Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

Mapping Earthquake Induced Landslides Using Machine Learning: A Case of Dolakha District in Post Gorkha Earthquake Context

Ajaya Pyakurel [a], Dipendra Gautam [b], Bhim Kumar Dahal [c]

[a, c] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[b] *Department of Civil Engineering, Thapathali Campus, IOE, Tribhuvan University, Nepal*
[a] pyaku.ajay@gmail.com, [b] dipendra01@tcioe.edu.np, [c] bhimd@pcampus.edu.np

Abstract

This study performs earthquake induced landslide susceptibility mapping and evaluates the underlying causative factors using Extremely Gradient Boosting algorithm (XGB) as a machine learning model. 14 landslide causative factors were taken to develop earthquake induced landslide susceptibility map. By using Gorkha earthquake induced landslide inventory map over the study area, 1275 landslide polygons used to develop training and testing data sets. Similarly, non landslide points are created randomly over the study area using QGIS. Training and testing data were in 70/30 ratio. XGB algorithm is trained using training data set and found that accuracy of training is 100%. While testing the accuracy of model to predict unknown points, testing accuracy found to be 89.97%. In the same way, area under the ROC curve (AUC) and kappa coefficient were calculated and values were obtained as 0.959 and .799 respectively.Finally, using XGB, susceptibility map developed and result shows 90% of areas were feebly susceptible, 5%, and 6% of the areas were moderately and highly susceptible, respectively. Furthermore 81% of building were exposed in the low susceptibility class, where as 5% of building were exposed to very high and high susceptibility class. This provides a handy information for urban planner, land use planning process, and others government authorities to make an effective mitigation and prevention action plan.

Keywords

earthquake induced landslide, machine learning, susceptibility, kappa coefficient

1. Introduction

Landslides are considered to be the most damaging geological hazard in mountainous regions of world [1, 2]. A landslide is the movement of a slope forming materials like rock and soil down a slope.Landslide is one of the widespread natural hazards in the hilly region of Nepal.Both natural and anthropogenic factors such as steep terrain, young and fragile geology, high rainfall intensity, deforestation, and unplanned human settlements are the major causes of landslides. Anthropogenic activities like improper land use further exacerbate landslide risk, encroachment into vulnerable land slopes, and unplanned development activities such as constructing roads and irrigation canals without proper protection measures in the vulnerable mountain belt [3]. In order to minimize the damage due to landslides, it is necessary to evaluate the factors responsible for the landslides. These factors include geology, geomorphology, land-use land cover, topography, rainfall, seismicity, and man made activities [1, 4, 5]. Landslide is complex interactions and have a complex relationship with causative factors, even though they may not be equally significant to landslide occurrences. Hence to produce an efficient landslide susceptibility map (LSM), it is fundamental and crucial to decide and select whether to include all the causative factors [6, 7]. Different study shows that Nepal is lie in the high seismic risk zone and ranked in 11th position in term of Global earthquake occurrence and risk [8]. According to deterministic seismic hazard assessment, peak ground acceleration can vary from 0.07g to 0.88g in different part of Nepal [9]. To minimize the earthquake induced landslide impacts on socioeconomic condition and livelihood, it is necessary to prepared the susceptibility map to locate the high hazard area and to take mitigation measure and disaster risk reduction.

Objective of study The main focus of this study is to prepare earthquake induced landslide susceptibility map and associated vulnerability assessment of buildings within the study area.

Study area There have been occurred several landslides in the Dohakha district 1.The 2,191 km2 study area is situated in central Nepal between latitudes 27.47 and 28.17 N and 85.88 and 86.55 E. With its steep terrain and young, weak geological formation, it is more prone to landslides. Following a 7.8 magnitude earthquake near Gorkha, an aftershock of 7.2 magnitude struck the Dolakha district.

