INTERNATIONAL JOURNAL OF GEOMATICS AND GEOSCIENCES
Volume 7, No 1, 2016

Research article ISSN 0976 – 4380

Landslide susceptibility mapping of southern part of Marsyangdi River basin, West Nepal using logistic regression method

Prakash Pokhrel, Dinesh Pathak
Central Department of Geology, Tribhuvan University, Kirtipur, Kathmandu, Nepal
dpathaktu@gmail.com

ABSTRACT

Landslide susceptibility mapping has been carried out around the Sundar Bajar to Besi Shahar area in the southern part of Marsyangdi river basin covering an area of about 221 Km^2. Logistic regression model has been used to generate the landslide susceptibility map of the area. This method assumes that the future landslides will have the same causal factors as the landslides initiated in the past. The landslide inventory map was prepared using the Google Earth images with the field identification and verification. Thirteen intrinsic causative factors and one extrinsic factor i.e. rainfall were used in analysis. Slope, aspect, curvature, elevation and relief were the Digital Elevation Model (DEM)-derived parameter whereas the thematic layers like lithology, distance to road, distance to river, lineament density, stream density, wetness index and land-use were prepared in the different GIS environment. Receiver operating characteristic (ROC) curves in SPSS was used to validate the model using the training and validation landslide data. The area is classified into five classes i.e. very low, low, medium, high and very high susceptibility using the ROC curve of the validation data. Most of the area belongs to the very low to low susceptibility class. The prepared susceptibility map can be used for the landslide hazard management.

Keywords: Marsyangdi river basin, Landslide susceptibility, Logistic regression, Receiver operating characteristic curve

1. Introduction

In Nepal, landslides are the very common natural hazards due to the steep and rugged topography, improper use of land cover, tectonic activity and adverse climatic condition. The International Landslide Centre of the University of Durham (2007) recorded that the most seriously affected country was China with 695 landslide-induced deaths, followed by Indonesia (465), India (352), Nepal (168), Bangladesh (150) and Vietnam (130). In order to reduce the damage caused by the landslide activity, a proper study of such phenomena is needed, one of which is the preparation of the landslide susceptibility map. Worldwide 89.6% of the fatalities were caused by landslides triggered by intense and/or prolonged precipitation (Petley, 2008). Other triggering processes were construction, mostly undercutting of slopes (3.4%), mining and quarrying (1.8%) and earthquakes (0.7%), while no cause would be identified for 3.4% of the landslides. Landslide hazard or susceptibility zonation is defined as the mapping of areas with an equal probability of occurrence of landslide within a specified period of time (Varnes, 1984). It demonstrates the spatial and temporal occurrence of future landslides in terms of probability. A region is considered to be susceptible to landslide when the terrain conditions at the site are similar to those in the region where a slide has occurred (van Westen, 2000). Landslide susceptibility is the tendency or preference of an area to undergo landsliding. It is a function of the degree of inherent stability of the slope as indicated by the factor of safety (Crozier and Glade, 2005).

The terminology used in the landslide management study has different practice in different country and among the researchers. For example use of the term 'landslide hazard' and 'landslide susceptibility'. In landslide hazard assessment practice, the term 'landslide susceptibility mapping' is addressed without considering the extrinsic factors in determining the probability of occurrence of a landslide event (Dahal et al., 2008). But after appreciating guidelines prepared by JCT-1 (Fell et al., 2008), rainfall distribution is also considered as one of the parameters for landslide susceptibility analysis.

Logistic regression analysis (LRA) was introduced by (Cox, 1958) and is used to investigate a binary response from a set of measurements. The technique which regresses a dichotomous dependent variable on a set of independent variables that can be interval, dichotomous or categorical is widely used to predict success or failure of a process based on a set of measurements. Instead of using a linear relationship between the independent variables and the responses, a logarithmic model relationship is used. Logical regression allows forming a multivariate regression relation between a dependent variable and several independent variables that might affect the probability of the searched situation. If the searched variable is a dichotomous outcome logistic regression has been shown to be best prediction (Dai et al., 2001).

