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

A Study of Random Forest and Support Vector Machine Algorithms for Landslide Susceptibility Mapping at Bhotekoshi Rural Municipality

Dipesh Pokharel [a], Ram Chandra Tiwari [b], Kshitij Charana Shrestha [c], Umesh Jung Thapa [d]

[a, b, c] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[d] *Department of Civil Engineering, Kathmandu University, Nepal*
[a] deeface0@gmail.com

Abstract

The study aims to evaluate and compare the performances of Support Vector Machine (SVM) and Random Forest (RF) machine learning methods for landslide susceptibility mapping of Bhotekoshi rural municipality. Spatial data prepared using remote sensing and GIS tools were applied to machine learning algorithms using python code in Jupyter Notebook. Eleven landslide variables were generated which include slope, aspect, elevation, distance from road, distance to river, plan curvature, profile curvature, Total Wetness Index (TWI), Total Ruggedness Index (TRI), landcover and NDVI. The performance of the MLTs was evaluated, validated, and compared using the area under the curve (AUC-ROC) method. AUC values for RF=90%and SVM=89% was obtained. According to AUC value, random forest method was found to have the best performance for our study area. Furthermore, landslide susceptibility map of Bhotekoshi rural municipality was prepared which can be very helpful for planning and mitigation of landslide hazards.

Keywords

landslide susceptibility, machine learning algorithm, geographical information system, Random Forest, support vector machine

1. Introduction

Landslide, a recurrent hazard in hilly and mountainous region of Nepal, is defined as outward and downhill movement of mass of rock, earth,or debris due to gravity [1]. It has resulted in large number of casualties and economic loss in mountainous region all over the world [2, 3, 4]. Global Climate Risk Index ranks Nepal fourth in terms of climate risks [5] which highlights the importance of identifying landslide prone areas for effective disaster planning preparation and mitigation.

Landslide susceptibility is the chances of occurrence of landslides in an area based on different slope failure conditioning factors [6]. The spatial distribution and characteristics of future landslides can be predicted using landslide inventory and data of the factors that affect landslide formation. Susceptibility mapping is important as field survey and dynamic monitoring is challenging in undeveloped areas. Various methods have been studied, from qualitative to quantitative methods , like direct mapping [7], heuristic approach [8], deterministic model [9], probabilistic method [10] and machine learning models [11]. The direct mapping and heuristic approach are highly subjective and expert dependent. The deterministic method involve physical model development which is very complex and suitable for small area only. Machine learning methods are very popular in recent years due to development of algorithms and availability of data from remote sensing and survey sources.

Random forest and support vector machine are quite popular for binary classification model [12, 13] and have been used for solving various sorts of problems, but only a few number of these researches were for landslide susceptibility for Nepal. Further, comparing machine learning techniques for susceptibility mapping is crucial as properties of each method is unique and choosing the best method has significant effect on real applications. The objective of the study was to compare and evaluate RF and SVM methods for landslide susceptibility mapping. Bhotekoshi rural municipality lies in one of the highly landslide affected region of Nepal and is chosen as the study area. The remote sensing data and GIS tools were used for data generation and interpretation. The findings of the study can be used to assess and lessen the risks related to landslide hazards.

2. Study area

Bhotekoshi rural municipality 1, located in the northeast of Sindhupalchowk District of Province 3, has area of 273.62 km2 and altitude ranges from 1100m to 5000m above sea level. It has five wards and is bordered to China on North and Jugal Rural Municipality, Barabise Municipality and Dolakha district on west, south and east respectively. There are five hydropower projects in the area including operational and under construction ones. The rural municipality is famous tourist destination for adventurous activities like rafting, bungee jump, canoeing etc. There are 10 major landslides recorded in the area in last two years (DRR portal) making it highly vulnerable to landslide hazard.

