

Landslide susceptibility mapping of Phewa catchment using multilayer perceptron artificial neural network

Ananta Man Singh Pradhan and Yun-Tae Kim*

Department of Ocean Engineering, Geosystems Engineering Laboratory, Pukyong National University, KOREA

Abstract

The aim of this study was to prepare landslide susceptibility mapping technique using multilayer perceptron artificial neural network (MLP-ANN) and then to apply this method to Phewa catchment in western Nepal. To determine the effect of causative factors on landslides, data layers of aspect, elevation, slope, internal relief, slope shape, drainage proximity, drainage density, stream power index, topographic wetness index, sediment transport index, land cover and geology were analysed in R-statistical package and final map was produced using geographical information system environment. A GIS-based landslide inventory map of 88 landslide locations was prepared using data from previous reports and satellite image interpretation. A MLP-ANN model was generated from a training set consisting of ~70% randomly selected landslide in the inventory map, with the remaining ~30% landslides used for validation of the susceptibility map. According to analysis, the model had a success rate of 82.1% and the prediction accuracy of 91.4%, indicating a good performance.

Key words: Artificial neural network, GIS, Landslide susceptibility

Introduction

Landslides generate greater annual losses of property than any other type of natural disaster (García et al., 2008). Landslides constitute a major natural hazard in Nepal mainly due to the unique combination of active tectonic setting, high rates of weathering and abundant rainfall, aggravated by human interference in the form of rapid urbanization and infrastructure development (Deoja et al., 1991; Dhital, 2000). In Nepal, a significant number of landslides occur each year (as many as 12,000). Furthermore, due to climatic changes and increase in the frequency of extreme events landslide problems have aggravated (Shiwakoti, 2000). Landslides and floods alone claimed more than 8061 human lives during the period 1983–2011 (DWIDP, 2012). The mountainous terrain in Nepal Himalaya is generally characterized by steep slopes, high relief, highly weathered and densely jointed rock with unfavorable hydro-geological conditions with respect to slope stability. The stability condition of the slopes is further affected by intense human activities and improper land use practices (Sikrikar et al., 1998).

In order to mitigate or control the problems caused by landslides, a systematic study of landslides, including inventory mapping, susceptibility mapping, hazard mapping and risk assessment have to be undertaken (Upreti & Dhital, 1996). Researchers have developed different methods to assess landslide susceptibility or landslide hazard and risk. The basic concept of determining

landslide-prone areas (zones) using the spatial distribution of factors related to instability processes (without temporal factors) was first introduced by R adbruch (1970). Various methods to prepare landslide susceptibility and hazard maps (models) using statistical methods and GIS tools have been developed over the last decade (Guzzetti et al., 2005; Van Westen et al., 2003). Among recent models for landslide susceptibility mapping, a soft computing technique such as artificial neural networks is widely used in pattern recognition (Lee & Pradhan, 2007).

Phewa Lake is the travelers' focal point in Pokhara, and is the second largest lake in Nepal. Phewa Lake has national as well as local importance because of its rich biodiversity, proximity to the Pokhara city, and significance for socioeconomic. According to Pokharel (2008), the area of the lake has reduced from 10 km² in 1957 to 4.4 km² in 1998. It is assumed that the high amount of sediment intake during past big landslide events is one of the causes for lake area reduction (FEED Nepal P Ltd., 2014).

This study focuses on the use of a MLP-ANN to quantitatively model the relationship between landslide occurrence and causative factors (CFs), which could be used for land use planning and as a base for further developments of research related to landslide hazard.

*Corresponding author, email address: yuntkim@pknu.ac.kr

Materials and Methods

Study area

The Phewa Lake watershed area is situated west of Pokhara City about 200 km northwest of Kathmandu (Fig. 1), covering an area of about 123 km². The catchment area is a NW-SE elongated Valley, lying between latitude 28°11' to latitude 28°18' and longitude 83°48' to longitude 83°59'. The main tributary river to Phewa Lake is Harpan Khola. Measuring 2509 m. Panchase is the highest peak, situated at the western part of the catchment boundary and Naudanda and Sarangkot are other important mountain peaks in the northern rim of the catchment.

The study area is located within the Pokhara Valley, an intramontane basin situated at the foot hills of the Annapurna Himalaya range. Several episodes of debris flow events during the Holocene have dammed the river channels, forming a number of lakes in the valley. Phewa is the largest of these lakes, formed by damming of Harpan Khola (Yamanaka et al., 1982). The slopes around the valley are generally very steep and deeply incised by the rivers.

The unprecedented debris flows of September 2007 devastated the Sedi Bagar area. There was a heavy downpour from 4th of September and it intensified further between 7 and 9 September 2007 leading to that disaster (Fig. 2). The debris flow claimed a life and a huge loss of properties.

Spatial database

In this study, the data required for susceptibility mapping were obtained from topographic maps and aerial photographs together with extensive field survey of the area. The effects of causative factors on slope stability were mapped on topographical maps of 1:25,000 and 1:50,000 and aerial photographs of 1:20,000 scales. Geological data were obtained from engineering geological map (1:50,000) published by Department of Mines and Geology.

