Proceedings of 15th IOE Graduate Conference
Peer Reviewed
Year: 2024 Month: May Volume: 15
ISSN: 2350-8914 (Online), 2350-8906 (Print)

Analysis of Climate Induced Landslide Susceptibility Map Using Machine Learning Algorithm: A Case Study of Nuwakot District, Nepal

Prem Raj Bhattarai [a], Bhim Kumar Dahal [b]

[a] *Department of Applied Sciences and Chemical Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[b] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[a] 078msccd009.prem@pcampus.edu.np, [b] bhimd@pcampus.edu.np

Abstract

This paper investigates the climate change influence on landslide susceptibility mapping (LSM) in Nuwakot district, under the future climate change scenarios. The study employs 11 factors as a landslide contributing factor, such as terrain slope, aspect, curvature, elevation, geology, TWI, SPI, distance to stream, distance to road, land use/ land cover, and rainfall. Among them, rainfall is considered a dynamic climate factor. In this study, the machine learning algorithm called XGBoost was selected in order to map landslide susceptibility within the context of baseline (1995-2019) and for the future: near future (2021-2045), mid future (2046-2070) and far future (2071-2095) under SSP245 and SSP585 scenarios, which are based on the three Coupled Model Inter-comparison Project Phase 6 (CMIP6) global climate model ensembles. Data preparation and normalization are performed using QGIS. Based on the results, future annual rainfall is expected to rise under both scenarios, with SSP585 exhibiting more significant climatic changes. The AUC value of 97.25% indicates that XGBoost is an effective classifier for LSM in the study area, and evaluation metrics such as accuracy, recall, precision, Mathew's correlation coefficient, and Kappa Coefficient are used to measure the quality of the model. Altogether, seven LSMs were generated, including baseline and future scenarios, in which future scenarios have an increase in high and very high-class values compared to baseline susceptibility map. Thus, the result indicates that the far future LSM under the SSP585 zonation is impacted more significantly due to the effect of climate change.

Keywords

Landslide Susceptibility, Climate Change, CMIP6, XGBoost, Extreme Rainfall

1. Introduction

Climate change and global warming seem to be the most critical environmental challenges of the 21st century[1]. Climate change, which is predicted to impact the frequency and severity of rainfall events, creates the conditions for more landslides in many areas worldwide[2][3][4]. Based on the Global Climate Risk Index, Nepal is placed fourth among the countries facing climate risks[5]. The aberrant temperatures and rainfall drainage occurring due to climate change can intensify the probability of landslide rates[6].

Landslides refer to natural processes that cause earth materials to move down-slope at varying speeds. Nepal Disaster Risk Reduction Portal reported that from 2010 to 2020, Nepal experienced 2386 landslide incidents, making it the third most prevalent type of natural disaster in the country[7]. Annually, the mountainous regions of Nepal face significant challenges due to their rugged terrain, steep slopes, unpredictable geological formations, and the presence of fragile rock formations. These factors, coupled with intense and prolonged rainfall during the monsoon season, contribute to the occurrence of severe landslides and associated phenomena[8].

Landslide susceptibility is the degree to which landslides are likely to be triggered by a localized slope[9]. Consequently, the production of susceptibility maps would play a great part in delineating areas susceptible to future landslide eventualities. Extensive research has been carried out on susceptibility mapping using a range of tools[10], including direct mapping[11], heuristic techniques[12], deterministic models[13], probabilistic methods[14], and machine learning models[15]. Among these methods is the application of machine learning algorithms, which have gained popularity in recent years because of advancements in the field of algorithms and remote sensing data and survey sources. In this research, the XGBoost algorithm is used for landslide susceptibility mapping. Pyakurel et al. assessed machine learning algorithms for predicting landslide susceptibility caused by earthquakes, and they reported that XGBoost proved to have the best performance in terms of achieving high accuracy.

The Shared Socioeconomic Pathways (SSPs), which comprise the SSP245 and SSP585 scenarios, were designed to examine the implications of different levels of global warming and socioeconomic situations on future climatic scenarios. The SSP245 scenario assumes modest socioeconomic growth and provides a radiative forcing of 4.5 W/m^2 at the end of the century. This is consistent with major, but not excessive, mitigation actions [16]. SSP585, with a radiative forcing of 8.5 W/m^2, is considered a high-end route due to its high greenhouse gas emissions and inadequate mitigation efforts [17]. These scenarios are critical for climate modeling.

The aim of this study is to use CMIP6 global climate model data combined with the landslide susceptibility models in order to explore the near future (2021-2045), mid future (2046-2070), and far future (2071-2095) landslide conditions in the Nuwakot District—specifically in respect of the change in extreme rainfall under SSP245 and SSP585 scenarios.

