EasyChair Preprint

№ 9449

Improved Landslide Susceptibility Mapping Using Statistical MLR Model

K C Niraj, Ankit Singh and Dericks Shukla

December 11, 2022

Improved Landslide Susceptibility mapping using statistical MLR model

Niraj K.C
Ph.D. Scholar, School of Civil and Environmental Engineering (SCENE), Indian Institute of Technology (IIT)-Mandi, Himachal Pradesh, India,
d18058@students.iitmandi.ac.in

Ankit Singh
Ph.D. Scholar, School of Civil and Environmental Engineering (SCENE), Indian Institute of Technology (IIT)-Mandi, Himachal Pradesh, India,
d21111@students.iitmandi.ac.in

*Dericks Praise Shukla**
Associate Professor, School of Civil and Environmental Engineering (SCENE), Indian Institute of Technology (IIT)-Mandi, Himachal Pradesh, India,
dericks@iitmandi.ac.in

***Abstract*— In this study, the accuracy of the Landslide Susceptibility Maps (LSMs), prepared using the GIS-based statistical multiple linear regression (MLR) model, is improved by incorporating time-dependent factors. The LSM is prepared to identify landslide-prone areas based on the 108 training landslide points and 10 causative factors. The accuracy is evaluated with the area under the curve (AUC) using 56 testing landslide points. The results showed that landslides were more likely to occur at closer distances to the road, in areas having severely fragmented and jointed strata. The prepared LSM shows that, when annual mean NDVI (Normalized Difference Vegetation Index) is considered as a causative factor, areas with dense and sparse vegetation were less likely to experience landslides, whereas locations that were classified as barren land were more prone to landslides. The accuracy (AUC) of the LSM produced using the MLR model was 86% when NDVI is not considered and it increases to 92% when NDVI is considered as the causative factor. We, therefore, suggest using mean NDVI as a landslide causative factor in the landslide predictive model while working in dense vegetation and tropical areas.**

Keywords*— *Landslide Susceptibility Map (LSM); multiple linear regression (MLR) model; NDVI; Causative factors; AUC

I. INTRODUCTION

In recent years, the preparation of causative factors (thematic layers), landslide inventory, and identifying landslide-prone areas using GIS-enabled landslide predictive models have become possible because of the availability of several Remote-Sensing datasets [1;2]. Most of these landslides susceptible areas are referred to as crucial geomorphic processes, which often play an essential role in the landscaping of hilly areas [2]. Landslides typically occur in steep locations and are hidden by dense vegetation, making it difficult to map them. GIS-based statistical models are also frequently used to create landslide susceptibility maps (LSMs) using remote sensing data and thematic layers as the causative factors, such as slope, aspect, soil, lithology, NDVI (Normalized Difference Vegetation Index), land cover, distance to drainage, precipitation, distance to fault and distance to road, etc [4;1;5]. The efficiency of GIS-based statistical models for landslide prediction will vary depending on how well we prepared the landslide inventory and which thematic layers we took into consideration [1;2;3;5]. GIS-based statistical models, such as analytical hierarchy process (AHP), multiple linear regression (MLR) weighted linear combination (WLC), frequency ratio (FR), information value (IoV), etc., have been used to determine the relationships between the locations of the detected landslides and these causative factors [2;3;4;5]. The MLR is a multivariate statistical method commonly used for estimating the best-fitted line among the multiple causative thematic layers to map the Landslide susceptibility [6;15]. It assesses how landslide susceptibility changes with a change in the standard deviation of independent and predictor variables. Thus, in this work MLR, a multivariate statistical method is used instead of a bivariate method.

Moreover, many of these landslide susceptibility maps have been studied using MLR models, and the results have been validated using the under-the-curve (AUC)-ROC (Receiver operating characteristic) curve [7;8]. Also, several landslide susceptibility mapping techniques are used with and without considering NDVI as a causative factor for landslide susceptibility mapping [9;10]. Therefore, the objective of this study is to understand how NDVI impacts in multiple linear regression (MLR) statistical landslide model for preparing landslide susceptibility maps in tropical climates and areas with substantial vegetation.