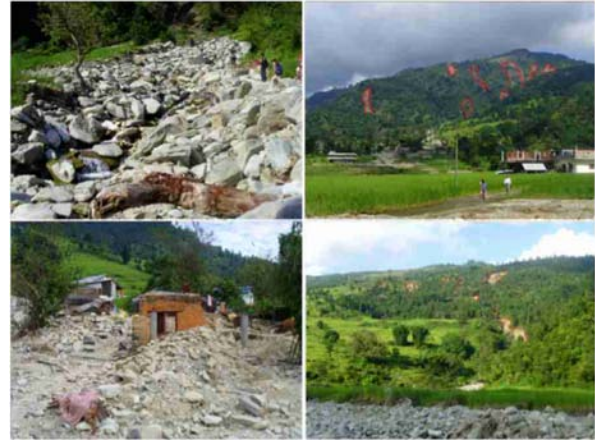


Figure 2 Landslide observed in Phewa catchment

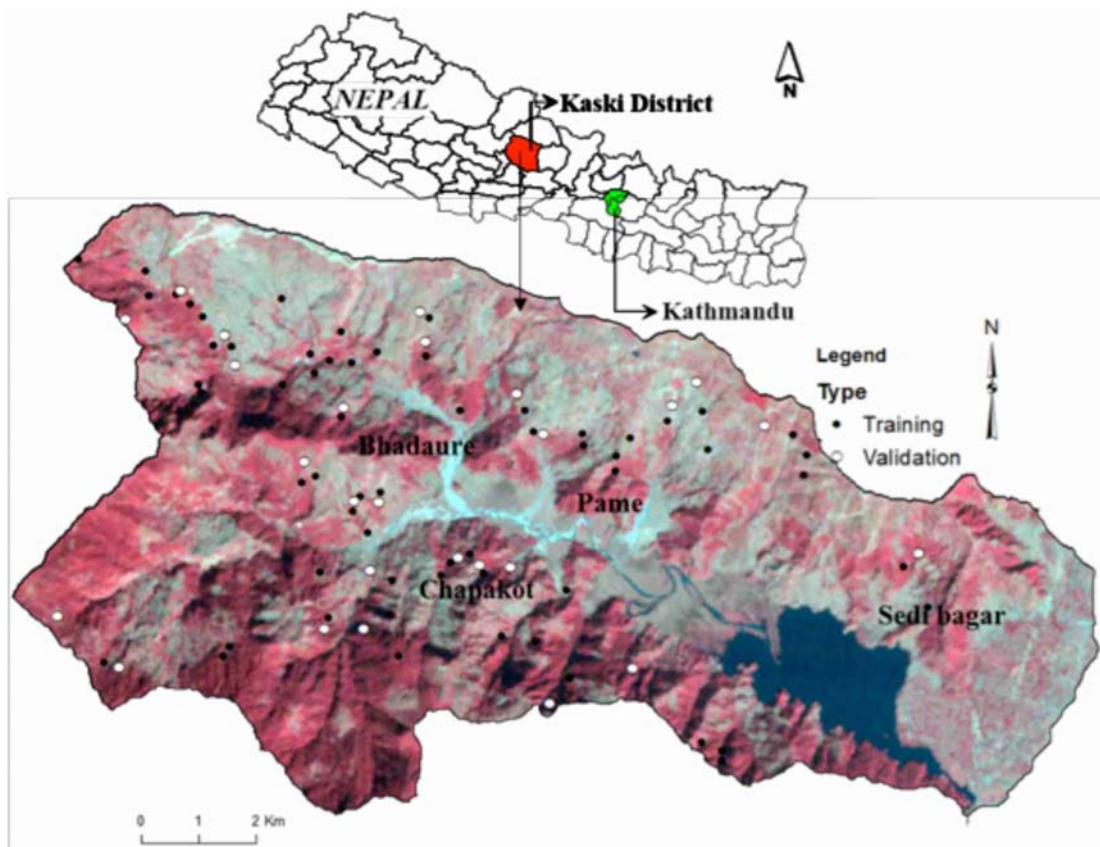


Figure 1 Location of Phewa catchment and distribution of landslides blended on Landsat image FCC 432

The landslides in the study area were identified from comparing LANDSAT 7 ETM+ (15 m) satellite images of 2008 and 2009 (landsat.gsfc.nasa.gov), and they were verified in the field. In total 88 landslides were mapped (Fig.1) and digitized for further analysis. Before analyzing the data, the training landslide data should be selected (Pradhan et al., 2016b). However, there is no exact mathematical rule to determine the required minimum size of the training data-set (Caniani et al., 2008). From these landslides 66 (~70%) randomly selected landslides (training data) were taken for making landslide susceptibility model and 26 (~30%) were used for validation of the model.

Physically, the aspect is related to parameters such as the orientation of discontinuities controlling landslide. The aspect or the direction of maximum slope of the terrain surface is presented in Fig. 3a. In general, elevation is usually associated with landslides by virtue of other factors such as slope gradient, lithology, weathering, precipitation, ground motion, soil thickness and land-use. Richter

et al. (2009) stated that landslides represent an important ecosystem disturbance. The elevation map (Fig. 3b) was produced using the DEM with 30 X 30 m grid size. The slope in the study area varies from 0° to 66.4°. This map was produced automatically in ArcGIS using the DEM with 30 X 30 m grid size (Fig. 3c). Curvature analysis allows dividing the area into concave, convex, and flat surfaces and, consequently, may help to identify zones that exhibit proneness to landslide (Mancini et al., 2010). The curvature was derived from the DEM in ArcGIS 10.2 (Fig. 3d). Internal relief corresponds to the local differences in height within a unit area; the internal relief provides the indication of the potential energy for mass wasting and soil erosion (Ghimire, 2011). It shows major breaks in the slopes of study area. Rank order filter was used in DEM to obtain maximum and minimum altitudes, and internal relief was computed as the difference between the maximum and minimum altitudes per 100 m² in study area (Pradhan & Kim, 2016a). In the study area, relative relief ranges from 0 to 111.78 m (Fig. 3e).