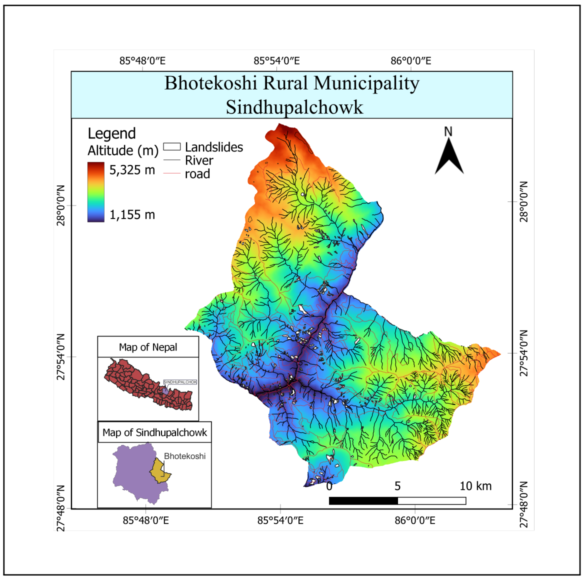


Figure 1: Bhotekoshi rural municipality location map

3. Materials and Methods

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

Data were collected from the satellite images, aerial photographs, research papers, government reports and websites.1 shows the data and their sources used in this study.The selection and preparation of the LCFs database is a crucial step in achieving high accuracy of the landslide susceptibility model in predicting landslide risk areas. In this work, different landslide conditioning factors are used such as slope, aspect, elevation, plan curvature, profile curvature, TWI, TRI, distance from roads, distance from river, normalized difference vegetative index, landcover data are used. QGIS was used to process DEM data to prepare Slope, Aspect, Plan and profile curvatures, TRI, TWI and elevation data. QGIS was used again to process topographical data of road and river to create distance to road and distance to river data.1 shows the different landslide causative factors taken, their resolutions and their sources. The landslide causative factors having

Table 1: Landslide causative factors

S.N.	LCFs	Resolution	Source
1	Slope	12.5 x 12.5	DEM
2	Aspect	12.5 x 12.5	DEM
3	Elevation	12.5 x 12.5	DEM
4	Distance to river	12.5 x 12.5	GIS
5	Distance to road	12.5 x 12.5	GIS
6	NDVI	12.5 x 12.5	Sentinel-2
7	TWI	12.5 x 12.5	GIS
8	TRI	12.5 x 12.5	GIS
9	Profile curvature	12.5 x 12.5	GIS
10	Plan curvature	12.5 x 12.5	GIS
11	Landcover	30 x 30	ICIMOD

significant contribution on landslides have been selected based on literature review. The slope gradient has a significant impact on subsurface flow and soil moisture concentration, both of which are directly related to the incidence of landslides[14]. Aspect is another important factor as wind directions, precipitation patterns, sunshine influence, discontinuity orientations, hydrological processes, evapotranspiration, soil moisture concentration, vegetation, and root development are all factors that have direct and indirect effects on landslides which can be impacted by Aspect[15]. The profile curvature influences flow acceleration and slowdown, as well as erosion and deposition. The plan curvature has an impact on flow convergence and divergence. We can better understand the flow through a surface if we take both plan and profile curvature into account. The saturation of the materials is affected by the slope's proximity to the drainage structure. Distance from drainage was taken into account when modeling the impact of runoff on landslide occurrence [16]. On both the topography and the heel of the slope,

