Proceedings of 10th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2021 Month: October Volume: 10

Evaluation and Assessment of Landslide Susceptibility of Sindhupalchowk District using Artificial Neural Network in Regional Scale

Bikesh Manandhar [a], Ananta Man Singh Pradhan [b], Pawan Kumar Bhattarai [c]

[a,c] *Institute of Engineering, Pulchowk Campus, Lalitpur, Nepal*
[b] *Ministry of Energy, Water Resources and Irrigation, Water Resources Research and Development Centre, Government of Nepal, Pulchowk, Lalitpur 44700, Nepal*
Corresponding Email: [a]bikesh.manandhar12@gmail.com, [b]anantageo@hotmail.com

Abstract
Landslides are one of the world's most dangerous natural hazards. On Nepal's steep hills, preventing landslides is very difficult. Furthermore, there are few real-time monitoring systems, either to collect field data or connect to warning systems. Techniques that allow for the rapid identification of landslide-prone areas while avoiding the need for substantial field data are necessary in such a situation. The evaluation of landslide hazard is a complex procedure that frequently depends on an index, a statistical connection, or a physical process. Owing to developments in computer science, machine learning approaches have lately been effectively used in identifying landslide dangers with greater precision. Due to its high number of steep hills, the Sindhupalchowk district is one of the areas of Nepal that are most prone to landslides. The landslide susceptibility map of the research area was generated using the optimal input parameters and independent variables in the form of conditional factors using the multi layered perceptron python script, which normalizes the susceptibility indices from 0.1 to 0.9. This script prepares data, tunes settings for best results, and generates the susceptibility map as a deliverable in three phases. Further, four landslide susceptibility classifications were defined using three threshold percentages (56%, 28%, and 10%) of digitized landslide occurrence (training and validation datasets).The ROC analysis was performed to evaluate the accuracy of the models. The calculated AUC values for success and prediction rate were 84.83% and 84.8%, respectively. According to the categorization, the rate is excellent (0.8-0.9), indicating that the landslide susceptibility is fairly accurate. Landslide susceptibility mapping is an important technique for predicting the likelihood of landslides in steep terrain. As a result, reliable landslide prediction models are essential. This study was conducted for the Sindhupalchowk area and yielded positive results that may be extended to other places after retrieval of trustworthy parameters.

Keywords
Artificial Neural Network,Factor Selection, Landslide susceptibility

1. Introduction

Landslides are among the most catastrophic natural hazards globally, causing substantial financial loss and hundreds of deaths and injuries each year [1].Due to the country's geographical diversity and geological features, and significant rain during the monsoon season, landslides, debris flows and floods are widespread [2]. It is essential to conduct a thorough examination of landslide processes, comprising susceptibility mapping, hazard mapping, and risk assessment, to prevent or monitor concerns caused by mass movement patterns.However, Nepal lacks real-time monitoring systems for gathering field data or connecting to warning systems. As a result, the country's socioeconomic position remains a crucial impediment to assessing susceptibility and risk [3]. In such a circumstance, techniques that allow for fast identification of landslide-prone regions while eliminating the need for extensive field data are required. In general, the safest precautionary strategies are prevention and avoidance [4]. In addition, since most Nepalese people are financially weak and have little property, avoidance is also unlikely. Moreover, for the time being, alerting current settlements and preventing the future could be Nepal's best option. Evaluating landslide hazards is a complex procedure that frequently depends on an index, a statistical connection, or a physical process. Thanks to developments in computer science, machine learning approaches have lately been effectively used in identifying landslide hazards with greater precision. However, there is no agreement on which method or set of methods can provide the most accurate prediction [5, 6]. The most accurate landslide assessment for a given location is dependent on more than just the quality of the data used [7, 8]. It also depends on the modelling techniques employed [9]. However, due to a dearth of historical landslide databases and current geospatial data, relatively little research in Nepal, particularly in mountainous regions,have employed the machine learning methods like neural network, random forest, logistic regression, frequency ratio methods and different other methods. Furthermore, the current findings are stand-alone and have not been integrated into any national or regional databases.