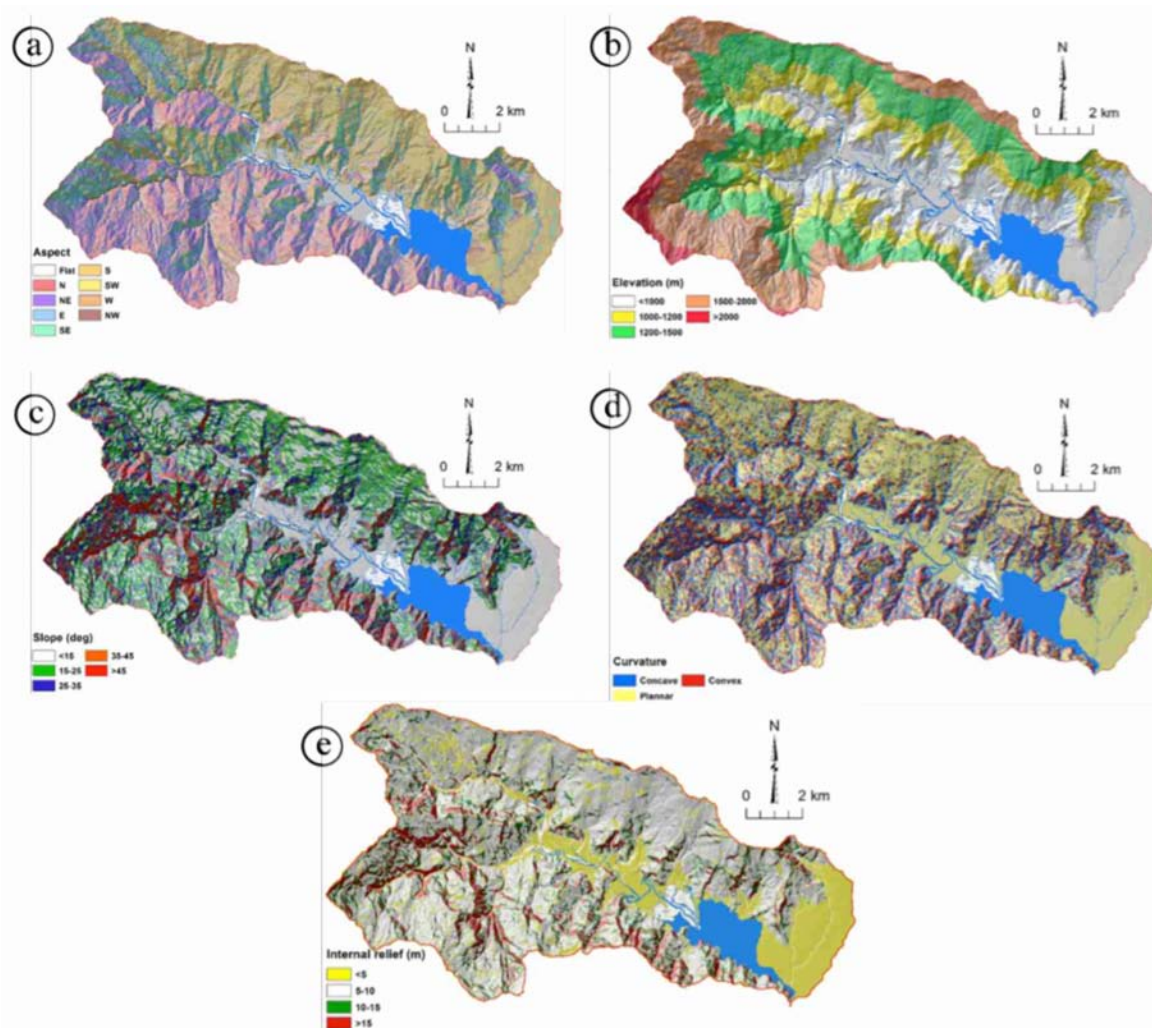


Figure 3 Topographic causative factors: a) aspect; b) elevation; c) slope; d) curvature and e) internal relief

The river map obtained from the Department of Survey of Nepal was used in the ArcGIS 10.2 software to get the proximity to rivers (Fig. 4a). The stream may adversely affect stability by eroding the slope or by saturating the lower portion of the material. Drainage density is the ratio of the total length of the stream to the area of the drainage basin (Fig. 4b). Stream power index (SPI) measures the erosion power of the stream and is also considered as a factor contributing toward stability within the study area (Fig. 4c). The slopes of an area have two components that are slope length (L) and slope steepness (S). The sediment transport index (STI) is a dimensionless sediment transport capacity index that is a nonlinear function of specific discharge and slope was derived by considering the transport capacity limiting sediment flux and catchment evolution erosion theories (Moore & Wilson, 1992) as presented in Fig. 4d. The topographic wetness index (TWI), which combines local upslope contributing area and the entire slope, is commonly used to quantify topographic control on hydrological processes (Fig. 4e).