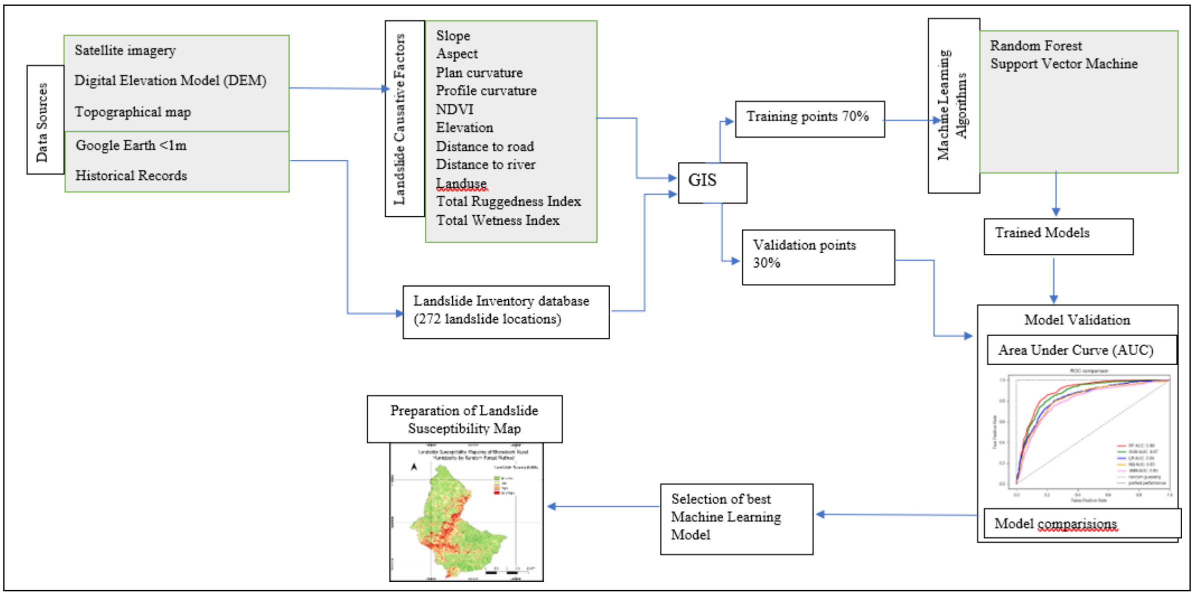


Figure 2: Flow chart of methodology

constructed roads reduce the load. As a result, superficial collapses occur along the roadside in both uphill and downhill slopes, posing a serious road hazard in [17]. Slopes that are left barren are more prone to landslides and so landuse has been used as a factor for susceptibility mapping [18]. Landslide susceptibility is frequently assessed using elevation. Different environmental factors, such as vegetation kinds and rainfall, may be influenced by elevation change. Secondary geomorphometric parameters such as Topographic Wetness Index (TWI) and Terrain Ruggedness Index (TRI) are used to define and measure local relief. TWI and TRI offer new insights into the morphology of landslides, particularly when describing their depositional components [19].

3.2 Landslide Inventory

Satellite images from Google Earth, historical records and database prepared by ICIMOD in 2019 were integrated to prepare landslide inventory. 272 landslides were mapped of which smallest landslide was of size 0.181 sq.km. The data was divided into training and testing sets in the ratio 70% and 30% [11]. The positive and negative values for landslides were labelled as 1/0 respectively.

3.3 Modelling using machine learning techniques

3.3.1 SVM

SVM is a supervised learning model which can handle linearly non-separable and high-dimensional data sets and it deals with binary classification model [12] Consider a dataset $(x_i, y_i)|x_i\, \varepsilon R^n, y_i\, \varepsilon -1, 1_{i=1}^m$. For the linear sepearable data, a separating hyperplane is defined as:

$$\mathrm{y_i(w \times x_i) + b \geq 1 - \xi} \tag{1}$$

where w is coefficient vector determines the orientation of the hyper plane, b is intercept of hyperplane and ξ_i is the positive slack variable. The optimal hyperplane is obtained by solving the Lagrangian multipliers,

$$Minimize \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i x_j) \tag{2}$$

3.3.2 RF

A random vector i_k is generated and distributed to all other trees while employing classification problem in RF method. Bagging technique is applied to the decision trees. Predictions for unseen samples x' can be created after training by summing the predictions from all of the separate regression trees on x':

