laid over the high susceptibility area (orange). The results of calculation for probability of landslide occurrences are shown in Table II.

TABLE II

TOTAL AREA (IN PIXEL) AND THEPERCENTAGE OF PROBABILITY OF LANDSLIDE OCCURRENCES USING LR METHOD

Landslide Susceptibility Classes	**No of Pixel (Area)**	**%**
Very Low Susceptibility	0	0.000
Low Susceptibility	0	0.000
Moderate Susceptibility	20676531	76.231
High Susceptibility	6446869	23.769
Very High Susceptibility	0	0.009

Based on the Table II, only two classes of susceptibility have been identified in the study area; High Susceptibility and Moderate Susceptibility. High Susceptibility class involves a total area of 6446869 pixels (23.769%) while Moderate Susceptibility class involves an area of 20676531 pixels (76.231%).

The results of the assessment of landslide susceptibility were validated using existing landslide location acquired from Google Earth aerial view and field survey. At that time, there were 50 active landslides recorded and added to the inventory for validation purposes. Existing landslide inventory map which consists of existing and active landslide was overlaid with the landslide susceptibility map and the results is shown in Table III.

TABLE III.

EXISTING ACTIVE LANDSLIDE LOCATION

Landslide Susceptibility Classes	**Existing Landslide (%)**
Very High Susceptibility	0
High Susceptibility	48
Moderate Susceptibility	42
Low Susceptibility	0
Very Low Susceptibility	0

Table III represent the percentage of existing landslide happened in each susceptibility class. From the table, 48% landslides location lies

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

within the high susceptibility area and 42% landslides happened within the moderate susceptibility area which. By logic, active landslide should be lies in the high susceptibility or very high susceptibility area. If any case, an active landslide falls within the low susceptibility area, it can be concluded that the susceptibility map is inaccurate.

B.Area under Curve (AUC)

The rate curves have been created and their areas under curve have been calculated. This curve is called as area under curve. This explains how accurate one model is including the factor to predict landslide.

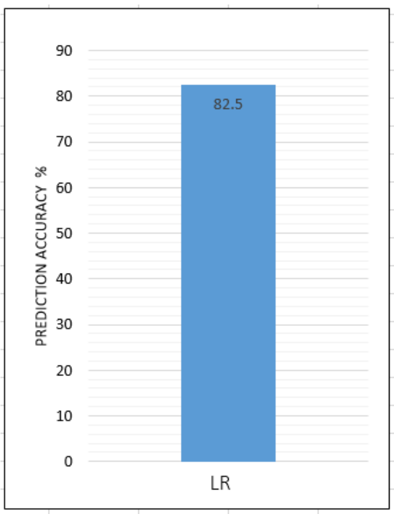

Fig.5 The area under curve method for landslide susceptibility map

Fig. 5 show the areas under curve that represent the accuracy of the prediction. In this study, the accuracy obtained for the landslide prediction is 82.5% percent.

IV. CONCLUSION

Cameron Highlands is one of the main attractions of tourists for its cold weather and various visiting places. Hence, rapid development in agriculture sector, land clearing, and construction within this district, coupled with natural triggering factors led to serious landslides hazard. A study has been carried out in order to identify susceptible areas and guide risk management. Logistic regression (LR) and geographical information system (GIS) was applied to develop landslide susceptibility maps for Cameron Highland. Ten variables have been considered for the assessment of landslide susceptibility including elevation, slope aspect, angle of slope, curvature, distance from fault, distance from drainage, geology, land use, distance from road and rainfall data. From the assessment, there are 23.8 percent area has been identified as high susceptibility area while 76.2 percent were identified as moderate susceptibility. From the validation performed, the landslide prediction accuracy obtained is 82.5% which is considered as a good prediction. Recommendation and mitigation measures to avoid landslide can be formulated based on this result. It will support the local authorities to notify and warn the public regarding the related issues.