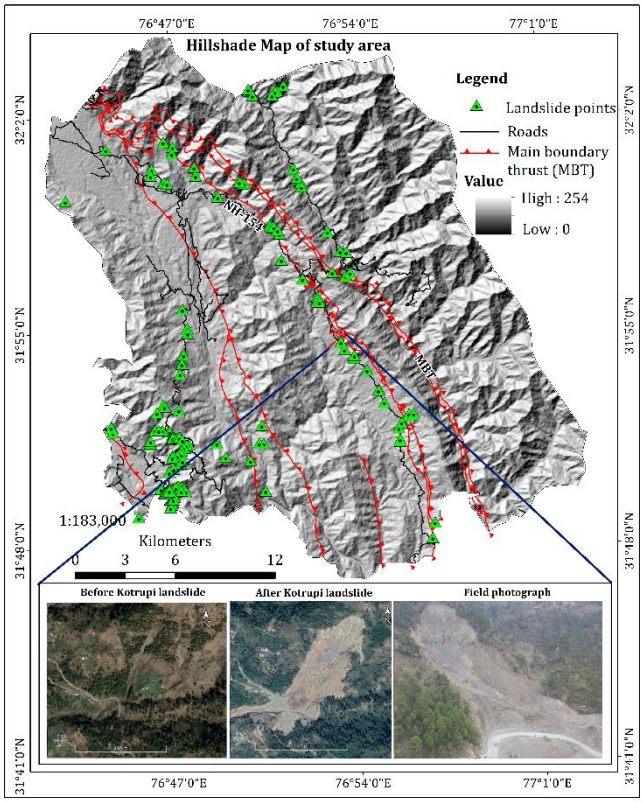

Figure 1: The study area map is drawn on hill-shade showing the presence of thrusts and the landslide points like Kotrupi landslide in this area.

II. Study Area

In this study, the Joginder Nagar area of Mandi district Himachal Pradesh is selected to assess the effect of NDVI on the MLR model for LSM of this area. Numerous landslides have occurred in the Joginder Nagar area in the past and present due to heavy rain, earthquakes, and road construction [14].

The Kotrupi landslide, whose run-off zone is enclosed by a red circle (figure 1), is the catastrophic landslide that occurred on August 13, 2017, in the Joginder Nagar area [15] that passes by near the Main Boundary Thrust (MBT) as shown on the Hill-shade map (figure 1). The yearly average rainfall in the Joginder-Nagar varies greatly from place to place, ranging from 700 mm to more than 2000 mm at Joginder Nagar [14,16]. The most common rock types in and around the Joginder-Nagar include sandstone, red and purple shale, mudstones, and dolomite of the Siwalik group, as well as dolomite of the Shali formation [14,17]. The area is situated in a thrust contact between the Siwaliks and the Shali group of rocks [17,18]. These rocks are weak, and when they are subjected to thrust displacement, they are prone to landslides [16;17;18;19].

III. Data and Software Used

In this work we used Landsat 8 OLI ETM+ dataset of November 15, 2015, to calculate the NDVI (Normalized difference vegetation index) of the study area. We used ASTER-DEM with a spatial resolution of 30 meters to determine the elevation, slope, aspect, curvature stream network, and SPI (stream power index) map. Additionally, data about lithology, lineament, and geomorphology were obtained from the Bhukosh portal and were modified based on the fieldwork. Moreover, the initial landslide inventory was obtained from GSI and was updated based on a field visit and high-resolution data from Google Earth and Planet images. Landslide susceptibility maps (LSMs) were created and analyzed using ArcMap 10.5, ENVI 5.3, and R-studio 4.2.0.