Much effort has gone into understanding the susceptibility of shallow landslides and the rainfall requirements for giving early warnings that can assist in avert landslides [10]. Previous research has used various methods to assess shallow landslide susceptibility, including heuristic, mathematical, and deterministic approaches, assuming that prospective slope collapses have higher chances of occurring under similar issues that caused previous and current instability [11, 12, 13, 14]. This topic is not new, but it has not been thoroughly researched, particularly in Nepal's mountainous regions. However, this study differs from previous studies. It uses a multi-layer perceptron (MLP) neural network script in conjunction with a GIS user interface to perform data preparation, MLP parameter tuning, and susceptibility map and success rate estimation in a relatively short time without sacrificing reliability.

2. Study area

The Sindhupalchowk District is situated in central Nepal, at an elevation of 750–7080 meters, and includes a total area of 2542 km^2. The yearly average rainfall is around 2500 mm, and the temperature fluctuates between 7.5° and 32° Celsius. The terrain is hilly and rich in natural resources, and the people rely primarily on agriculture for a living. The terrain has steep to moderate slopes and geological characteristics that cause slope instability [15]. The location map of the study area is presented in Figure 1.

Figure 1: Map showing the location of Sindhupalchowk District

3. Methods and Data sources

Artificial Neural Network Neural networks are robust machine learning systems consisting of neurons or processing elements arranged in layers. By addressing the non-linear relationship between a landslide inventory and causative factors (CFs), an artificial neural network (ANN) may be utilized to categorize landslides [16]. An ANN comprises many layers along with varying numbers of neurons, and the output value is weighted before being fed into other neurons [17]. MLP neural network (NN) training consists of two phases. Firstly, to compute the difference, the inputs are sent through the hidden layers to generate output values, then compared to pre-values. Second, the attachment weights are fine-tuned to get the most significant outcomes with the slightest fluctuation. Prediction can be improved

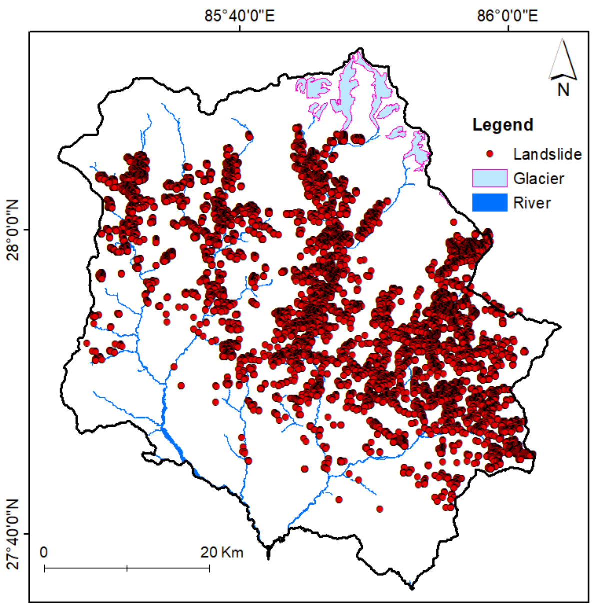

Figure 2: Landslide distribution in study area

by assessing many CFs, although minor factors can lead to over-fitting. Multicollinearity screening can be used to reduce the number of CFs by reducing high-dimensional data [18].

Rainfall triggered landslide inventory The Water Resources Research and Development Centre (WRRDC), Government of Nepal, provided the data for the historic landslide inventory . Other landslide locations were retrieved from satellite imagery and Google Earth^{TM}. 7159 polygons in the research area were digitized in GIS and added to the database as landslides. The pixel bordered by the polygon with the highest elevation value was considered the location of landslide initiation and was used for further analysis. The landslide points taken into account for machine learning are shown in Figure 2.