Land use also plays a significant role in the stability of slope. The land use map was obtained from topographic map published by Department of Survey, Government of Nepal. In the study area, forest and cultivation land are predominantly present which is followed by grass, bush and others (Fig. 5).

The rocks exposed in the northern area of the catchment belong to the Kunchha Formation (DMG, 1998). The formation comprises mostly of gritty phyllites, greenish gray phyllites, talcosicphyllite and gray banded gray quartzite (Fig. 6). The greenish gray banded quartzite and phyllites occur in the upper part of the Formation. Fagfog Quartzite is well exposed in the western part of the catchment. The characteristic feature of this formation is abundant ripple marks. In this study, geological structures were not used for analysis.

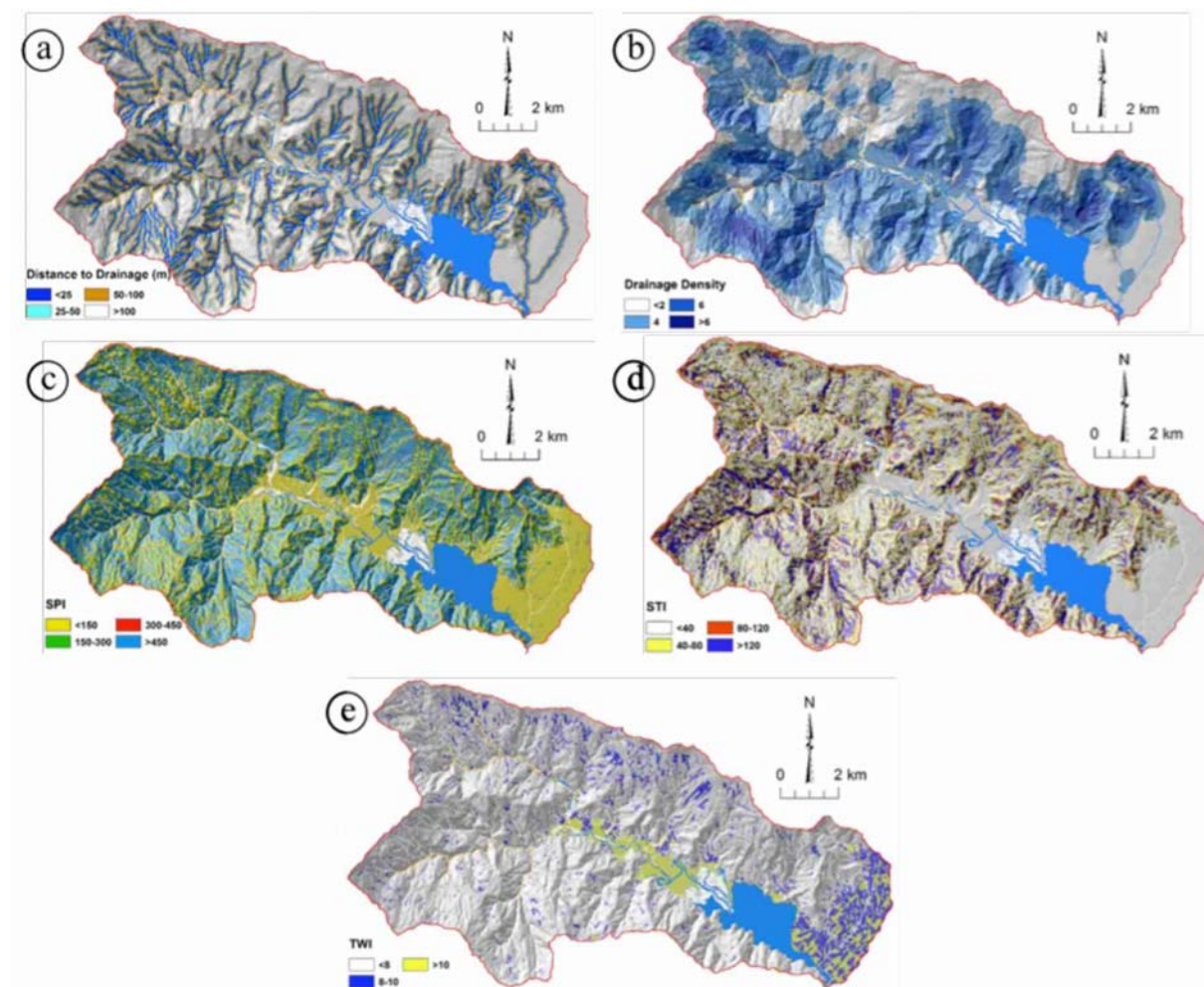


Figure 4 Hydrologic causative factors: a) drainage proximity; b) drainage density; c) SPI; d) STI; e) TWI