IV. Methodology

The landslide data (164 locations) was divided into training (108 locations i.e. 70%) and testing (56 locations i.e. 30%) for carrying out the analysis. We took 10 causative factors namely aspect, curvature, distance to road, distance to stream, distance to lineament, elevation, slope, SPI,

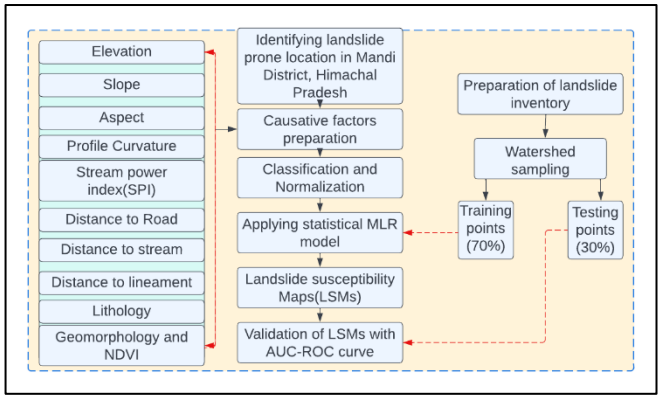

Figure 2: Flowchart of the procedure used for this study.

geomorphology, and lithology that were considered time-independent as these factors do not change a lot within a year. Then we took the NDVI data which is a time-dependent parameter to assess its effect on LSM. In this study, the Landsat 8 OLI image is first radiometrically and atmospherically corrected using the FLAASH setting, and then the NDVI is determined [11]. The flowchart showing the method used is shown in figure 2.

4.1 Thematic Layer preparation

The thematic (or causative) layers are first systematically classified and normalized using the natural break classification approach to develop landslide susceptibility maps (LSMs). The aspect was generated using DEM data, which was then divided into 10 classes using the natural break classification method: flat, north, northeast, east, southeast, south, southwest, west, northwest, and north (figure 3a). Curvature, which is also the surface profile, is a significant factor in landslide causation [12]. Convex, planar, and concave curvature are the three different types of curvature. In this study, ArcMap 10.5 curvature tool used DEM data to determine the curvature of the study area. The distance to the

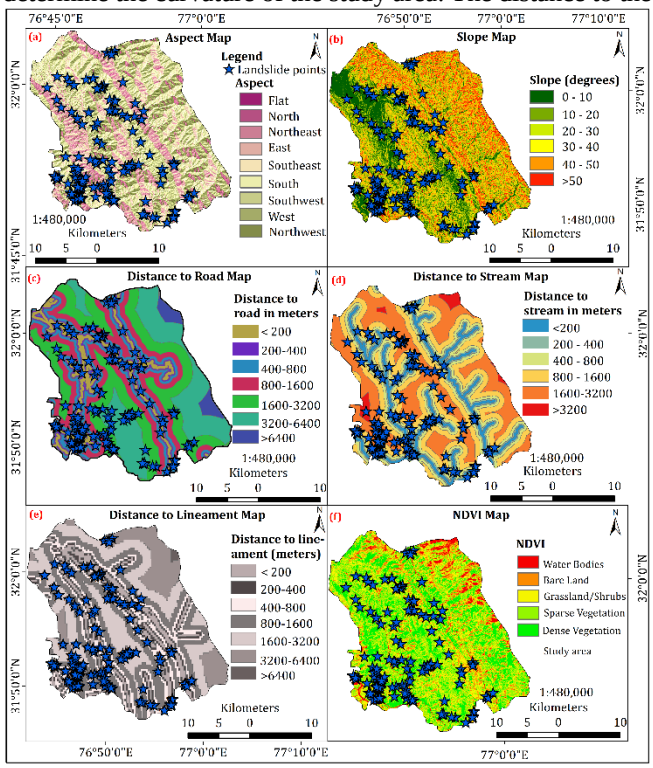

Figure 3: The maps of some causative factors used in this study.