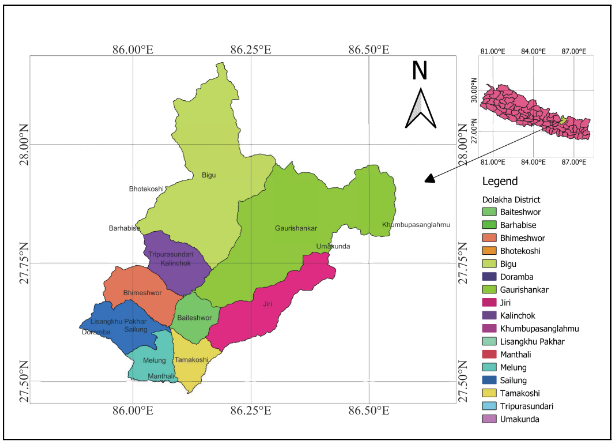

Figure 1: Study area map

2. Data and Methodology

2.1 Landslide inventory map

To make a prediction model for landslide susceptibility, two data sets were needed landslide inventory map and landslide causative factors (LCFs). During the preparation of landslide susceptibility prediction and mapping model, it is assumed that the landslide will be occurred under the same environment as it happened in the past [10].Using the past landslide inventory map, LCFs were analyzed using different statistical or machine learning (ML) models and a complex relationship were established with landslide occurrence or non-occurrence. Landslide inventory map is prepared from historical earthquake induced landslide developed by [11] (see Figure 2.

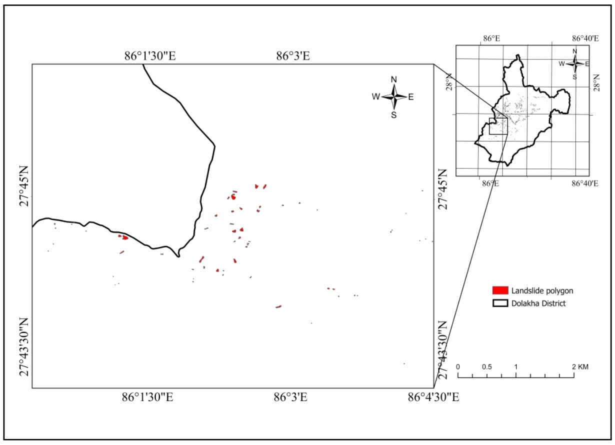

Figure 2: Earthquake induced landslide inventory map(source [11])

2.2 Landslide causative factors (LCFs)

Landslide inventory map and landslide causative factors are the mandatory data for susceptibility mapping. While preparing landslide susceptibility map it is considered that landslide will occurred in the similar environment as it occurred in past. 14 LCFs were developed using different data and the data used in this research are secondary data which were acquired from different data sources as mention in Table 1

Table 1: Detail of data sources

SN	Data	Data Source
1	Elevation	SRTM DEM (USGS)
2	Slope	Derived from DEM
3	Aspect	Derived from DEM
4	Profile curvature	Derived from DEM
5	Plan curvature	Derived from DEM
6	Distance to river	Derived from DEM
7	Distance to road	ICIMOD road data
8	Distance to fault	DMG
9	Distance to rupture	USGS shake map
10	Peak ground acceleration	USGS shake map
11	Geology	DMG
12	Lithology	GLim
13	Land use land cove	ESRI
14	Topographic ruggedness index	Derived from DEM

2.3 Methodology

LCFs were developed in the raster formats using QGIS software. Landslide polygon is used to create the centroid point inside the polygon and the attributes of these landslide points is updated with selected LCFs. Similarly, the non-landslide points were created randomly and these points attributes also updated. All the data preparation process is carried out in QGIS. These landslide and non-landslide points is divided into two data set as training (70%) and testing data set (30%). Extremely Gradient Boosting (XGB) is used as the machine learning algorithm for landslide susceptibility prediction and mapping process. Developed model is evaluated using training and testing accuracy (ACC), Kappa coefficient and area under the ROC curve (AUC). Furthermore, after making earthquake induced landslide susceptibility map, exposure of building structure is carried out. Building used in this study area is taken from open-street map.