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x') \tag{3}$$

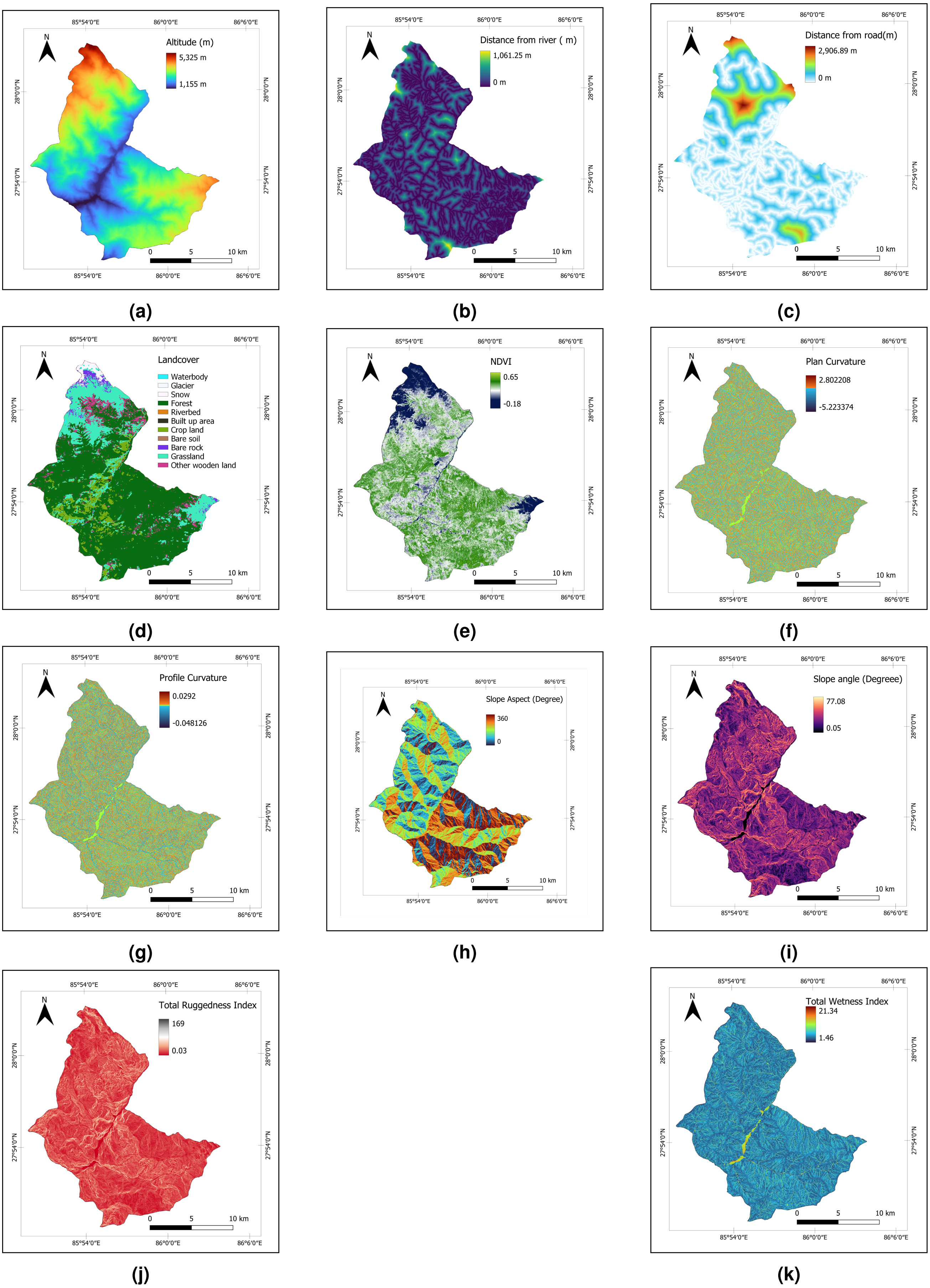


Figure 3: Landslide Causative Factors (a) Altitude (b) Distance from river (c) Distance from road (d) Landcover (e) NDVI (f) Plan Curvature (g) Profile Curvature (h) Slope aspect (i) Slope angle (j) TRI (k) TWI

3.4 Model performance analysis

In this study, two machine learning models were established using training and testing landslide datasets and landslide susceptibility index for every pixel was determined and visualized with the help of GIS software. Prediction accuracy and performance of the models was done quantitative and graphically by the creation of confusion matrices and their extracted statistics. ROC-AUC curve and RMSE score were applied to evaluate the predictive performance of the models. ROC curve is the curve of specificity vs (1-specificity).