2. Study Area

Nuwakot district is located in the Bagmati province of Nepal as shown in figure 1. It covers an area of 1,121 km^2 within latitudes 27°48' N to 28°06' N and longitudes 84°58'E to 85°30"E. Its altitude ranges from 518m to 4876 m above sea level[18]. Sindhupalchowk and Kathmandu districts are to the east; Dhading to the west; Rasuwa to the north; and Dhading and Kathmandu to the south. Geologically, the area lies in the Lesser Himalayan and Higher Himalayan Zones of central Nepal. The area has a very distinct geological feature. The study area constitutes hilly terrain exhibiting rugged topography with a diversity of land forms, which is characterized by the elevated mountains and the deep river valleys.

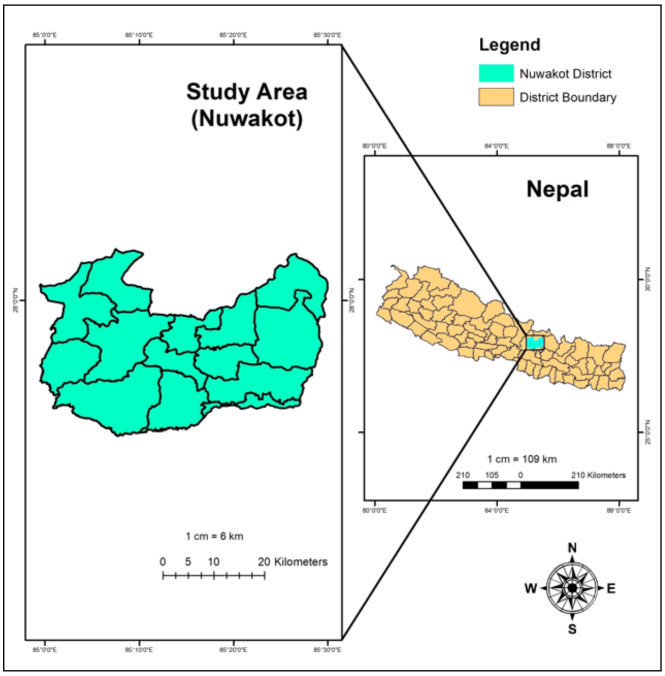


Figure 1: Location of study area

3. Materials and Methodology

3.1 Landslide Inventory

A landslide inventory map provides crucial data for assessing landslide hazards or risks on a regional level for probabilistic analysis of landslide susceptibility [19]. Satellite images were integrated to prepare landslide inventory. Polygons were constructed in Google Earth to map previous landslides and then processed in a QGIS. Figure 2 illustrates that in the study area, 491 landslides were mapped which covers 2.84 km^2 area.

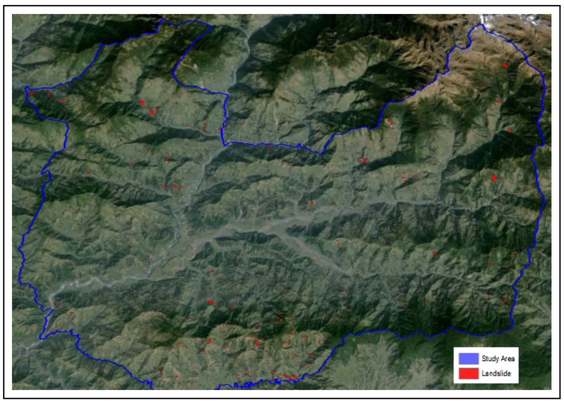


Figure 2: Landslide Inventory Map

3.2 Data Collection and preparation of Landslide causative factors (LCFs)

Data collection for this study was accomplished through the use of satellite imagery, aerial photography, scholarly articles, official government documents, and online resources. Choosing and preparing the LCFs database is essential for attaining