Causative factors Causative factors are properties that influence the driving and opposing forces of landslide direction of travel and their equilibrium. On a regional scale, there are different geological, geomorphological, and environmental features of the ground. In other words, conditioning variables set the stage for the landslides to occur. Causative factors in landslide investigations are often chosen based on examining of the landslide types and the features of the research region [19]. Elevation, slope angle, plan curvature, and distance to drainage networks are all standard causative variables [20]. However, most researchers produced landslide susceptibility maps by arbitrarily and subjectively selecting causative elements such as geological, geomorphological, hydrological, and anthropogenic factors. As a result, the selection of landslide causative variables and their classifications is critical in LSM research. 16 CFs were considered in this research to determine the landslide susceptibility feeding as input to the ANN model.The conditioning factors were analysed using 30m resolution maps. Table 1 displays the considered data and their several sources.

4. Results and Discussion

The MLP python script was used to produce the landslide susceptibility map of the research region utilizing the ideal input parameters and independent variables in the form of CFs. This script performs in three stages: data preparation for input, parameters optimization process for maximum performance, and outputs the susceptibility map as a deliverable.

Factor selection To choose optimum variables, we tested the 16 CFs for multicollinearity.Variance inflation (VIF) and tolerance (TOL) are two often used multicollinearity indicators [29]. A TOL of 0 and high VIF values suggest a significant multicollinearity issue. Because all 16 CFs were independent of one another, they served as independent variables in the ANN analysis; shallow slide position was utilized as the dependent variable. The filtering resulted in 12 CFs that passed the multicollinearity tests. Their tolerance values and VIF are shown in Table 2.

Table 2: Final multicollinearity test table for filtered 12 CFs

Statistic	R^2	Tolerance	VIF
DEM	0.625	0.375	2.667
Slope	0.301	0.699	1.432
Profile curvature	0.253	0.747	1.338
Plan curvature	0.445	0.555	1.803
Vertical distance to channel	0.230	0.770	1.299
Fault proximity	0.629	0.371	2.694
Drain density	0.247	0.753	1.328
TWI	0.518	0.482	2.074
SPI	0.053	0.947	1.056
Drainage proximity	0.261	0.739	1.354
Landuse	0.001	0.999	1.001
Geology	0.257	0.743	1.347

Table 1: Landslide data layers and their features

Type	Causative factors	Significance	Source
Topographic	Elevation	Weather, vegetative cover and potential energy are all subject to change with elevation resulting in variation in the likelihood of landslides [21]	Department of survey, Government of Nepal/DEM in GIS
	Slope	Steeper slopes have less friction, and landslides are more likely to happen[22]	
	Plan Curvature	The convergence or divergence of slide material and water in the path of landslide velocity is influenced by plan curvature. [23]	
	Profile Curvature	The driving and resistive forces inside a landslide are influenced by profile curvature in the direction of movement. [24]	
	Relative Relief	Relative relief depicts significant fissures in slopes and reflects the energy available for slope failures and soil degradation [25]	
	Topographic Position Index	Identification of a location as a valley, ridge, or flat [11]	
Hydrologic	Drain Proximity	Distance from river lineament	Department of survey, Government of Nepal or Euclidean analysis in GIS
	Stream Power Index	The measure of the stream's erosive strength [26]	
	Sediment Transport Index	depicts the erosion and sedimentation processes [26]	
	Topographic Wetness Index	The impact of topography on hydrological cycle [27]	
	Drainage Density	illustrates the balance of erosive strength of surface runoff and resilience of surface geological formations [11]	
	Vertical distance to streams	Distance from stream centerline in the vertical direction	
Landuse	Landuse	Landslides have an impact on the rate of water flow and the ability of the soil to store water [28]	(Karra et. al. 2021, ESRI INC.)
Geologic	Geology	Each lithological unit is associated with a particular degree of weathering	Department of Mines and Geology
	Fault Proximity	Distance from fault lines	Department of Mines and Geology, GIS

Table 3: MPL modifying parameters

Tuning parameters	Selected values
Hidden layer number: It is possible to utilize one or more number of layers. If just one layer is utilized, this number must be a neuron number in the integer format.	25
Solver: As a solver function, lbfgs, sgd and adam may be utilized.	adam
Activation: identity,logistic, tanh, relu can be selected	tanh
Alpha: parameter of convolution	0.001
Learning rate: Selection between constant, invscaling, and adaptive can be done	adaptive
Learning rate init: It is employed in optimization of weights	0.001
Max iterations: Highest value of sample for convergence	2000
Momentum: It is employed in the gradient descent iteration.	0.6