ACKNOWLEDGMENT

This study was supported by grant FRGS/1/2018/STG08/UNITEN/02/1 for software, data collection and data processing using SPSS from 20180101YCU/20. Authors would like to thank the FEG group from IEI,iRMC and YCU, UNITEN.

REFERENCES

[1] H.Miura,“Fusion Analysis of Optical Satellite Images andDigital Elevation Model for Quantifying Volume inDebris Flow Disaster.” Remote Sens. 2019, 11, 1096 2. 2019.

[2] WSDOT (Washington State Department of Transportation), “Landslide Mitigation Action

November-December 2019
ISSN: 0193-4120 Page No. 3308 - 3315

Plan. Final Report." Washington State Department of Transportation (WSDOT), Rail Division. 2014.

[3] S. L. K. Tyng, "GIS Based Determination of Landslide Location." Dissertation. UniversitiTeknologi PETRONAS. 2014.

[4] M. A. Torres-Vera, "Spain: Natural Hazards in the Country."Reference Module in Earth Systems and Environmental Sciences. 2018.

[5] B. K.Bujang, T.Hua, F.Ali, H. B.David, H.Singh, &H.Omar, "Landslides in Malaysia: Occurrences, Assessment, Analyses and Remedition." Serdang: Universiti Putra Malaysia Publisher. 2008.

[6] H. A. Rahman, &J. Mapjabil, "Landslides Disaster in Malaysia: An Overview." Health and the Environment Journal, 2017, Vol. 8, No. 1. 2017.

[7] J. B.Nsengiyumva, G. Luo, L.Nahayo, X.Huang, &P. Cai, "Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation Model in Rwanda." International Journal of Environmental Research and Public Health. 2018.

[8] J. Mirnazari, "Evaluation of Different Techniques for Generating Landslide Susceptibility Map." Thesis. Faculty of Geoinformation and Real Estate. UniversitiTeknologi Malaysia. 2015.

[9] P. Reichenbach, M. Rossi, D. B. Malamud, M. Mihirb, &F. Guzzetti,"A review of statistically-based landslide susceptibility models."Earth-Science Reviews,Volume 180, May 2018, Pages 60-91. 2018.

[10] T. L. Ngyuen, "Landslide Susceptibility Mapping of the Mountainous Area in a Luoi District, ThuaThien Hue Province, Vietnam." Faculty of Engineering. Department of Hyrology and Hydraulic Engineering. University of Brussels. 2008.

[11] N. Intarawichian, & S. Dasananda, "A Comparative Study of Analytical Hierarchy Process and Probability Analysis for Landslide Susceptibility Zonation in Lower Mae Chaem Watershed, Northern Thailand. Suranaree J Sci Technol. 17. 2010.

[12] C.Y.Oh, K. J. Cheung &U.C. Chul, "The Comparative Research of Landslide Susceptibility Mapping Using FR, AHP, LR, ANN." Google Scholar. 2003.

[13] B. Prandhan, "An Assessment of the Use of an Advanced Neural Network Model with Five Different Training Strategies for the Preparation of Landslide Susceptibility Maps. Journal of Data Science 9(2011), 65-81. 2011.

[14] B.Prandhan, "Remote Sensing and GIS-based landslide hazard analysis and crossvalidation using multivariate logistic regression model on three test areas in Malaysia." Advances in Space Research 45 (2010) 1244-1256. 2010.

[15] C. J. Van Westen, "GIS in landslide hazard zonation: A review, with examples from the Andes of Colombia." In: M. F. Price & D. I. Heywood (eds), Mountain Environments and Geographic Information Systems, Taylor and Francis Publishers, 135-165 pp. 1994.

[16] A.Matori, A. Basith, &I. Harahap, " Study of Regional Monsoonal Effects on Landslide Hazard Zonation in Cameron Highlands, Malaysia." Arabian Journal of Geosciences - ARAB J GEOSCI. 5. 1-16. 10.1007/s12517-011-0309-4. 2011.