In landslide susceptibility investigation, the response is the presence or absence of landslides in each mapping unit and the independent variables are the set of n environmental factors X_1, X_2, …….., X_n available for each mapping unit. In LRA, the quantitative relationship between the occurrence of landslides in a mapping unit (dependent variables and its dependency on the set of several variables (independent or predictor variable) is expressed as

$$\ln (p/1-p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots.. + \beta_n X_n$$

Where, p= probability of an event occurring

X_i= set of factors

B_j= regression co-efficient

LRA involves fitting above equation to the data and then expressing the probability of the presence or absence of landslides in each mapping unit. The relative contribution of each mapping unit to logistic function can be obtained by looking at the significance of each regression parameter. Logistic regression has widely adopted for predicting binary responses for landslide prediction. One of the advantages of LRA compared to other method is the range of resultant values. While using linear regression to predict a categorical variable of 0 or 1, the predicted values are transformed into probabilities which fall between 0 and 1. The average of the fitted values agrees with the average of the dependent variable in the data being modeled. Therefore many studies have used logistic regression for the assessment of landslide (such as Eeckhaut et al., 2006; Lei and Jing-Feng, 2007; Chen and Wang, 2007; Akgün and Bulut, 2007; Long, 2008; Chauhan et al., 2010; Dahal et al., 2012).

2. Study area

The Sundar Bajar-Besi Shahar area is located in the western part of Nepal in Lamjung district and covers an area of about 221 km^2 belonging to part of Marsyangdi river basin (Figure 1). This area is characterized by the elevated mountains and the deep river valleys. The elevation of the area ranges from about 500 m in the south to about 2000 m in the north. Marsyangdi River is the main river flowing through the area whereas Dordi Khola and Paudi Khola are its main tributaries.

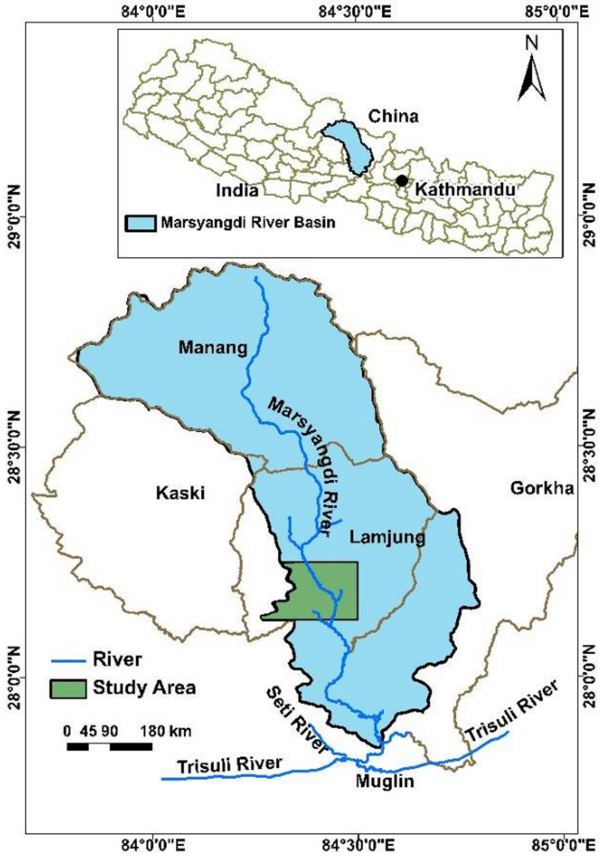

Figure 1: Study area

Geologically, this area lies in the Lesser Himalayan terrain in the northern flank of the Gorkha-Kuncha Anticlinorium (counter part of Mahabharat Synclinorium in central Nepal) near to the MCT zone (Pokhrel, 2015).