Tuning MLP The AUC score was used to choose the optimal machine learning parameter for the highest success rate of both training and testing data. The greater the AUC score, the more optimized the simulation. Because of the massive amount of pixels in the research region, our simulated hidden layer size was restricted to a maximum of 25 layers. More extensive layer size simulations would need more powerful computer equipment. Additional parameters were adjusted in trial and error while keeping the layer size constant to improve the simulation. After assessing the AUC values of the input parameters for machine learning, the following parameters were selected for the simulation, as shown in Table 3.

Shallow slide susceptibility Using the MLP python script and the ideal input parameters and independent variables in the form of CFs, the landslide susceptibility map of the research region was generated, which normalizes the susceptibility values from 0.1 to 0.9, as shown in Figure 3. The accuracy of the models was evaluated using ROC analysis. The true positive rate (sensitivity) is shown against the false positive rate in ROC curves (1-specificity).The range of AUC values generally lies in the range of 0.5 to 1. According to Yesilnacar and Topal [30],there is a relationship between predictive performance and AUC value, which may be characterized as 0.5–0.6 (poor), 0.6–0.7 (average), 0.7–0.8 (good), 0.8–0.9 (outstanding), and 0.9–1(excellent).

The ROC analysis script requires susceptibility map input (in raster form), train values (polygon format), and test values as input variables (polygon format). The outputs are the ROC curve and AUC indices of test and train data. The prediction rate was computed from landslide validation. The success rate of LSM was determined by analyzing landslide training data. The LSM was reclassified to generate 100 equivalent areas in a GIS context for this technique. For each class, pixel counts for train and validation landslides were defined. The TPR and FPR were computed using the cumulative pixel values of zones. The AUC was then calculated using the trapezoidal formula[31]. The prediction rate and success rate is shown in Figure 4. The AUC values for success rate and prediction rate were estimated to be 84.83% and 84.8% respectively. According to the categorization, the rate comes into the category of outstanding (0.8-0.9), indicating that the accuracy of the landslide susceptibility is relatively high.

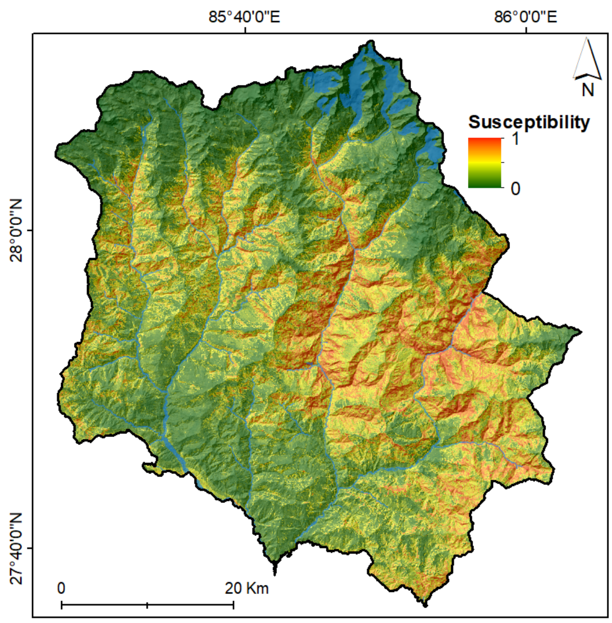

Figure 3: Landslide susceptibility of study area obtained from ANN

Susceptibility classification In this study, the landslide susceptibility map ranges of 0-1 shows a low to high chance of landslide occurrence. The cumulative landslide susceptibility index (LSI) data