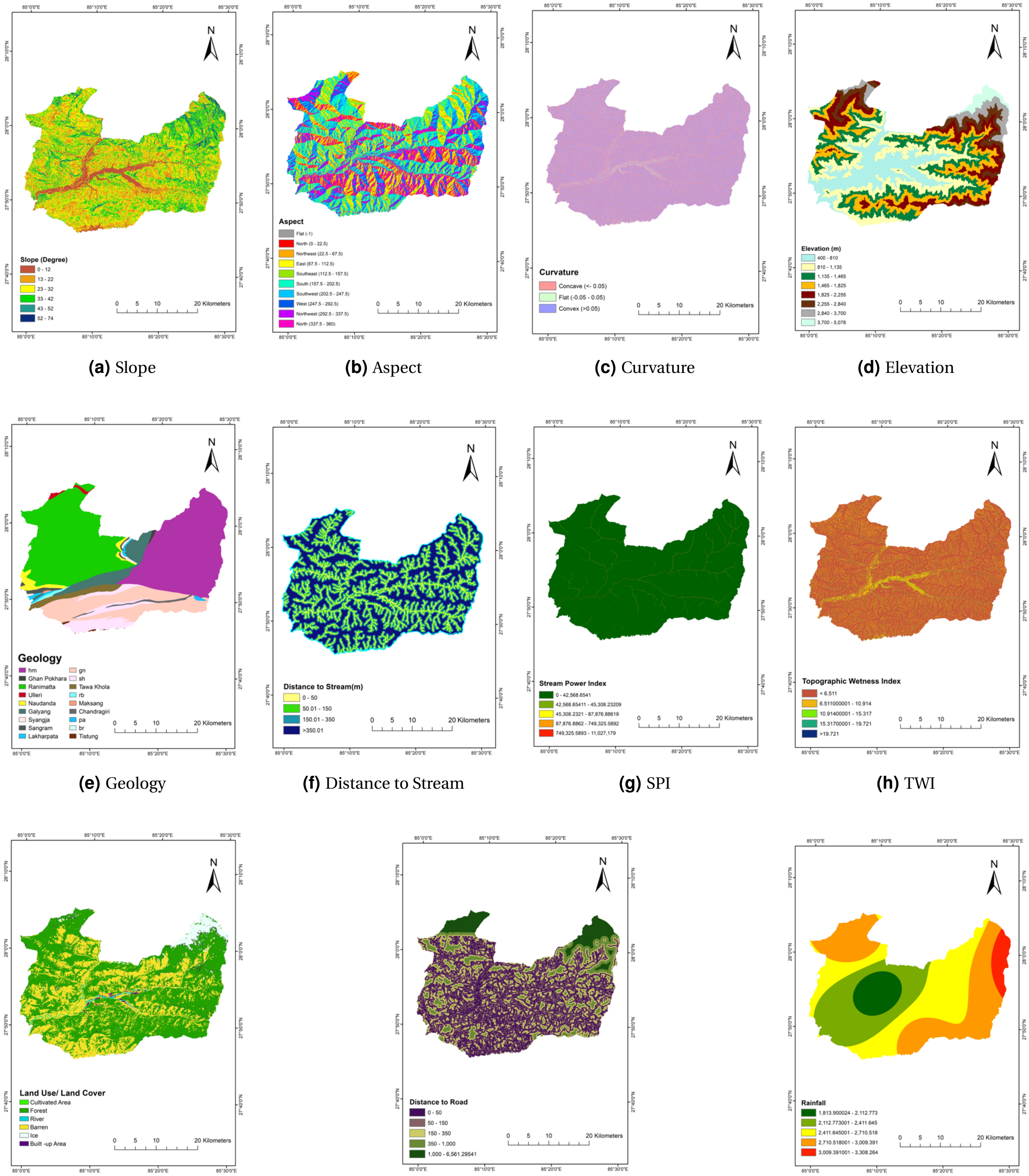
precise accuracy in the landslide susceptibility model for predicting areas at risk of landslides. In this work, 11 different landslide conditioning factors are used and these factors are process through using QGIS as shown in figure 3. Table 1 illustrates the different landslide causative factors taken, their resolutions and their sources.

Table 1: Landslide causative factors and data source

S.N.	Landslide causative factors	Source
1	Slope	DEM
2	Aspect	DEM
3	Elevation	DEM (USGS)
4	Curvature	USGS (DEM)
5	Geology	DMG
6	Distance to Stream	DEM
7	Stream Power Index	USGS (DEM)
8	Topographic Wetness Index	USGS (DEM)
9	Distance to Road	DOR
10	LULC	ICIMOD
11	Rainfall	DHM

3.3 Methodology

Landslide conditioning factors (LCFs) were created on the basis of various data sources. Landslide and non-landslide dataset were produced from the landslide inventory map which gathers data of the affected areas using historical data, field surveys and remote sensing. The projections for future precipitation were formed from three GCM such as EC-Earth3, MPI- ESM1- 2HR and NorESM2- MM and under SSP245 and SSP585 for different period NF (2021-2045), MF (2046-2070) and FF (2071-2095). All these data for different period are assembled using QGIS. From which landslide points and non- landslide points dataset are extracted in the form of Excel file then processed in python. These points are divided into training (70%) and testing (30%) data sets. Machine learning algorithms called Extremely Gradient Boosting (XGBoost) is used for landslide susceptibility prediction and mapping process. The performance of the model was assessed utilizing metrics such as Accuracy, Precision, Recall, Matthew's Correlation Coefficient (MCC), Kappa Coefficient, and Area under the ROC curve. Then model resulted the Landslide Susceptibility Mapped for Baseline, Near Future (NF), Mid Future (FF) and Far Future (FF). Figure 4 summarized the methodology of this study.