road map was taken to emphasize the impact of road proximity on landslides in the study area. The road network was buffered into intervals of 200m, 200-400m, 400-800m, 800-1600m, 1600-3200m, 3200-6400m, and >6400m using ArcMap 10.5 multiple ring buffer tool (figure 3c). The multiple-ring buffer method was used to divide buffers into six distinct categories to measure how much the streams affected the slopes. As a result, the DEM is used first to delineate the stream network, which was buffered to <200m, 200-400m, 400-800m, 800-1600m, 1600-3200m, and >3200m using the multiple ring buffer method to determine the distance to a stream of the study area (figure 3d). The lineaments mapped were buffered to obtain distances to the

lineament map of <200m, 200-400m, 400-800m, 800-1600m, 1600-3200m, 3200-6400m, and >6400m (figure 3e). The elevation map was divided into 4 classes using the natural break classification method: <1000 meters, 1000-1500 meters, 1500-2000 meters, and >2000 meters. In this study, the slope was calculated using DEM and further divided into 6 classes using the natural break classification method: <10°, 10°–20°, 30°–40°, 40°–50°, and >50°. The stream power index (SPI), which measures the stream capacity for erosion also affects the occurrence of landslides. The SPI is calculated as SPI = AS tan(β), where AS is the area of a particular catchment and β is the local slope gradient expressed in degrees [12]. Four categories were created from the SPI map of the study area: <-4.5, -4.5 to -0.2, -0.2 to 2.2, and >2.2. In the study area, there are 12 lithological and geomorphological formations. The Normalized Difference Vegetation Index (NDVI) is frequently used to calculate the amount of vegetation present in any region [11;13]. The value of NDVI varies from -1 to 1, and we classified NDVI into five groups i.e. waterbodies (NDVI: -1 to 0), barren land (NDVI: -0.1 to 0.1), grassland (NDVI: 0.1-0.4), sparse vegetation (NDVI: 0.4-0.7), and dense vegetation (NDVI: >0.7) as shown in figure 3f.

4.2 Multiple linear regression (MLR)

The mathematical expression for the multiple linear regression, which was used in this study, shows how the susceptibility of landslides increases as the independent variables and predictors standard deviation changes [14]. The MLR value for each factor class derived is taken into consideration as the input for additional regression analysis [6]. An equal number of non-landslide points were obtained randomly [6]. Non-landslide points were counted as 0 for analysis purposes while landslide points were counted as 1 [6;14]. Additionally, R studio 4.2.0 software was used to carry out the regression analysis. The following mathematical equation was used to get the LSI [6;14].

$$LSI = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \ldots\ldots + \beta_n x_n \quad (1)$$

where β_0 is the intercept while β_1, β_2, etc is the coefficient value of each factor class, and x1, x2 is the causative factor.

4.3 Accuracy Assessment

We used the area under the curve (AUC)-ROC (Receiver operating characteristic) to assess the accuracy of the LSM prepared. If the AUC value is more than 0.90, then the LSM map is considered to be highly accurate, and less accurate if the AUC value is 0.5 or lower [15;16]. Also, AUC values above 0.7 are generally regarded as satisfactory in many landslide studies.

V. Results and Discussion

5.1 Landslide susceptibility maps without NDVI layer

The landslide susceptibility maps (LSMs) were prepared using 11 causative factors (some causative factors are shown in figure 3) and 108 training landslide points based on MLR models. The predicted LSMs were divided into five classes by the natural break classification method: very low, low, medium, high, and very high. The MLR model was used to determine the coefficient and intercept values for each layer, which were then used to prepare the LSI.