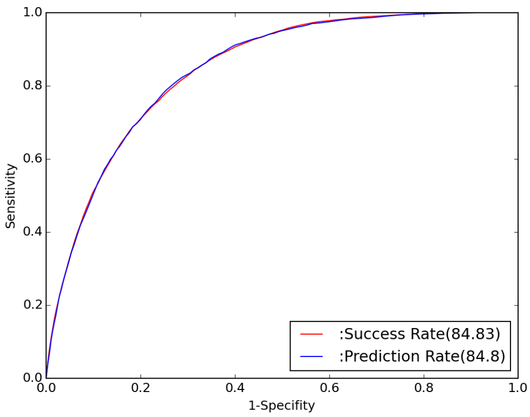

Figure 4: ROC curve for susceptibility simulation

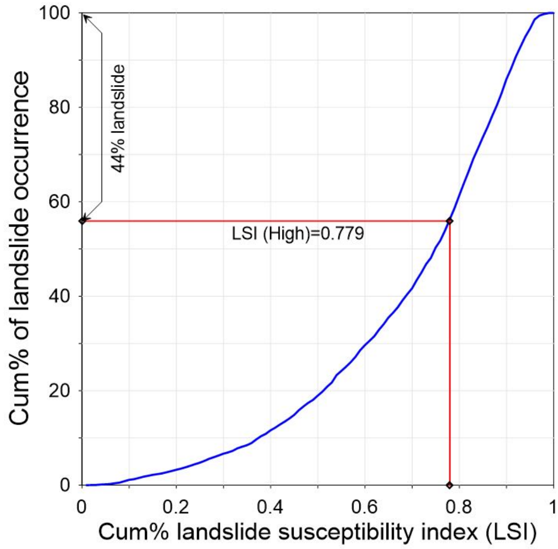

Figure 5: Landslide susceptibility classification using 44% cutoff value

were split into various landslide susceptibility classes to make the map simpler to read. However, this is not an easy task because of lack of statistical criteria for automatically categorising continuous data. Furthermore, the magnitude of categorising continuous data is uncertain since most researchers heavily focus on their judgment to determine class boundaries. Galang's [32] widely used manual classifier approach is utilized to divide the LSI values into four different susceptibility zones (Figure 6 and Figure 5). According to this categorization technique, higher landslide susceptibility classes should account for most landslide events, and successive threshold values are halved from prior levels [32]. As illustrated in Figure 5,, the landslide occurrence map was contrasted to the landslide susceptibility map, and the overall result of recorded landslides vs accumulated LSI numbers was calculated. Four landslide susceptibility classes were identified using three threshold percentages (56%, 28% and 10%) of observed landslide incidence (training and validation datasets).A very high susceptibility area was defined as an area that accounted for 56% of the overall landslide incidence data. The high susceptibility area was denoted by an LR cutoff value of 0.779, which was utilized for the landslide runout study. The four classes divided are labelled as very low (< 10%), low (10-28%), moderate (28-56%) and high (> 56%). A total of 62.52% of the region was classified as very low, 17.27% as low, 13.00% as moderate and 7.20% of area as highly vulnerable. The regions of the four susceptibility classes, as well as their shallow slide distributions, are listed in Table 7.

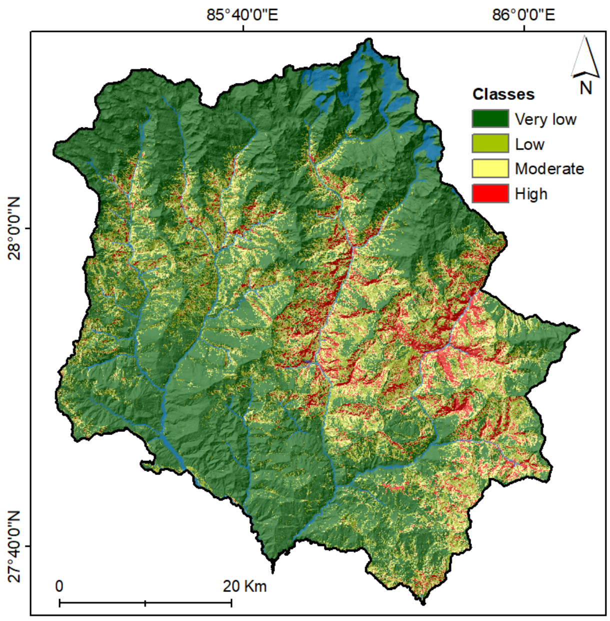

Figure 6: Classified shallow slide susceptibility map

Shallow sliding susceptibilities were estimated in this work utilizing well-known back-propagation ANN machine-learning methods. Depending on the location of the shallow slide and the CFs, the susceptibility mapping depicts the geographical likelihood of a shallow slide. The current study, however, has some limitations: (1) The study area has limited parametric records, with shallow landslides being one of them; and (2) As in other geographic modeling studies focusing on natural break categorization, the shallow slide susceptibility characterization is relative.; (3) Similarly, TWI and SPI categories derived from DEM