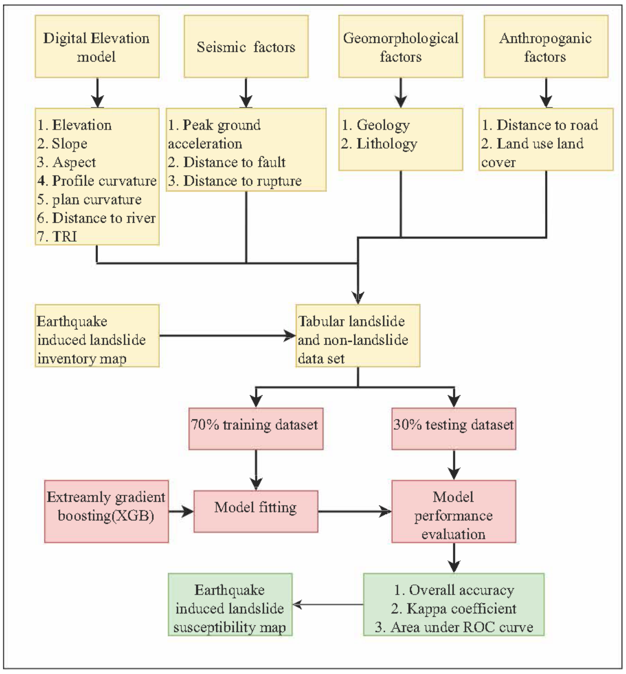

Figure 3: Research frame work

Extremely Gradient Boosting (XGB) is on of the most popular machine learning algorithm, which is used by data scientist, civil engineering and others many fields. It used effectively gradient tree boosting with minimizing following objective

$$\Gamma(\phi) = \sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \qquad (1)$$

Here, the difference between the forecast $\hat{y}_i$ and the target y_i is measured by the differentiable convex loss function ℓ. The model's complexity is penalized by the second term, Ω. Traditional optimization techniques in the Euclidean space cannot be used to improve the tree ensemble model in Equation 1 since it has functions as parameters. The model is instead trained in an additive way. Formally, if $\hat{y}_i^{(t)}$ is the prediction of the i^{th} instance at the t^{th} iteration, then f_k must be added to minimize the goal below. [12].

$$\Gamma(\phi) = \sum_i^n \ell(\hat{y}_i, y_i^{t-1}) + f_t(X_i) + \Omega(f_t) \qquad (2)$$

3. Result and Discussion

Landslide causative factors importance This study found that out of 14 LCFs, lithology play significant role in model output and followed by topographic ruggedness index (TRI), elevation, and so on as shown in Figure 4

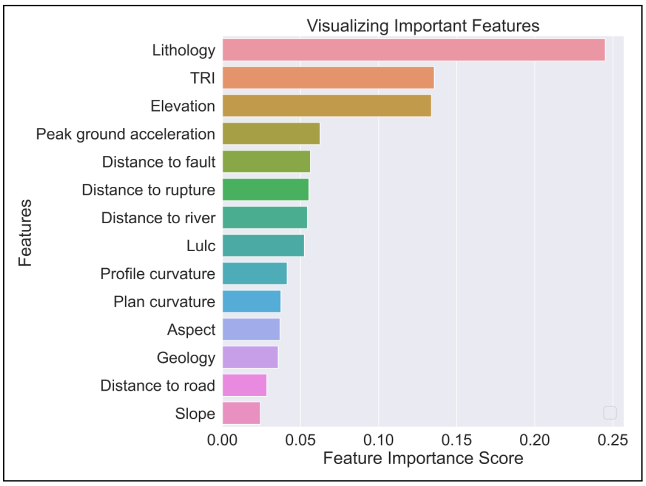

Figure 4: Features importance using XGB model

For the earthquake induced landslide susceptibility mapping, peak ground acceleration, distance to fault, distance to rupture also pay significant role. Based on the assessment of the relationship between the presence of landslides and the slope class, it is implied that the majority of landslides occur in a single slope class.