(a) Slope

(b) Aspect

(c) Curvature

(d) Elevation

(e) Geology

(f) Distance to Stream

(g) SPI

(h) TWI

(i) LULC

(j) Distance to Road

(k) Rainfall

Figure 3: Map of Landslide Causative Factor

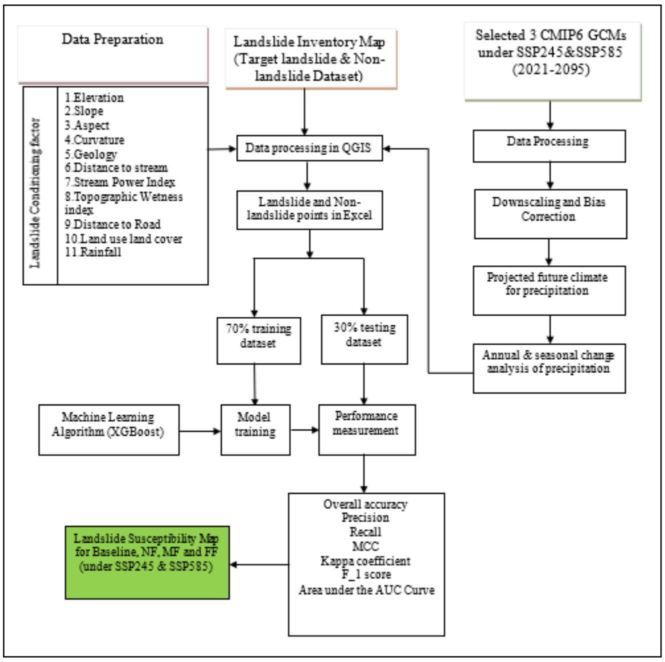


Figure 4: Methodological Framework

3.3.1 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a machine learning algorithm that belongs to the class of ensemble learning methods, where multiple models are combined to improve predictive performance. It builds a strong predictive model by sequentially adding weak learners, typically decision trees, and optimizing a predefined objective function. It used effectively gradient tree boosting with minimizing following objective

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

Here, the difference between the forecast y^i and the target yi is measured by the differentiable convex loss function ℓ. The model's complexity is penalized by the second term, Omega. 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 y^i (t) is the prediction of the i th instance at the t th iteration, then fk must be added to minimize the goal below[20].

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

4. Result

4.1 Performance Measurement of Model

The calculated performance measuring parameter as accuracy, precision, recall, MCC, Kappa and AUC of the XGBoost model is summarized in Table 2. By using XGBoost Model, accuracy, precision and recall values were found to be 93.2%, 84% and 88.19% respectively. Kappa Coefficient is used to find the strength of agreement of classification which found 81.55% coefficient value. The AUC value of 97.25% indicates that XGB is an effective classifier for landslide susceptibility mapping in the study area as shown in Figure 5.

Table 2: Model Performance Parameter

S.N.	Performance Parameter	Value
1	Accuracy	93.2%
2	Precision	84%
3	Recall	88.19%
4	MCC	81.59%
5	Kappa Coefficient	81.55%
6	AUC	97.25%

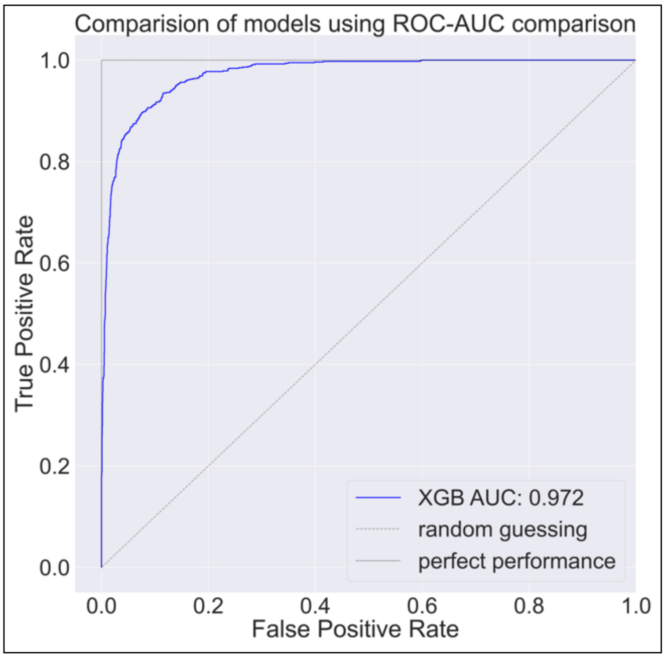


Figure 5: Area under the ROC Curve

4.2 Landslide susceptibility map

A baseline susceptibility map for study area is created following XGBoost model evaluation as shown in Figure 6. Likewise, the proportion of area occupied by each susceptibility class across the entire study area was determined. It was observed that (34.63 km^2) 2.94% of the total area falls within the categories of very high and high susceptibility. The majority of the study area is categorized as having very low susceptibility, encompassing (1039.39 km^2) 88.12% of the total area which is followed by low and moderate classes, as illustrated in Table 3.