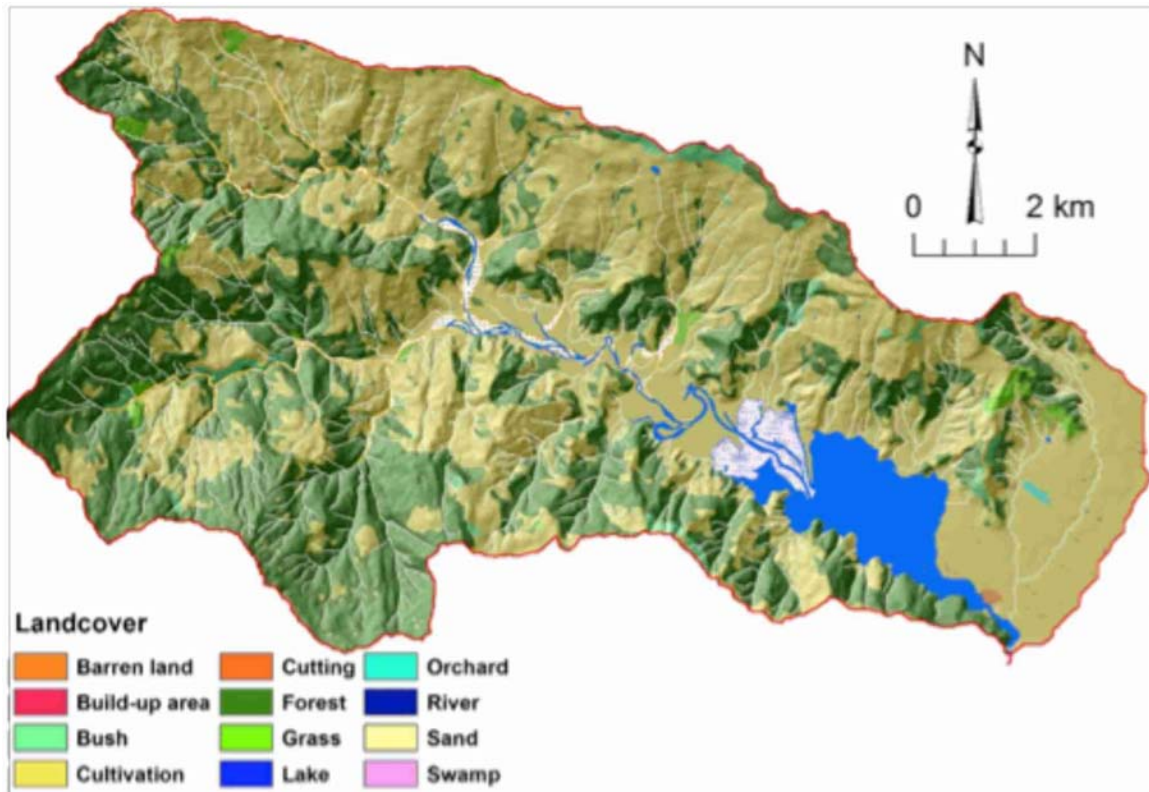


Figure 5 Land-use map of Phewa catchment

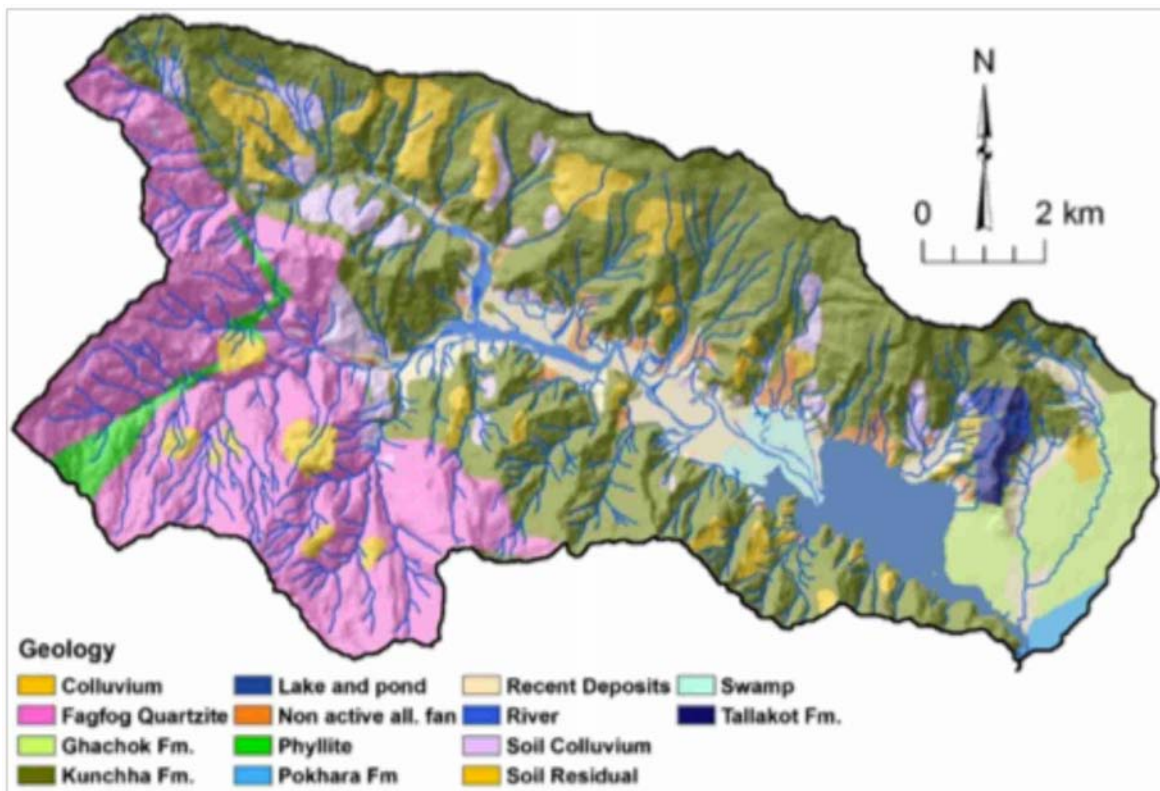


Figure 6 Geological map of Phewa catchment (Modified after DMG, 1998)

Methods

A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs (Garrett, 1994). An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. This method utilizes a supervised learning technique called back propagation for training the network, and a modification of the standard linear perceptron (Rosenblatt, 1961). This technique can distinguish data that are not linearly separable.