Model performance: As mentioned above, the model performance is evaluated using training and testing accuracy, kappa coefficient and AUC value see Table 2

Table 2: Model evaluation parameters

SN	Parameter	Value
1	Training accuracy	100%
2	Testing accuracy	89.97
3	Kappa coefficient	0.799
4	Area under ROC curve	0.959

The training accuracy indicated that during training the model, it identified each target label accurately with accuracy 100%. Similarly while predicting the testing data, XGB model 89.97% of target label are accurately classify. To find the strength of agreement of classification, kappa coefficient is used and it is found that there is a strong strength of agreement of classification with 0.799 coefficient value. Area under the ROC curve also widely used to measure the model performance. XGB model with 0.959 AUC value indicated that for earthquake induced landslide susceptibility mapping, XGB is a good classifier for the study area.

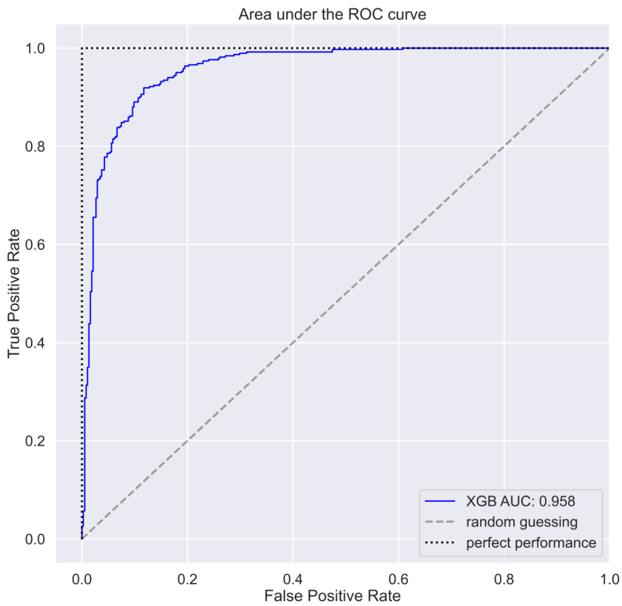

Figure 5: Area under the ROC curve using testing data set

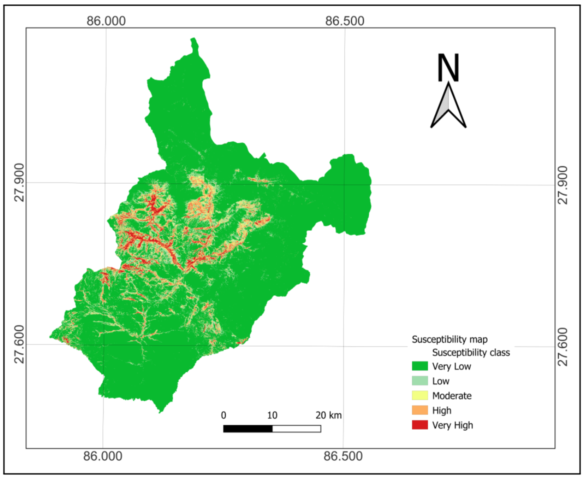

Figure 6: Earthquake induced landslide susceptibility map using XGB algorithm

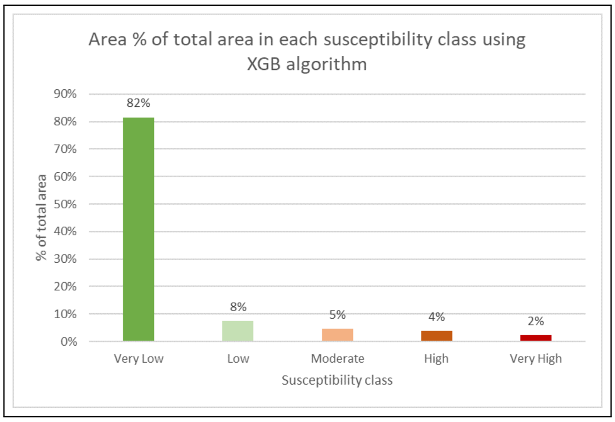

Figure 7: Percentage area of total study area in each susceptibility class

Earthquake induced landslide susceptibility map: After evaluating the model, earthquake induced landslide susceptibility map is prepared (Figure 6). Similarly, area percentage of each susceptibility class over the whole study area is calculated and it is found that the 6% of total area lie in very high and high susceptibility class as shown in Figure 7.

Furthermore, it is found that the 3,775 building out of 79,075 lie in the very high and high susceptibility (Figure 8). This indicated the the 5% of building were exposed to the earthquake induced landslide. Similarly, 81% of building lie in the very low susceptibility class Figure 9. It is observed that area of high susceptibility class in Jiri municipality is less as compared to Bigu rural municipality Figure 8.

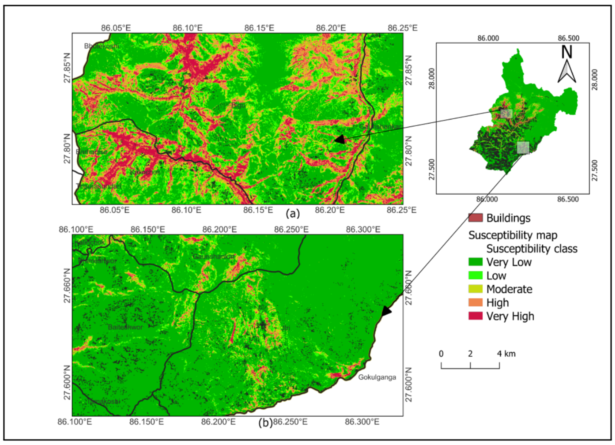