4. Results

4.1 Accuracy assessment and comparison

The model performance was evaluated by the calculation of AUC and RMSE statistics. From the AUC and RMSE method (Table 3, Figure 4), the variation in model performance among MLTs was considerably high. RF (AUC= 90%) and SVM (AUC=89%) had the relatively higher accuracy . The RMSE value also found to be 0.43 and 0.44 where lower the RMSE value better is the result. In our study, by both AUC method and RMSE method, RF model was found to be best model for landslide prediction. Also from gini index method for random forest, elevation was found to have importance factor score of 0.226 followed by NDVI = 0.222, TRI = 0.099, aspect = 0.072, slope = 0.071 and others as shown in Figure 5.

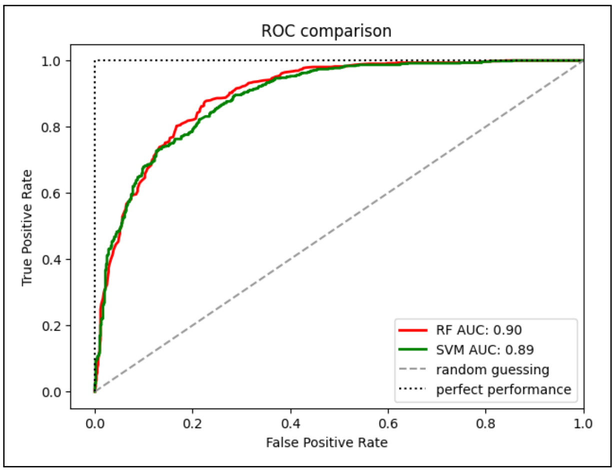
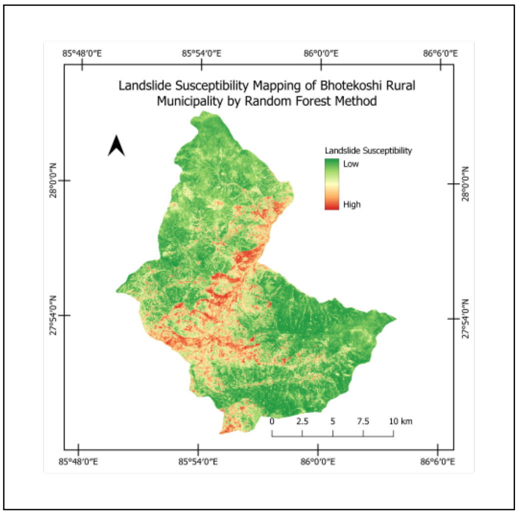


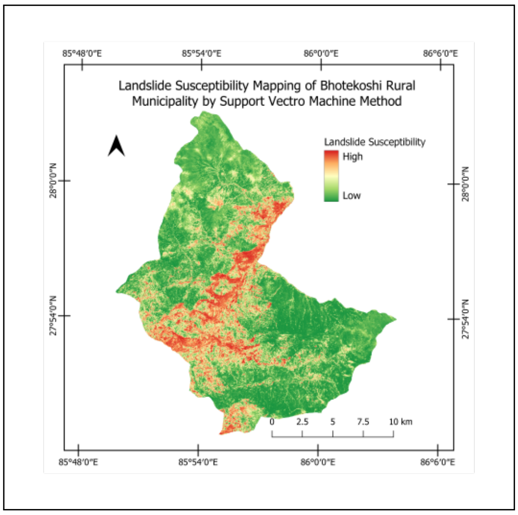
Figure 4: AUC Curve of RF and SVM

4.2 Landslide susceptibility map

From the analysis, Random Forest was selected as the best model for classification of landslide. Using that model and LCFs parameters in raster data for the study area, landslide susceptibility map is prepared. The landslide susceptibility index were visualized in color gradient with green indicating lower susceptibility and red indicating higher landslide susceptibility.