Zone	Susceptibility value	Percentage of shallow slide	Percentage of susceptible area
High	>0.779	44	7.20
Moderate	0.586-0.779	28	13.00
Low	0.379-0.586	18	17.27
Very low	<0.379	10	62.52

Figure 7: Shallow slide and terrain distributions based on shallow slide susceptibility values

will inherently include DEM mistake; (4) In addition, causative factors are scarce for the study field, from which impactful factors are sorted and used in the research [11].

The primary benefit of the suggested model is (1) Developing neural network models require less statistical training than conventional models, and (2) it is an essential and quick approach for preliminary forecasting. The proposed model might be utilized as a starting point for developing preliminary warning systems, but long-term verification and subsequent upgrading are necessary.

5. Conclusion

Despite several research, there is no one-size-fits-all approach for either predicting or preventing its incidence. We may, however, avoid losses by following susceptibility maps based on the available causal variables. Landslide susceptibility mapping is an essential technique for predicting the likelihood of landslide events in mountainous terrain. As a result, high-quality landslide prediction models are critical. The Sindhupalchowk district is one of the areas of Nepal that are most prone to landslides due to its high number of steep hills. In hilly areas of Nepal where there is a lack of accessible current data, machine learning approaches can be a viable way to evaluate landslide susceptibility since they reduce the influence of subjectivity and allow for better repeatability. The artificial neural network approach of machine learning is used in this work to provide a thorough landslide hazard assessment. The AUC method in the study gave high success and prediction rates indicating that the susceptibility model was reliable and appropriate in the geographic location. This study was conducted for the Sindhupalchowk area and yielded positive results that may be extended to other places if reliable parameters are obtained. In future rounds of the study, the approach must be validated. Local authorities can use hazard information to manage their land use and avoid landslide-prone areas. The resulting landslide hazard zone map will be a significant resource in the Sindhupalchowk district's regional planning. It will be extremely beneficial in determining human relocation and settlement depending on the frequency of landslide occurrence.

References

[1] Michael J Crozier and Thomas Glade. Landslide Hazard and Risk: Issues, Concepts and Approach. In *Landslide Hazard and Risk*, pages 1–39. 2012.

[2] Bhubanesh Kumar Pradhan. Disaster Preparedness for Natural Hazards: Current Status in Nepal. Technical report, 2007.

[3] Irasema Alcántara-Ayala. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology*, 47(2-4):107–124, 2002.

[4] Elliott C. Spiker and Paula L. Gori. National landslide hazards mitigation Strategy - A framework for loss reduction. *US Geological Survey Circular*, (1244):1–54, 2003.

[5] F Guzzetti, M. Cardinali, P. Reichenbach, and A. Carrara. Comparing landslide maps: A case study in the upper Tiber River basin, central Italy. *Environmental Management*, 25(3):247–263, 2000.

[6] Alberto Carrara and Richard J. Pike. GIS technology and models for assessing landslide hazard and risk, 2008.

[7] Dieu Tien Bui, Biswajeet Pradhan, Owe Lofman, and Inge Revhaug. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and nave bayes models. *Mathematical Problems in Engineering*, 2012, 2012.

[8] Mustafa Neamah Jebur, Biswajeet Pradhan, and Mahyat Shafapour Tehrany. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment*, 152:150–165, 2014.

[9] Işik Yilmaz. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Computers and Geosciences*, 35(6):1125–1138, 2009.

[10] Pietro Aleotti. A warning system for rainfall-induced shallow failures. *Engineering Geology*, 73(3-4):247–265, jun 2004.