Figure 8: Building footprints overlay with susceptibility map: (a) Bigu rural municipality, (b) Jiri municipality

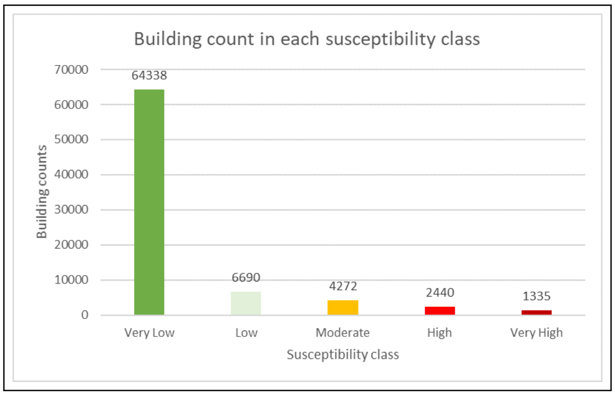

Figure 9: Number of building in each susceptibility class

Accuracy and AUC obtained in this study 89.97% and 0.959 were compared with the result obtained by Can et al., 2021 [13] study and it was observed the accuracy value less then the XGB(90.18%) and AUC (0.96). Similarly, another study carried out for the earthquake induced landslide susceptibility mapping by Gautam et al., [14] in the upper Indrawati Watershed using logistic regression and it is found that the success rate (0.843) and prediction rate (0.832).

4. Conclusion

Nepal is located in a very high seismic zone, and earthquakes frequently occur of different magnitudes. Identifying the earthquake-induced landslide susceptibility is necessary to minimize the impact of earthquake-induced landslides. Furthermore, to make a susceptibility map, it is crucial to find the model with the best performance. This study found that for the earthquake induced landslide susceptibility mapping, XGB model is suitable for the Dolakha district. AUC value is obtained using testing data set is found 0.959 which shows that the XGB is excellent model classifier for the earthquake induced landslide classification problem. Further more Kappa coefficient (0.799) indicated that there is a strong agreement in classification and output of model. After making susceptibility map, assessment of building exposure is carried out and it is found that 5% of total building in the Dolakha district located in very high and high susceptibility class and 81.36% of buildings located very low susceptibility class. Finally, the result of this studies provides the handy information to the urban planner as well as the governmental authority to formulate the mitigation, preparedness plan and land use planning. Furthermore this information can be utilized to minimize the risk associated with earthquake induced landslide and its secondary hazard like landslide dam outburst flood.