3. Methodology

There are various methods of landslide susceptibility mapping. Landslide susceptibility mapping employs both direct and indirect methods of assessment. In direct method, the area is categorized based on density of recorded landslide. In the second method, several factors which contribute to the initiation of slope failure are considered. Landslide susceptibility assessment in parts of the Marsyangdi river basin was carried out based on the primary data collected from the field as well as the information extracted from the Google Earth images. All these information were integrated in the Geographic Information System (GIS) database and used to prepare several thematic layers like DEM, slope, aspect, curvature, relief, land use, stream density, lineament density, wetness index, sediment transport index, geological and landslide inventory maps. Digital elevation model (DEM) was prepared from the contours available in the digital topographic map produced by the Department of Survey, Government of Nepal. The thematic layers were further analyzed and processed in GIS, Statistical Package for Social Science (SPSS) and Excel in order to generate landslide susceptibility map with different susceptibility zones.

Most of the landslides were identified through the traverses along main roads, foot trails and valleys. Some landslides were identified from Google Earth images at more remote location. The landslides identified on the image are probably of rather recent origin as older landslides will are less visible. The landslide inventory map prepared with the distribution of both training and validation data are shown in Figure 2.

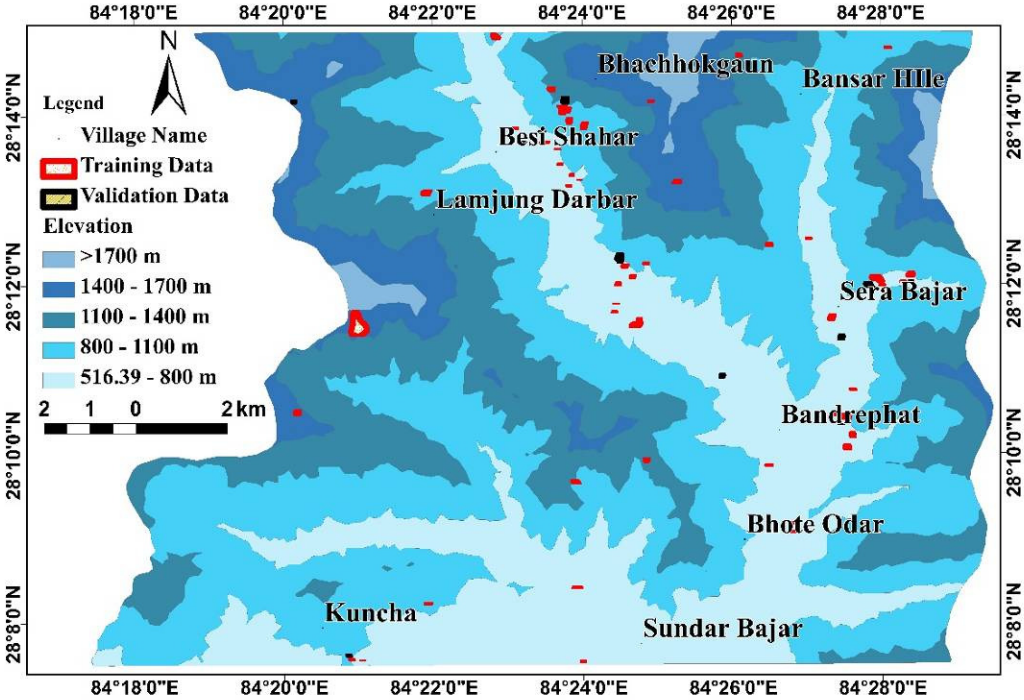

Figure 2: Landslide inventory map showing both training and validation data of landslide and elevation map of the study area in background

About 20% of total identified landslides were randomly selected as a validation data and remaining were treated as training data. The landslides used as training data and validation data fairly represented the different geological and geomorphic units as well as elevation ranges within the Marsyangdi river basin. It was necessary to ensure that the analysis covered the varying geological, geomorphological, climatic and land use condition existing in the basin.

The observed landslide map was converted into the numerical map by attributing the value 1 to observed landslide areas and the value 0 to the other areas. The DEM with a pixel size of 20 m*20 m was prepared and used for the derivation of other factors map. Different landslide factors maps used in the landslide susceptibility mapping in the present study is shown in Figure 3. The numerical values to SPSS were exported for each pixel of the map for subsequent analysis. Then the estimated results were imported into a raster map and finally the map was classified into the different susceptibility classes. Hosmer-Lemeshow test for the goodness of fit, Cox and Snell R^2 and Negelkerke R^2 for reliability of model, and Receiver Characteristic (ROC) curve for validation and classification of map into different susceptibility zones were used for the analysis of data.