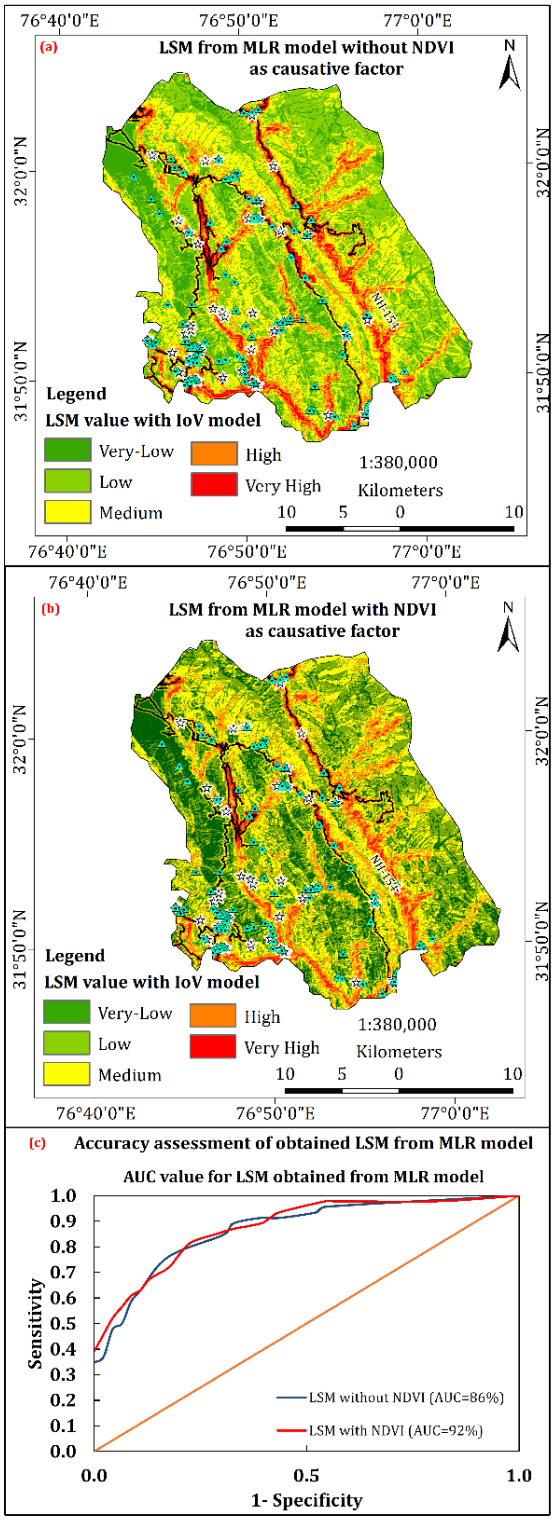

Figure 4: LSM maps prepared using MLR model (a) with and (b) without NDVI as causative factors. (c) The AUC-ROC curves for both are shown.

As a result, the following mathematical expression is used to calculate the LSI (3).

$$Z = -0.1788 + 0.1069 * Asp + (-0.0059) * Cur + 0.0357 * Dtl + 0.454 * Dtr + 0.242 * Dts + 0.242 * Ele + (-0.0718) * Geo + 0.1443 * Lith + 0.4112 * Sl + (-0.2077) * SPI \quad (2)$$

The LSM obtained using the MLR model ranges from very low (-0.21 to -0.19), low (019 to 0.35), medium (0.35 to 0.51), high (0.51-0.71), and very high (0.71-1.36) as shown in figure 4a. Where high and very high susceptibility areas made up 20% of the total area. The 56 landslide testing points were used for calculating the AUC-ROC curves. Hence, the accuracy of the landslide susceptibility map (LSM) was obtained to be 86% while using the MLR model when NDVI is not used as shown in figure 4c.

5.2 Landslide susceptibility maps with NDVI layer

The results showed that upon the inclusion of NDVI as a causative factor class the distribution of the landslide-susceptible classes changed significantly (figure 4b). In the case of LSI obtained using MLR methods the high and very high landslide-susceptible zones consisted of only 23% of the study area (figure 5B). However, after taking NDVI as a causative factor 21% of the areas consisted of high and very high susceptible zones. The LSM obtained using the MLR model ranges from very low (-0.18 to -0.24), low (024 to 0.40), medium (0.40 to 0.57), high (0.57 to 0.79), and very high (0.79-1.56) as shown in figure 4. The accuracy of the landslide susceptibility map (LSM) was obtained to be 92% while using the MLR model when NDVI is used (figure 4c).

5.3 Comparison

The findings showed that distances closer to the road had a higher likelihood of having landslides than those farther away from it. Furthermore, the LSMs demonstrate that locations with sparse and thick vegetation were less susceptible to landslides, but areas that were classified as barren land were more vulnerable to landslides when the NDVI layer is used as a causative factor. The fact that vegetated areas are less prone to landslides implies that the trees have an efficient soil-holding capacity that can prevent landslides. Additionally, the AUC-ROC curve illustrates how accuracy suddenly increases when we use NDVI as one of the causative factors in the preparation of LSMs. So, without NDVI as a causative factor, the AUC values for LSM obtained using MLR was 86%, which increased to 92% when the NDVI layer is considered. This clearly shows the effect of NDVI in the preparation of LSM using MLR statistical landslide model. But the choice of time for taking NDVI data needs to be explored more.