Acknowledgments

The authors would like to thank the Department of Mines and Geology for providing geological data.

References

[1] Tarun Kumar Raghuvanshi, Jemal Ibrahim, and Dereje Ayalew. Slope stability susceptibility evaluation parameter (ssep) rating scheme – an approach for landslide hazard zonation. *Journal of African Earth Sciences*, 99:595–612, 11 2014.

[2] Filagot Mengistu, K.V. Suryabhagavan, Tarun Kumar Raghuvanshi, and Elias Lewi. Landslide hazard zonation and slope instability assessment using optical and insar data: A case study from gidole town and its surrounding areas, southern ethiopia. *Remote Sensing of Land*, 3:1–14, 1 2019.

[3] ADCP. Nepal innovations for climate adaptation and resilience current status and needs assessment 2021, 2021.

[4] Bhim Kumar Dahal and Ranjan Kumar Dahal. Landslide hazard map: tool for optimization of low-cost mitigation. *Geoenvironmental Disasters*, 4:1–9, 12 2017.

[5] Leulalem Shano, Tarun Kumar Raghuvanshi, and Matebie Meten. Landslide susceptibility evaluation and hazard zonation techniques – a review. *Geoenvironmental Disasters 2020 7:1*, 7:1–19, 5 2020.

[6] A. N. Matori and A. Basith. Evaluation of landslide causative factors towards efficient landslide susceptibility modelling in the cameron highlands, malaysia. *WIT Transactions on Engineering Sciences*, 73:207–217, 5 2012.

[7] Jie Dou, Dieu Tien Bui, Ali P. Yunus, Kun Jia, Xuan Song, Inge Revhaug, Huan Xia, and Zhongfan Zhu. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and gis data in parts of niigata, japan. *PloS one*, 10, 7 2015.

[8] MOHA. Nepal disaster report 2015, 2015.

[9] Dilli Ram Thapa, DR Thapa, X Tao, G Wang, and F Fan. Deterministic seismic hazard assessment for nepal seismic hazard analysis and recognition of earthquake-prone areas in nepal and adjacent regions view project deterministic seismic hazard assessment for nepal, 2017.

[10] Ahmed Mohamed Youssef, Hamid Reza Pourghasemi, Ahmed Mohamed Youssef, and Hamid Reza Pourghasemi. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at abha basin, asir region, saudi arabia. *Geoscience Frontiers, 2021, Vol. 12, Issue 2, Pages: 639-655*, 12:639–655, 3 2021.

[11] Kevin Roback, Marin K. Clark, A. Joshua West, Dimitrios Zekkos, Gen Li, Sean F. Gallen, Deepak

Chamlagain, and Jonathan W. Godt. Map data of landslides triggered by the 25 april 2015 mw 7.8 gorkha, nepal earthquake. *U.S. Geological Survey*, 4 2017.

[12] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system.

[13] Recep Can, Sultan Kocaman, and Candan Gokceoglu. A comprehensive assessment of xgboost algorithm for landslide susceptibility mapping in the upper basin of ataturk dam, turkey. *Applied Sciences (Switzerland)*, 11, 6 2021.

[14] Pawan Gautam, Tetsuya Kubota, and Aril Aditian. Evaluating underlying causative factors for earthquake-induced landslides and landslide susceptibility mapping in upper indrawati watershed, nepal. *Geoenvironmental Disasters*, 8:1–19, 12 2021.