4. Results and discussion

The analysis of the training data of the landslide shows value of Hosmer-Lemeshow test is less than 0.05. Higher the value for this test, the higher is the significance of the work. The Hosmer-Lemeshow test for the validation data of landslide showed that the goodness of fit of the equation can be accepted, since the value is greater than 0.05 (0.106). In the Hosmer-Lemeshow test, if significance of Chi-Square value is less than 0.05, then the parameters used in the analysis miss-specifies the model. The value of Cox and Snell R^2 and Negelkerke R^2

gives the reliability of the model. Higher is the value of R^2, more the model is reliable and vice-versa. The Negelkerke R^2 value obtained from the regression analysis of training data is 0.309 and 0.432 for validation data. The R^2 values for the training data of landslide especially by Negelkerke R^2 showed that the independent variables can explain the dependent variable in an acceptable way.

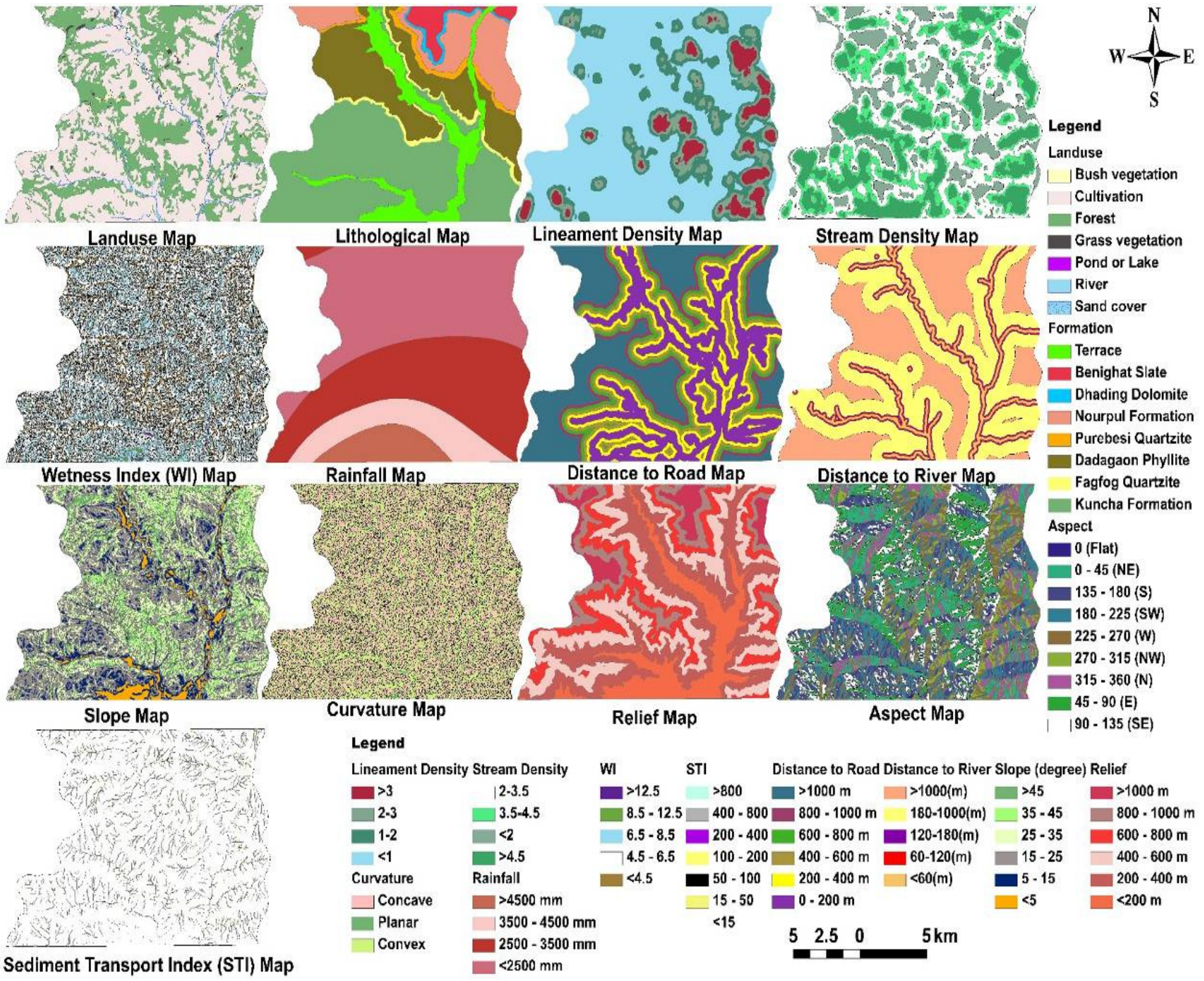

Figure 3: Different landslide factor maps used in the susceptibility analysis

A Receiver Operating Characteristic (ROC) curve is the plot of the true positive rate against the false positive rate of different possible cut points of a diagnostic test. The ROC curve approach is used to analyze the prediction accuracy of the proposed models. The ROC curve gives the area under the curve and it is a measure of goodness of fit. The ROC curve of the training data of landslide account the succession rate of the catchment model and the ROC curve of the validation data of landslide account the prediction rate obtained from the logistic regression. The ROC area