(a) Landslide susceptibility map by Random Forest Method



(b) Landslide susceptibility map by Support Vector Machine Method

5. Discussion

The landslide susceptibility assessment of Bhotekoshi rural municipality was done using remote sensing data, GIS tools and machine learning algorithms. The resulting two susceptibility maps exhibited a consistent spatial distribution pattern and did not differ significantly amongst the models. The area near

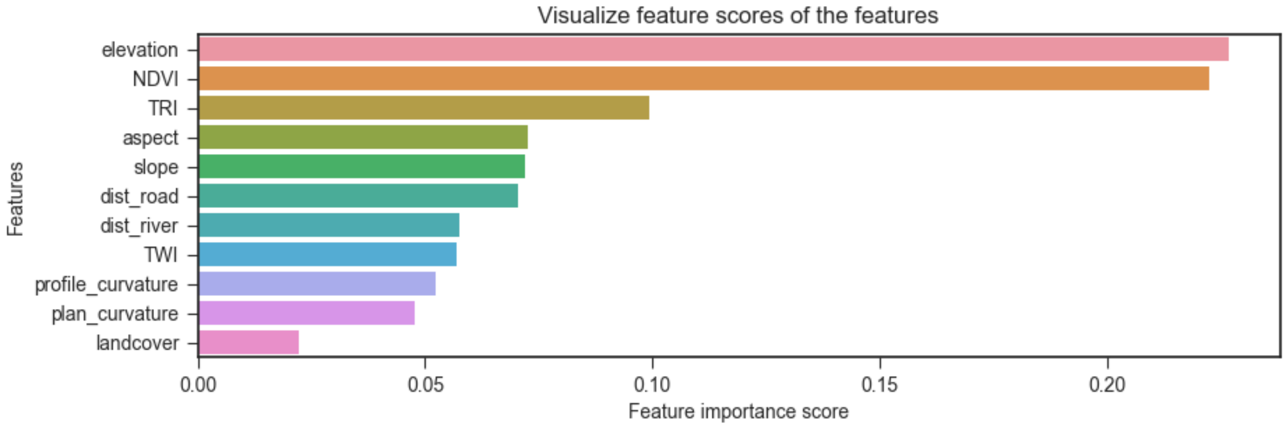


Figure 5: Feature importance

to the Sunkoshi river has higher landslide susceptibility. The construction of hydropower projects and high traffic road along the area may have contributed to the hazard susceptibility. The accuracy of SVM and RF models are very close with SVM having slightly higher accuracy in the comparison done in Sichuan province [20]. In the comparison study at Abha basin, Saudi Arabia, RF model was found to have more accuracy [11] which was in agreement with our result. From the feature analysis, elevation was observed to be the most important feature. This result is in accordance with other studies in other areas with similar topography as our study area [20, 21]. As elevation influences other features such as vegetation, climate, etc, this might have affected the result. Based on above analysis, it is recommended that government and decision makers should compare and analyse the multiple models for ideal susceptibility mapping for practical applications.

6. Conclusions

The comparative study of machine learning models is very useful to predict future landslides. In this study, both RF and SVM methods showed higher accuracy in classification and prediction of landslides with RF having higher accuracy. So RF is best method for Bhotekoshi rural municipality. This map showed that the area near to the Sunkoshi river are highly susceptible to landslides. Additionally, satellite imagery data and GIS tools provided important data for landslide susceptibility analysis. The study provides the overall distribution of landslide susceptibility in Bhotekoshi rural municipality which the decision makers could use to implement sustainable disaster mitigation and preparedness in landslide hazard.

Acknowledgments

The authors are grateful to Institute of Engineering [IOE], Pulchowk Campus, Department of Hydrology and Meteorology, Department of Mines and Geology for the co-ordination and guidance for data collection without which this paper would have been incomplete.

References

[1] David J Varnes. Landslide types and processes. *Landslides and engineering practice*, 24:20–47, 1958.

[2] Le Lin, Qigen Lin, and Ying Wang. Landslide susceptibility mapping on a global scale using the method of logistic regression. *Natural Hazards and Earth System Sciences*, 17(8):1411–1424, 2017.