If a multilayer perceptron has a linear activation function in all neurons, that is, a linear function that maps the weighted inputs to the output of each neuron, then it is easily proved with linear algebra that any number of layers can be reduced to the standard two-layer input-output model (Haykin, 1998).

This learning algorithm is a multi-layered neural network, which consists of an input layer, hidden layers, and an output layer. The hidden and output layer neurons process their inputs by multiplying each input by a corresponding weight, summing the product, and then processing the sum using a nonlinear transfer function to produce a result.

A neural network consists of a number of interconnected nodes. Each node is a simple processing element that responds to the weighted inputs it receives from other nodes (Paola & Schowengerdt, 1995). The arrangement of the nodes is referred to as the network architecture (Fig. 7).

The two main activation functions used in current applications are both sigmoid, and are described by equation 1

$$y_i = \frac{1}{1 + e^{-J_i}} \quad (1)$$

In which, the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape

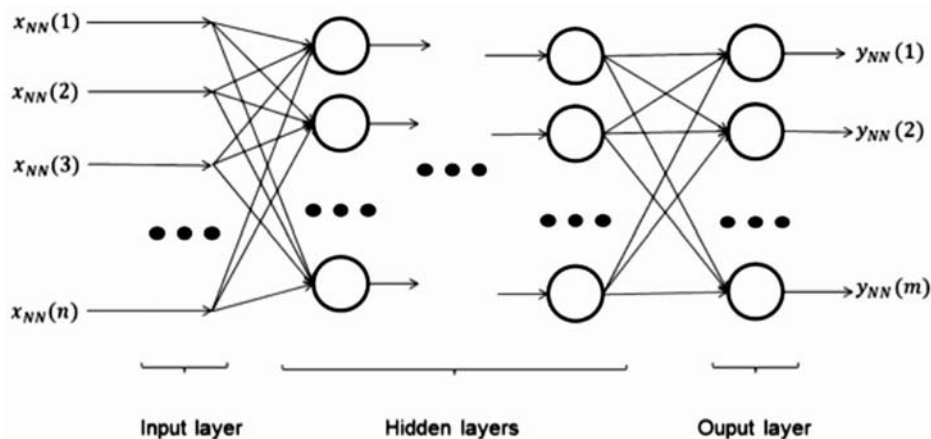


Figure 7 Multilayer neural network

but ranges from 0 to 1. Here y_i is the output of the i^{th} node (neuron) and J_i is the weighted sum of the input synapses. Alternative activation functions have been proposed, including the rectifier and soft plus functions. More specialized activation functions include radial basis functions which are used in another class of supervised neural network models.

Results and Discussion

Once subsets of features were developed using the methodology previously described, MLP-ANN is used to construct the model. The selection of the optimal number of hidden neurons was conducted by setting up a group of neurons in the range of 5 to 150 and examining the RMSEs of the training data. Fig. 8 illustrates that RMSE value of the training data decreases with the increasing number of neurons.

The calculated landslide susceptibility index values, computed using MLP-ANN, were converted into an ARC/INFO grid data. In order to facilitate the map interpretation, a map of landslide hazard zonation is established by dividing the LSI values into different landslide susceptibility classes. However, this is not straight forward, as there are no statistical rules which can categorize continuous data automatically. Moreover, the problem of dividing continuous data into classes remains always unclear, because the most researchers use their own expert opinion to develop class boundaries. Mathematical methods for data classification are usually available in GIS software. In this study, the manual classifier method was used to reclassify the landslide susceptibility index values into four different susceptibility zones. The classification method should satisfy the principle that higher landslide susceptibility classes should capture more or most of the landslide occurrences. The susceptible area which encompassed 70% landslide (training) occurrence data was considered as very high susceptible area. And remaining three classes, i.e. low, moderate and high were classified on the basis of natural break (Jenks) classification available in ArcGIS (Fig. 9).

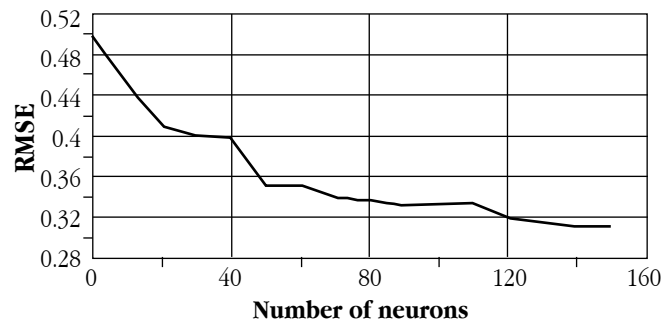


Figure 8 RMSE versus number of neurons for training data to select optimal hidden neurons