under curve value for the model with the training data of landslide shows the good succession rate of 89.8% (0.898). In this analysis, the test variable used is the predicted probability value obtained from the logistic regression of the validation data of landslide and state variable used is the training data of landslide with the value of 1. The larger values of the test result variables indicate stronger evidence for a positive actual state. The positive actual state is 1. The probability value given in each of the classes of parameter is due to the relative contribution of the each of the parameter.

From the logistic regression modeling of this study area, the following logistic regression equation was obtained:

Log (p / (1-p)) = -64.568 + (2.02 * Elevation) – (2.23 * Slope) – (16.58 * Aspect) + (0.513 * Curvature) + (4.31 * Distance to River) + (12.02 * Lineament Density) + (1.74 * Stream Density) + (11.70 * Relief) + (0.592 * Rainfall) + (1.84 * Wetness Index) + (6.97 * Sediment Transport Index) – (0.899 * Lithology) + (11.75 * Landuse)

The ROC area under curve value for the randomly selected validation data of landslide is 0.853. This value indicates the very high prediction rate of about 85.3% for the probability of occurrence of the landslide. In this analysis the test variable used is the predicted probability value obtained from the logistic regression of training data and the state variable used is validation data of landslide with value of the state variable 1. The prediction rate of model using ROC curve is shown in the Figure 4.

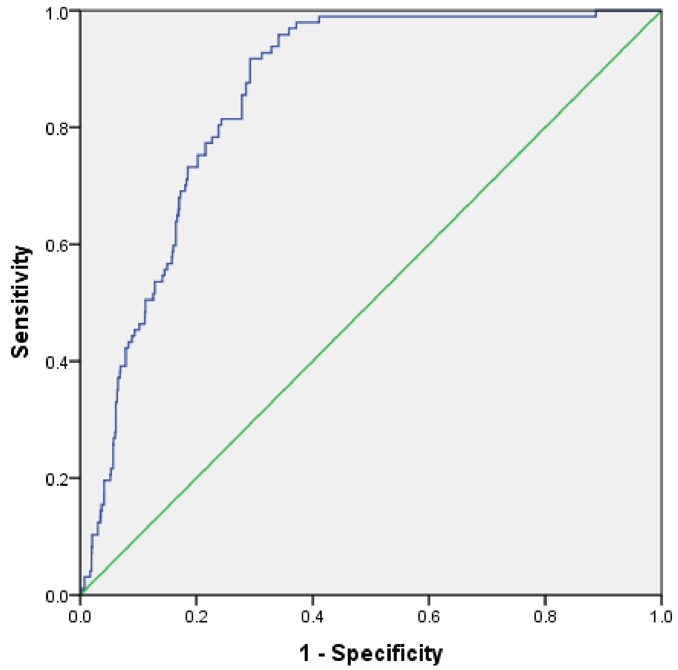

Figure 4: ROC area under curve for prediction rate of model (85.3%)