Table 3: susceptibility classes for Landslide affected area

S.N.	Susceptibility Class	Area(km^2)	Area(%)
1	Very Low	1039.39	88.12
2	Low	76.21	6.46
3	Moderate	29.26	2.48
4	High	19.25	1.63
5	Very High	15.38	1.30

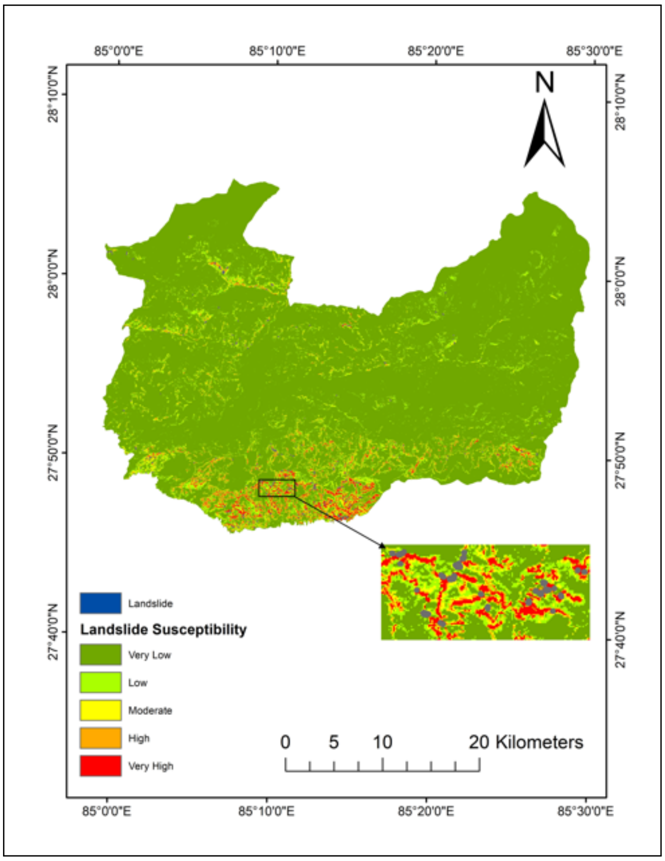


Figure 6: Baseline landslide Susceptibility Map

To determine which class ranges are more susceptible to landslides, the baseline susceptibility map was compared to thematic maps of the landslide conditioning factors. According to the analysis, areas with slopes ranging from 32° to 42° exhibit high susceptibility to landslides. South-facing aspects are particularly prone to landslides based on slope aspect, while concave areas are more susceptible in terms of curvature. Also, TWI, SPI, Elevation, Distance to Road and Distance to Stream susceptibility is notably high within the class range of 0-6.51, 0 - 42,568.65, 1135 – 1465 meters, 0-50 meters and >350 meters, respectively. In geological terms, the Ranimatta are the most susceptible. The majority of forest, and barren land cover classes fall within highly susceptible areas. For rainfall, susceptibility peaks within the ranges of 2,411.65 - 2,710.52mm.

4.3 Projected Future Climate for Precipitation

EC- Earth3, MPI- ESM1- 2HR and NorESM2- MM are the three GCM selected for the future climate projection for precipitation. From these GCM, values were projected and ensembled into a single value for the selected meteorological station and the annual average value was taken for near future (NF), mid future (MF) and far future (FF). Under SSP245 and SSP585 are the scenarios taken for each future period. The average annual precipitation across the entire study area is 15.80%, 17.84%, and 16.31% for NF, MF, and FF, respectively, under SSP245, with values increasing order. And, the seasonal (monsoon) precipitation for NF, MF, and FF is 13.01%, 16.10%, and 14.28%, respectively, is in increasing order. Similarly, seasonal changes under SSP245 values for winter, pre-monsoon and post-monsoon seasonal shown in Figure 7.

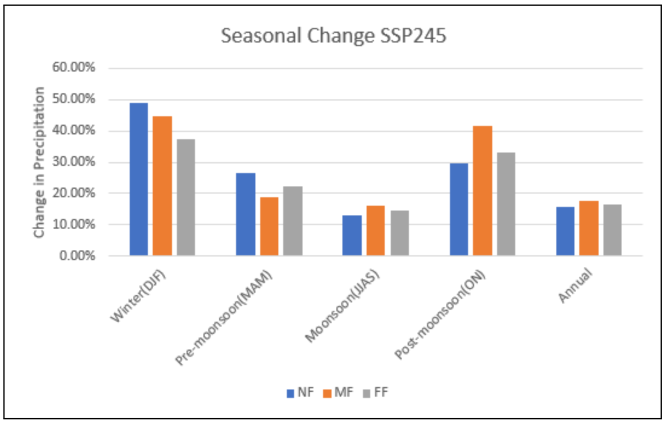


Figure 7: Projection of Seasonal Precipitation under SSP245

Likewise, Under SSP585, the average annual precipitation for the whole study area is 14.24%, 27.51%, and 44.70% for NF, MF, and FF, respectively, with values increasing order. Figure 8 illustrates the seasonal change under SSP585 in which the seasonal (monsoon) precipitation for NF, MF, and FF is 9.71%, 25.85%, and 44.69%, respectively, showing a raising trend. Pre-monsoon changes are projected to be 31.40%, 32.89%, and 34.03%, while winter precipitation changes are projected to be 48.33%, 41.45%, and 37.99% for NF, MF, and FF, respectively. Post-monsoon changes are similarly projected to be 55.45%, 44.51%, and 87.34%.