[11] H. R. Pourghasemi, H. R. Moradi, S. M. Fatemi Aghda, C. Gokceoglu, and B. Pradhan. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arabian Journal of Geosciences*, 7(5):1857–1878, 2014.

[12] Ananta Man Singh Pradhan and Yun Tae Kim. Relative effect method of landslide susceptibility zonation in weathered granite soil: A case study in Deokjeok-ri Creek, South Korea. *Natural Hazards*, 72(2):1189–1217, 2014.

[13] Paola Reichenbach, Mauro Rossi, Bruce D. Malamud, Monika Mihir, and Fausto Guzzetti. A review of statistically-based landslide susceptibility models, may 2018.

[14] C. J. Van Westen, N. Rengers, and R. Soeters. Use of geomorphological information in indirect landslide susceptibility assessment. *Natural Hazards*, 30(3):399–419, nov 2003.

[15] Suchita Shrestha, Tae Seob Kang, and Madan Krishna Suwal. An ensemble model for co-seismic landslide susceptibility using GIS and random forest method. *ISPRS International Journal of Geo-Information*, 6(11), 2017.

[16] H. Gómez and T. Kavzoglu. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Engineering Geology*, 78(1-2):11–27, apr 2005.

[17] Dawei Han, Terence Kwong, and Simon Li. Uncertainties in real-time flood forecasting with neural networks. *Hydrological Processes*, 21(2):223–228, jan 2007.

[18] Ananta Man Singh Pradhan, Seung Rae Lee, and Yun Tae Kim. A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan, Korea. *Landslides*, 16(3):647–659, mar 2019.

[19] Lulseged Ayalew, Hiromitsu Yamagishi, Hideaki Marui, and Takami Kanno. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology*, 81(4):432–445, 2005.

[20] Jie Dou, Hiromitsu Yamagishi, Hamid Reza Pourghasemi, Ali P. Yunus, Xuan Song, Yueren Xu, and Zhongfan Zhu. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Natural Hazards*, 78(3):1749–1776, may 2015.

[21] Murat Ercanoglu and Candan Gokceoglu. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Engineering Geology*, 75(3-4):229–250, 2004.

[22] F. Catani, D. Lagomarsino, S. Segoni, and V. Tofani. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13(11):2815–2831, 2013.

[23] Anthony Young, M. A. Carson, and M. J. Kirkby. Hillslope Form and Process. *The Geographical Journal*, 139(1):140, 1973.

[24] Gregory C. Ohlmacher. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology*, 91(2-4):117–134, may 2007.

[25] Ananta Man Singh Pradhan and Yun Tae Kim. An artificial intelligence-based approach to predicting seismic hillslope stability under extreme rainfall events in the vicinity of Wolsong nuclear power plant, South Korea. *Bulletin of Engineering Geology and the Environment*, 80(5):3629–3646, 2021.

[26] I. D. Moore and G. J. Burch. Sediment Transport Capacity of Sheet and Rill Flow: Application of Unit Stream Power Theory. *Water Resources Research*, 22(8):1350–1360, 1986.

[27] Taskin Kavzoglu, Emrehan Kutlug Sahin, and Ismail Colkesen. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3):425–439, 2014.

[28] R. K. Dahal, S. Hasegawa, A. Nonomura, M. Yamanaka, and S. Dhakal. DEM-based deterministic landslide hazard analysis in the Lesser Himalaya of Nepal. *Georisk*, 2(3):161–178, 2008.

[29] Ananta Man Singh Pradhan, Seung Rae Lee, and Yun Tae Kim. A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan, Korea. *Landslides*, 16(3):647–659, 2019.

[30] E. Yesilnacar and T. Topal. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, 79(3-4):251–266, jul 2005.

[31] Bernard Liengme. A Guide to Microsoft Excel for Scientists and Engineers. *IEEE Electrical Insulation Magazine*, 15(1):35–35, 2005.

[32] Jeffrey Sterling Galang and Stephen P Prisley. A Comparison of GIS Approaches to Slope Instability Zonation in the Central Blue Ridge Mountains of Virginia A Comparison of GIS Approaches to Slope Instability Zonation in the Central Blue Ridge Mountains of Virginia. *October*, 2004.