The predicted probability value obtained from the logistic regression of the training data of landslide were considered to prepare the landslide susceptibility zonation map of the study area. Considering the prediction rate of logistic regression from ROC curve of the study area (Figure 4), five classes of the landslide susceptibility were defined. The break values used in the classification of the landslide susceptibility zones are, less than 40% for Very Low Susceptibility zone, 40% to 60% for Low Susceptibility zone, 60% to 80% for Medium Susceptibility zone, 80% to 90% for High Susceptibility zone and 90% to 100% for Very High Susceptibility zone. Thus obtained landslide susceptibility zones for the study area is shown in Figure 5.

The higher susceptibility classes generally follow the river and road section. The rivers continuously affect the side slope through bank cutting while the road construction in the hilly region makes the natural and fragile mountain slopes more unstable (Pathak, 2014). In addition, the cultivated areas on slope are more susceptible to landslide occurrence.

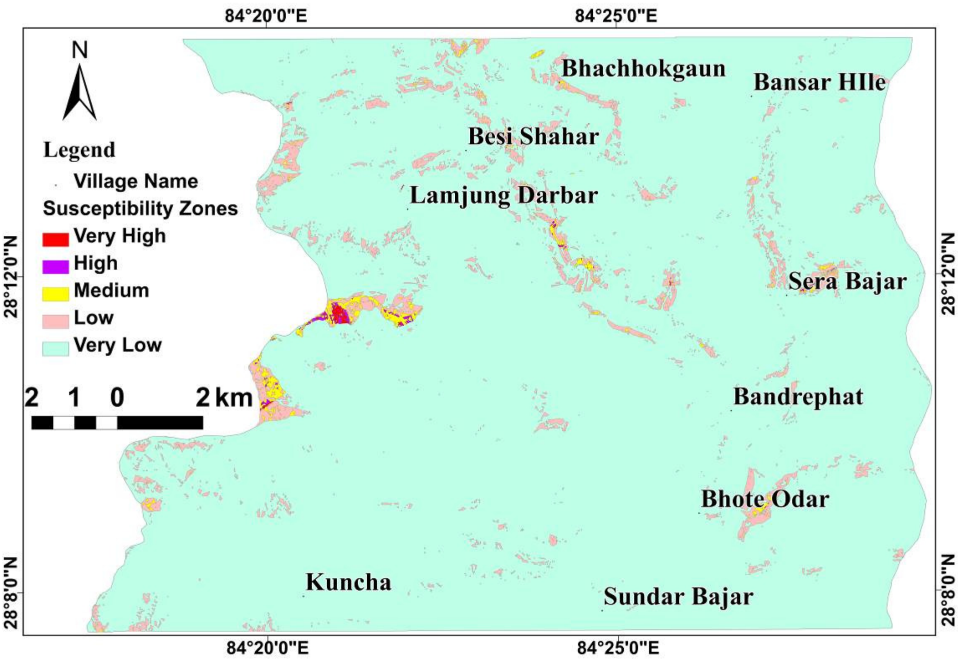

Figure 5: Landslide susceptibility zones for the study area

5. Conclusion

The ROC curve between the predicted probabilities of landslide occurrence from the training data shows the good succession rate of about 89% and from validation data shows the good succession rate of about 85%. The R^2 value according to the Nagelkerke R^2 is 0.309 (>0.05) suggesting the reliability of the model. Kuncha Formation consisting pelitic and psammatic schist is most susceptible to landslide than any other lithological domain in the study area. The cultivated area shows the highest landslide susceptibility index (LSI) among various land use classes. Smaller value of wetness index (WI) shows the higher value of LSI. Area nearer to the stream and lineament are more susceptible than other. Distance to road and distance to river also shows the similar trend. More than 90% of the study area falls in very low to low landslide susceptibility zone.

Acknowledgement

This paper is based on research financed under the Climate Change Research Grants Program implemented by the Nepal Academy of Science and Technology. The Program is part of the Mainstreaming Climate Change Risk Management in Development project. This project is a component of Nepal's Pilot Program for Climate Resilience and is executed by the Ministry of Science, Technology and Environment, (Nepal), financed by the Climate Investment Funds, administered by the Asian Development Bank with technical assistance from ICEM, METCON and APTEC.