[3] Abdul Rachman Rasyid, Netra P Bhandary, and Ryuichi Yatabe. Performance of frequency ratio and logistic regression model in creating gis based landslides susceptibility map at lompobattang mountain, indonesia. *Geoenvironmental Disasters*, 3(1):1–16, 2016.

[4] Omar Hamza, Tiago De Vargas, Fernando Eduardo Boff, Yawar Hussain, and K Sian Davies-Vollum. Geohazard assessment of landslides in south brazil: case study. *Geotechnical and Geological Engineering*, 38(1):971–984, 2020.

[5] David Eckstein, Vera Künzel, Laura Schäfer, and Maik Winges. Global climate risk index 2020. *Bonn: Germanwatch*, 2019.

[6] Fausto Guzzetti, Alessandro Cesare Mondini, Mauro Cardinali, Federica Fiorucci, Michele Santangelo, and Kang-Tsung Chang. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1-2):42–66, 2012.

[7] Khamarrul Azahari Razak, Michele Santangelo, Cees J Van Westen, Menno W Straatsma, and Steven M de Jong. Generating an optimal dtm from

airborne laser scanning data for landslide mapping in a tropical forest environment. *Geomorphology*, 190:112–125, 2013.

[8] Ranjan Kumar Dahal, Shuichi Hasegawa, Atsuko Nonomura, Minoru Yamanaka, Takuro Masuda, and Katsuhiro Nishino. Gis-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environmental Geology*, 54(2):311–324, 2008.

[9] Işık Yilmaz. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from kat landslides (tokat—turkey). *Computers & Geosciences*, 35(6):1125–1138, 2009.

[10] Biswajeet Pradhan. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, 38(2):301–320, 2010.

[11] 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*, 12(2):639–655, 2021.

[12] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[13] Biswajeet Pradhan. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using gis. *Computers & Geosciences*, 51:350–365, 2013.

[14] Paolo Magliulo, Antonio Di Lisio, Filippo Russo, and Antonio Zelano. Geomorphology and landslide susceptibility assessment using gis and bivariate statistics: a case study in southern italy. *Natural hazards*, 47(3):411–435, 2008.

[15] Krishna Chandra Devkota, Amar Deep Regmi, Hamid Reza Pourghasemi, Kohki Yoshida, Biswajeet Pradhan, In Chang Ryu, Megh Raj Dhital, and Omar F Althuwaynee. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in gis and their comparison at mugling–narayanghat road section in nepal himalaya. *Natural hazards*, 65(1):135–165, 2013.

[16] Prabin Kayastha, Megh Raj Dhital, and Florimond De Smedt. Landslide susceptibility mapping using the weight of evidence method in the tinau watershed, nepal. *Natural hazards*, 63(2):479–498, 2012.

[17] Ranjan Kumar Dahal, Shuichi Hasegawa, Takuro Masuda, and Minoru Yamanaka. Roadside slope failures in nepal during torrential rainfall and their mitigation. *Disaster mitigation of debris flows, slope failures and landslides*, pages 503–514, 2006.

[18] Mohammad Onagh, VK Kumra, and Praveen Kumar Rai. 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, 2012.

[19] Milena Różycka, Piotr Migoń, Aleksandra Michniewicz, et al. Topographic wetness index and terrain ruggedness index in geomorphic characterisation of landslide terrains, on examples from the sudetes, sw poland.

[20] Ke Xiong, Basanta Raj Adhikari, Constantine A Stamatopoulos, Yu Zhan, Shaolin Wu, Zhongtao Dong, and Baofeng Di. Comparison of different machine learning methods for debris flow susceptibility mapping: A case study in the sichuan province, china. *Remote Sensing*, 12(2):295, 2020.

[21] Subodh Chandra Pal and Indrajit Chowdhuri. Gis-based spatial prediction of landslide susceptibility using frequency ratio model of lachung river basin, north sikkim, india. *SN Applied Sciences*, 1(5):1–25, 2019.