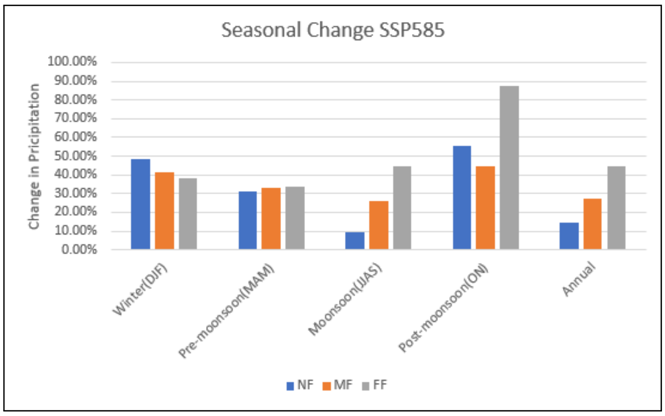
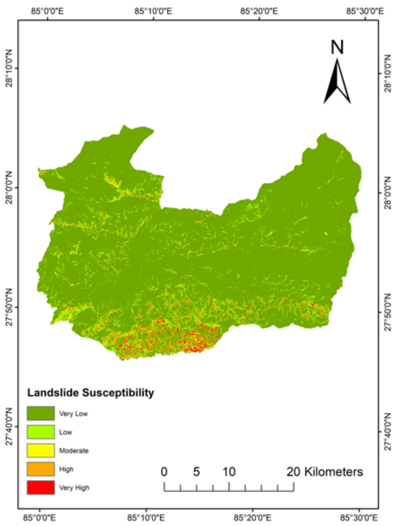


Figure 8: Projection of Seasonal Precipitation under SSP585

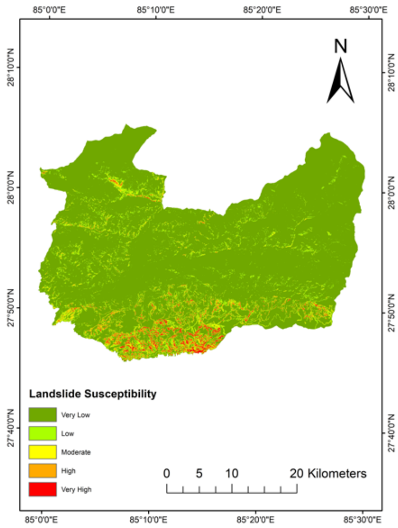
4.4 Landslide Susceptibility Maps Under Future Climate Scenarios

The landslide susceptibility maps for all future climate scenarios (NF, MF, and FF) generated from SSP245 and SSP585 were classified into five susceptibility classes: very low, low, moderate, high, and very high, as illustrated in Figure 9.

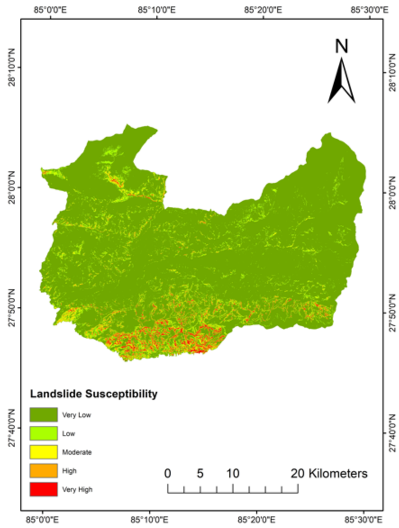
Table 4 presented the proportion of area occupied by each susceptibility class across the entire study area under SSP245 for NF (2021-2045), MF (2046-2070) and FF (2071-2095). The zones classified as high and very susceptible, comprising approximately 2.96%, 2.92%, and 3.11% of the study area, respectively, exhibited an increasing trend under the SSP245 scenario for NF, MF, and FF, in comparison to the baseline period (1995-2019).