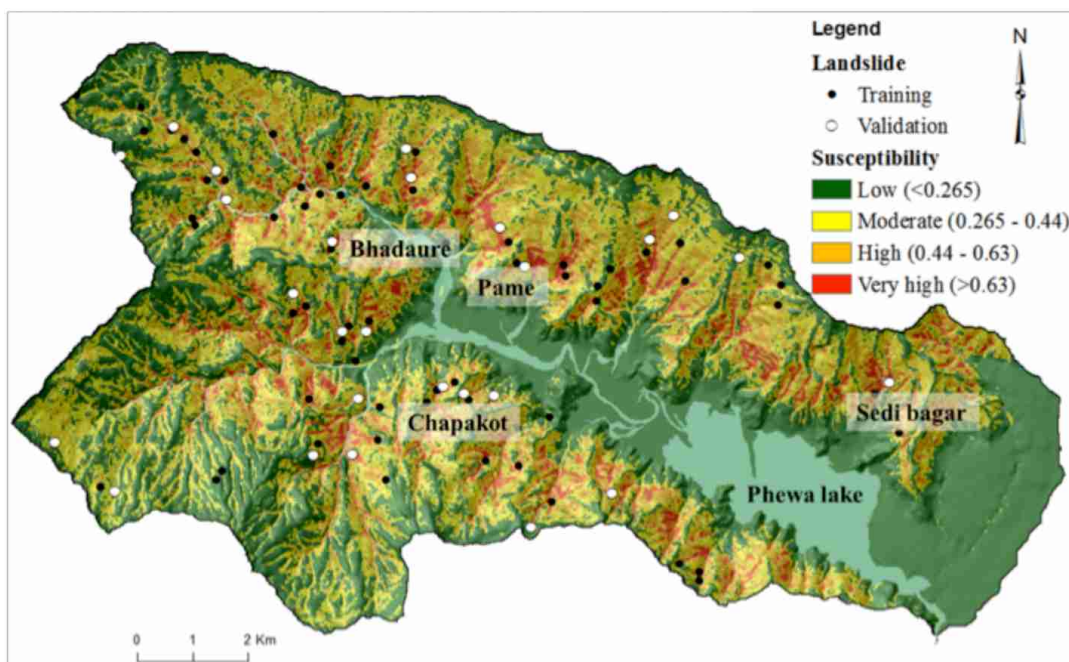


Figure 9 Landslide susceptibility map

Fig. 10 shows the area of the susceptibility class for the landslide susceptibility map prepared by MLP-ANN. A total of 36.52% of the area was categorized as having low susceptibility, 33.9% as moderately susceptible, 19.08% was classified as high and remaining 10.5% was considered to have very high susceptibility.

Verification is a fundamental step in the development of susceptibility and determination of its quality. The quality of a landslide susceptibility model is usually estimated by using independent information that is not available for building the model. Among 88 landslides, ~70% were used to train phrase, the rest of landslides were used as validate the susceptibility map. Receiver characteristic curve (ROC) is applied in many fields to test model performance. The ROC is a useful method to represent the quality of deterministic and probabilistic detection and forecast systems (Swets, 1988). When classifying a grid from the unstable

map, four outcomes are possible (Pradhan & Kim, 2016a). If landslides were observed in a cell that was calculated to be unstable, this was considered a true positive (TP); if landslides were not observed in the area that was calculated to be unstable; it was considered a false positive (FP); if landslides were observed in areas that were calculated to be stable cells, it was considered a false negative (FN); and if no landslides were observed in areas that were calculated to be stable cells, this was considered a true negative (TN). The ROC was calculated by plotting the fraction of positive outcomes that were correctly identified (i.e., the TNs) versus the fraction of positive outcomes that were not correctly identified (i.e., the FNs). The area under the ROC (AUC) can be used as a metric to quantify the overall performance of the model (Hanley & McNeil, 1982) such that the larger the area is, the better the performance of the model becomes.

Chung and Fabbri (2003) distinguished between success- and prediction-rate curves. The success-rate curve is based on a comparison of the susceptibility map with the landslides used in modelling (i.e., the training data). And the prediction-rate measurement is carried out with the validation landslide inventory. In Fig. 11, it can be seen that the AUC value of the success-rate curve was calculated as 82.1%, and AUC value of the prediction-rate was calculated as 91.4%, indicated that the result was excellent.

These values represented that the MLP-ANN had a good performance on the prediction of landslide. It also indicated that the susceptibility map which was obtained through this method was reasonable.