VI. Conclusion

In this study, the relative importance of classes and individual causative factors, the weightage and percentage of pixels in a class to the total pixels were extracted by integrating 11 causative factors, including aspect, curvature, distance to road, distance to stream, distance to lineament, elevation, slope, SPI, geomorphology, lithology, with/ without NDVI maps. The LSM is prepared using the statistical landslide multiple linear regression (MLR) model for the Joginder Nagar area. Only 20% of the study area was comprised of the high and very high landslide risk zones when LSM is prepared using the MLR model without NDVI. But after using NDVI as a causative factor, 23% of the area falls under high and very high susceptible zones. The AUC of the LSM improved significantly from 86% when NDVI was not considered to 92% when NDVI was considered a causative factor.

References

[1] Gupta, S. K., Shukla, D. P., & Thakur, M. (2018). Selection of weightages for causative factors used in preparation of landslide susceptibility zonation (LSZ). Geomatics, Natural Hazards and Risk, 9(1), 471-487.

[2] Gupta, S. K., & Shukla, D. P. (2022). Effect of scale and mapping unit on landslide susceptibility mapping of Mandakini River Basin, Uttarakhand, India. Environmental Earth Sciences, 81(14), 1-21.

[3] Conforti, M., Aucelli, P. P., Robustelli, G., & Scarciglia, F. (2011). Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Natural hazards, 56(3), 881-898.

[4] Kumar, D., Thakur, M., Dubey, C. S., & Shukla, D. P. (2017). Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. Geomorphology, 295, 115-125.

[5] Shukla, D. P., Gupta, S., Dubey, C. S., & Thakur, M. (2016). Geo-spatial technology for landslide hazard zonation and prediction. Environmental applications of remote sensing, 281-308.

[6] Mulder, H. F. H. M., & Van Asch, T. W. J. (1987). Quantitative approaches in landslide hazard analyses. Travaux de l'Institut de Géographie de Reims, 69(1), 43-53.

[7] Yariyan, P., Zabihi, H., Wolf, I. D., Karami, M., & Amiriyan, S. (2020). Earthquake risk assessment using an integrated Fuzzy Analytic Hierarchy Process with Artificial Neural Networks based on GIS: A case study of Sanandaj in Iran. International Journal of Disaster Risk Reduction, 50, 101705.

[8] Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112(1-2), 42-66.

[9] Ahmed, B. (2015). Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. Landslides, 12(6), 1077-1095.

[10] Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. International Journal of Environmental Science and Technology, 11(4), 909-926.

[11] Niraj, K. C., Gupta, S. K., & Shukla, D. P. (2022). A Comparison of Image-Based and Physics-Based Atmospheric Correction Methods for Extracting Snow and Vegetation Cover in Nepal Himalayas Using Landsat 8 OLI Images. Journal of the Indian Society of Remote Sensing, 1-19.

[12] Poudyal, C. P., Chang, C., Oh, H. J., & Lee, S. (2010). Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. Environmental Earth Sciences, 61(5), 1049-1064.

[13] Crippen, R. E. (1990). Calculating the vegetation index faster. Remote sensing of Environment, 34(1), 71-73.

[14] Onagh, M., Kumra, V. K., & Rai, P. K. (2012). Landslide susceptibility mapping in a part of Uttarkashi district (India) by multiple linear regression method. International Journal of Geology, Earth and Environmental Sciences, 2(2), 102-120.

[15] Sarkar, S., & Kanungo, D. P. (2004). An integrated approach for landslide susceptibility mapping using remote sensing and GIS. Photogrammetric Engineering & Remote Sensing, 70(5), 617-625.

[16] Lee, S. (2019). Current and future status of GIS-based landslide susceptibility mapping: a literature review. Korean Journal of Remote Sensing, 35(1), 179-193.