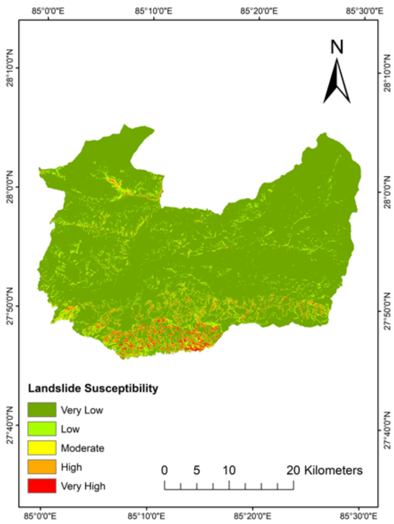
(a) Near Future



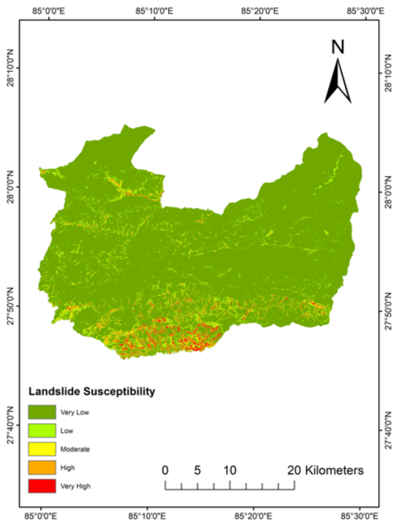
(b) Mid Future



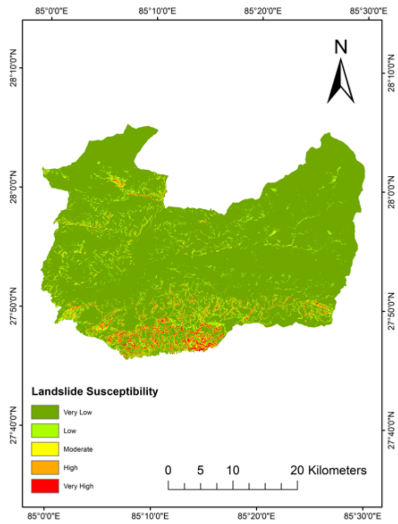
(c) Far Future



(d) Near Future



(e) Mid Future



(f) Far Future

Figure 9: Landslide susceptibility map: (a), (b), (c) under SSP 245 scenario; (d),(e),(f) under SSP 585 scenario

Table 4: Landslide susceptible area under SSP245

Susceptible Area	SSP 245 (km^2)		
	NF	MF	FF
Very Low	1035.50	1047.16	1041.92
Low	77.75	69.85	71.44
Moderate	30.83	27.47	28.89
High	20.00	19.29	21.23
Very High	15.42	15.72	16.01

Table 5: Landslide susceptible area under SSP585

Susceptible Area	SSP 585 (km^2)		
	NF	MF	FF
Very Low	1037.67	1034.03	1028.31
Low	76.94	77.33	80.09
Moderate	30.07	31.00	31.28
High	19.31	21.26	22.76
Very High	15.51	15.87	17.05

Table 5 shown the proportion of area occupied by each susceptibility class in the whole study area under SSP585 for NF (2021-2045), MF (2046-2070) and FF (2071-2095). The high and very susceptibility level zones, which represented around 2.91%, 3.1%, and 3.33% of the study region, respectively, showed a rising trend based on the SSP585 scenario for NF, MF, and FF compared with baseline (1995-2019).

The very high landslide susceptibility area under SSP585 for Near future, Mid Future and Far future is 15.51 km^2, 15.87 km^2 and 17.05 km^2 respectively, which is increasing trend. Also, high and moderate susceptible area for Near Future, Mid Future and Far future area in increasing order. It is observed that within susceptibility maps for all scenarios, the susceptibility map for the scenarios SSP585 FF demonstrates the highest susceptibility. This data shows that there can be a rise in vulnerability to landslides location if there will be climate change. Based on another study[21] conducted across the Nepalese valley regions of Bagmati and Madhesh, for future RCP4.5 and RCP8.5 climatic change scenarios, there would be more counties and areas with very high vulnerability than the baseline period.

5. Conclusion

In this study, landslide susceptibility modeling and mapping were conducted using the machine learning algorithm called XGBoost to identify areas prone to landslides under both present and projected future conditions. The AUC value derived from the testing dataset is calculated as 0.9725, indicating the XGBoost model's exceptional performance as a classifier for climatic-induced landslide classification. Additionally, the Kappa coefficient of 0.8155 suggests a substantial agreement between the model's classifications and its output. The result indicates a rise in the extent of high and very high landslide susceptible areas under both SSP scenarios compared to the baseline period, with the SSP585 far-future scenario exhibiting the most substantial increase in susceptibility area. As increasing trend rainfall is only a dynamic factor used in this study area, result increasing in landslide susceptibility map indicates an effect of climate change on landslide.

In the future, prediction accuracy will be influenced by various factors, including data quality such as the reliability of landslide inventory maps and the inclusion of dynamic factors like land use and land cover (LULC). Additionally, the choice of machine learning algorithm plays a crucial role. To address these limitations and improve predictions, future research efforts should focus on enhancing data quality, exploring alternative algorithms, and considering dynamic factors like LULC. Furthermore, expanding the regional and temporal focus of studies will be essential for a more comprehensive understanding of landslide dynamics. By broadening the scope to include diverse environmental factors and geographical contexts, future research can refine risk mitigation strategies and contribute to more effective land use planning, infrastructure initiatives, and emergency preparedness. This approach will enhance community resilience to the evolving impacts of climate change.