6. References

1. Akgün, A., and Bulut, F., (2007), GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region, Springer Verlag, Environmental Geology, 51, pp 1377-1387.

2. Chauhan, S., Sharma, M., and Arora, M.K., (2010), Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model, Springer-verlag, landslide, 7, pp 411-423.

3. Chen, Z., and Wang, J., (2007), Landslide hazard mapping unit using logistic regression model in Mackenzie valley, Canada, Natural Hazards, 42, pp 75-89.

4. Cox, D.R., (1958), The regression analysis of binary sequences, Journal of the Royal Statistical Society, Series B (Methodological), 20, 215-242.

5. Crozier, M.J., and Glade, T., (2005), Landslide Hazard and Risk: Issues, Concepts and Approach. In: Landslide hazard and risk, Glade, T., Anderson, M., and Crozier, M., Chichester, Wiley, pp 1-40.

6. Dahal, R.K., Hasegawa, S., Bhandary, N.P., Poudel, P.R., Nonomura, A., and Yatabe, R., (2012), A replication of landslide hazard mapping at catchment scale, Geomatics, Natural Hazard and Risk, pp 1-32.

7. Dahal, R.K., Hasegawa, S., Nonomura, A., Yamanaka, M, Masuda, T., and Nishino, K, (2008), GIS based weight of evidence modeling of rainfall-induced landslide in small catchments for landslide susceptibility mapping, Environmental Geology, 54, pp 314-324.

8. Dai, F.C., Lee, C.F., and Xu, Z.W., (2001), Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong, Environmental Geology, 40(3), pp 381-391.

9. Eeckhaut, M.V.D., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., and Vandekerckhove, L., (2006), Prediction of landslide susceptibility using rare events logistic regression: A case study in the Flemish Ardennes (Belgium), Science Direct, Geomorphology, 76, pp 392-410.

10. Fell, R., Corominas, J., Bonard, C., Cascini, L., Leroi, E., and Savage, W.Z., (2008), Commentary guidelines for landslide susceptibility, hazard and risk zoning for land use planning, Engineering Geology, 102(3–4), pp 85–98.

11. International Landslide Centre of the University of Durham, (2007), In: Long, N.T., 2008, Landslide susceptibility mapping of the mountainous area in a Loui district, Thiua Thien Hue province, Vietnam, PhD Thesis, Vrije Universiteit Brussel, Publications of Faculty of Engineering, Department of Hydrology and Hydraulic Engineering, 255.

12. Lei, Z., and Jing-Feng, H., (2007), GIS-based logistic regression method for landslide susceptibility mapping in regional scale, Journal of Zhejiang University Science, 12, pp 2007-2017.

13. Long, N.T., (2008), Landslide susceptibility mapping of the mountainous area in a Loui district, Thiua Thien Hue province, Vietnam. PhD Thesis, Vrije Universiteit Brussel, Publications of Faculty of Engineering, Department of Hydrology and Hydraulic Engineering, 255.

14. Pathak, D., (2014), Geohazard assessment along the road alignment using remote sensing and GIS: Case study of Taplejung-Olangchunggola-Nangma road section, Taplejung district, east Nepal, Journal of Nepal Geological Society, 47, pp 47-56.

15. Petley, D.N., (2008), The global occurrence of fatal landslides in 2007, Geophysical Research abstract, EGU General Assembly 2008, 10(3).

16. Pokhrel, P., (2015), Geological structure and landslide susceptibility analysis using the logistic regression method from Sundar Bajar to Besi Shahar Area. Master Thesis, Central Department of Geology, Tribhuvan University, 87.

17. Van Westen, C.J., (2000), The modeling of landslide hazards using GIS, Survey in Geophysics, 2, pp 241-255.

18. Varnes, D.J., (1984), International Association of Engineering Geology commission on landslides and other mass movements on slopes: Landslide hazard zonation: a review of Principles and Practice, UNESCO, Paris, 63.