Acknowledgments

The authors are grateful to Department of Hydrology and Meteorology, and Department of Mines and Geology for providing data.

References

[1] Corrado AS Camera, Greta Bajni, Irene Corno, Mattia Raffa, Stefania Stevenazzi, and Tiziana Apuani. Introducing intense rainfall and snowmelt variables to implement a process-related non-stationary shallow landslide susceptibility analysis. *Science of the total environment*, 786:147360, 2021.

[2] SL Gariano, G Rianna, O Petrucci, and F Guzzetti. Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale. *Science of the total environment*, 596:417–426, 2017.

[3] Yang Li, Jingyi Bai, Zhiwei You, Jun Hou, and Wei Li. Future changes in the intensity and frequency of precipitation extremes over china in a warmer world: Insight from a large ensemble. *PLoS One*, 16(5):e0252133, 2021.

[4] Valérie Masson-Delmotte, Panmao Zhai, Anna Pirani, Sarah L Connors, Clotilde Péan, Sophie Berger, Nada Caud, Y Chen, L Goldfarb, MI Gomis, et al. Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, 2(1):2391, 2021.

[5] David Eckstein, Marie-Lena Hutfils, and Maik Winges. Global climate risk index 2019. *Who suffers most from extreme weather events*, 36, 2018.

[6] Stefano Luigi Gariano and Fausto Guzzetti. Landslides in a changing climate. *Earth-Science Reviews*, 162:227–252, 2016.

[7] Nepal Disaster Risk Reduction Portal. Nepal disaster risk reduction portal, 2021.

[8] Tri Dev Acharya. Regional scale landslide hazard assessment using machine learning methods in nepal. *esis for: Doctor in EngineeringAdvisor,, esis for: Doctor in EngineeringAdvisor*, 2018.

[9] Earl E Brabb. Minimum landslide damage in the united states, 1973-1983. Technical report, US Geological Survey, 1984.

[10] Paola Reichenbach, Mauro Rossi, Bruce D Malamud, Monika Mihir, and Fausto Guzzetti. A review of statistically-based landslide susceptibility models. *Earth-science reviews*, 180:60–91, 2018.

[11] Antonini Carrara, Fausto Guzzetti, Mauro Cardinali, and Paola Reichenbach. Use of gis technology in the prediction and monitoring of landslide hazard. *Natural hazards*, 20:117–135, 1999.

[12] Faming Huang, Zhongshan Cao, Jianfei Guo, Shui-Hua Jiang, Shu Li, and Zizheng Guo. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, 191:104580, 2020.

[13] Federico Cervi, Matteo Berti, Lisa Borgatti, Francesco Ronchetti, Federica Manenti, and Alessandro Corsini. Comparing predictive capability of statistical and deterministic methods for landslide susceptibility mapping: a case study in the northern apennines (reggio emilia province, italy). *Landslides*, 7:433–444, 2010.

[14] Işık Yilmaz. Comparison of landslide susceptibility mapping methodologies for koyulhisar, turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environmental Earth Sciences*, 61:821–836, 2010.

[15] Ajaya Pyakurel, Bhim Kumar Dahal, and Dipendra Gautam. Does machine learning adequately predict earthquake induced landslides? *Soil Dynamics and Earthquake Engineering*, 171:107994, 2023.

[16] Brian C O'Neill, Claudia Tebaldi, Detlef P Van Vuuren, Veronika Eyring, Pierre Friedlingstein, George Hurtt, Reto Knutti, Elmar Kriegler, Jean-Francois Lamarque, Jason Lowe, et al. The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, 9(9):3461–3482, 2016.

[17] Keywan Riahi, Detlef P Van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C O'neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global environmental change*, 42:153–168, 2017.

[18] Harish Dangi, Tara Nidhi Bhattarai, and Prem Bahadur Thapa. An approach of preparing earthquake induced landslide hazard map: A case study of nuwakot district, central nepal. *Journal of Nepal Geological Society*, 58:153–162, 2019.

[19] Netra R Regmi, John R Giardino, and John D Vitek. Characteristics of landslides in western colorado, usa. *Landslides*, 11:589–603, 2014.

[20] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[21] I Putu Krishna Wijaya, Peeranan Towashiraporn, Anish Joshi, Susantha Jayasinghe, Anggraini Dewi, and Md Nurul Alam. Climate change-induced regional landslide hazard and exposure assessment for aiding climate resilient road infrastructure planning: a case study in bagmati and madhesh provinces, nepal. In *Progress in Landslide Research and Technology, Volume 1 Issue 1, 2022*, pages 175–184. Springer, 2023.