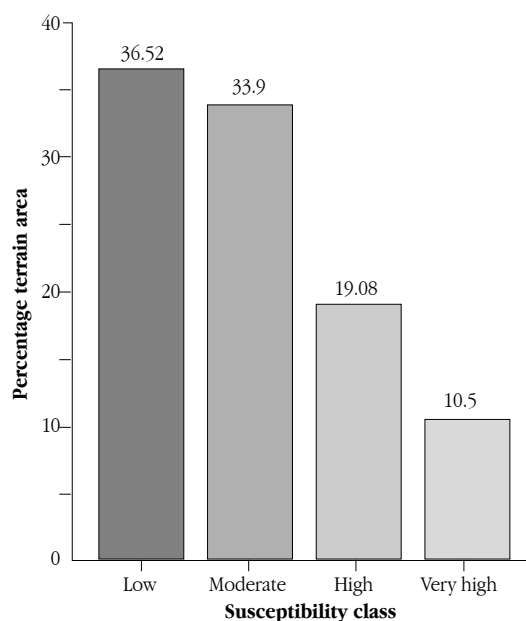


Figure 10 Area of susceptibility class

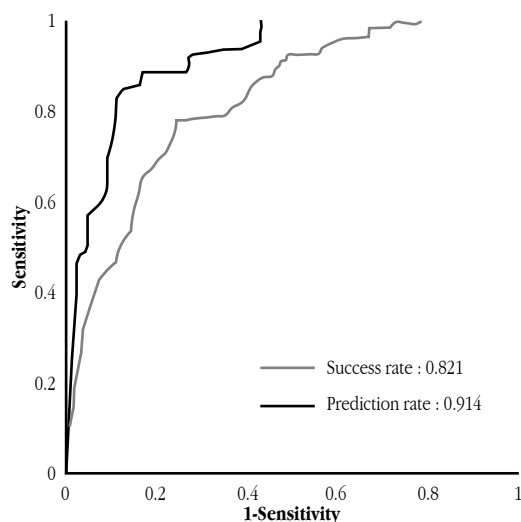


Figure 11 ROC plots for accuracy assessment of the model

Conclusion

Landslides are one of the most hazardous natural disasters, not only in Nepal but around the world. Government and research institutions worldwide have attempted for years to assess landslide hazards and their associated risks and to show their spatial distribution. A multilayer perceptron artificial neural network approach has been used to estimate areas susceptible to landslides using a spatial database for a selected study area in Phewa Lake catchment, Pokhara. The validation of landslide susceptibility maps indicated that artificial neural network can be applied for the landslide susceptibility mapping successfully. In order to get accurate susceptibility map, this model included slope, aspect, elevation, internal relief, and slope shape, drainage distance, stream power index, topographic wetness index, sediment transport index, land use and geology as the inputs. The AUCs of the success rate and prediction rate were 82.1% and 91.4% respectively, which indicate that artificial neural network is suitable for the prediction of landslides in the study area.

Acknowledgements

The authors would like to thank reviewers for their insightful comments on the paper, as these comments led us to an improvement of the work. This research was supported by the Public Welfare and Safety Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning (grant No. 2012M3A2A1050977), a grant (13S CIPS04) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea Government and Korea Agency for Infrastructure Technology Advancement (KAIA) and the Brain Korea 21 Plus (BK 21 Plus).

References

- Caniani, D., Pascale, S., Sdao, F., & Sole, A. (2008). Neural networks and landslide susceptibility: a case study of the urban area of Potenza. *Natural Hazards*, 45(1), 55-72.
- Chung, C.J. & Fabbri, A.G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30, 451-472.
- Deoja, B., Dhital, M.R., Thapa, B., & Wagner, A. (1991). *Mountain risk engineering handbook*. In: ICIMOD, Kathmandu, 857 pp.
- Dhital, M. R. (2000). An overview of landslide hazard mapping and rating systems in Nepal. *J Nepal Geol Soc*, 22, 533-538.
- DMG (1998). *Engineering and environmental geological map of Pokhara Valley, scale 1:50,000*. Department of Meteorology and Hydrology, Kathmandu.
- DWIDP (Department of Water Induced Disaster Prevention) (2012). *Annual Disaster Review 2011*. Ministry of Irrigation, Government of Nepal, Kathmandu, Nepal, 24 pp.
- FEED (2014). *Development of ecosystem based sediment control technique and design of siltation dam to protect Phewa lake: Herpan Khola watershed Kaski*. Kathmandu: FEED (P) Ltd.

- García-Rodríguez, M. J., Malpica, J. A., Benito, B., & Díaz, M. (2008). Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. *Geomorphology*, 95(3), 172-191.
- Garrett, J. (1994). Where and why artificial neural networks are applicable in civil engineering. *Journal of Computing in Civil Engineering*, 8, 129–130.
- Ghimire, M. (2011). Landslide occurrence and its relation with terrain factors in the Siwalik Hills, Nepal: case study of susceptibility assessment in three basins. *Nat Hazards*, 56(1): 299–320.
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72, 272–299.
- Hanley, J. A., & McNeil, B.J. (1982). The meaning and use of the area under receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall.
- Lee, S. & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia, using frequency ratio and logistic regression models. *Landslides*, 4, 33–41.
- Mancini, F., Ceppi, C., & Ritrovato, G. (2010). GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy. *Nat Hazards Earth SystSci*, 10, 1851–1864.
- Moore, I.D., & Wilson, J.P. (1992). Length–slope factors for the Revised Universal Soil Loss Equation: simplified method of estimation. *Journal of Soil Water Conservation* 47, 423–428.
- Pradhan, A.M.S., & Kim, Y.T. (2016a). 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.
- Pradhan, A.M.S., Kang, H.S., Lee, S., & Kim, Y.T. (2016b). Spatial model integration for shallow landslide susceptibility and its runoff using a GIS-based approach in Yongin, Korea. *Geocarto International*, 420–441.
- Paola, J.D. & Schowengerdt, R.A. (1995). A review and analysis of back propagation neural networks for classification of remotely sensed multi-spectral imagery. *International Journal of Remote Sensing*, 16, 3033–3058.
- Pokharel, R. K. (2008). *Nepal's Community Forestry Funds: Do They Benefit the Poor?*. South Asian Network for Development and Environmental Economics.
- Radbruch, D.H. (1970). *Map of relative amounts of landslides in California*. US GeolSurv Open-File Rep70-1485, p 36, map scale 1:500,000. *US GeolSurv Open-File Rep*, pp 85–585.
- Richter, M., Diertl, K., Emck, P., Peters, T., & Beck, E. (2009). Reasons for an outstanding Plant diversity in the tropical Andes of Southern Ecuador. *Landscape Online*, 12, 1–35.
- Rosenblatt, F. (1961). Principles of neurodynamics. Perceptron and the theory of brain mechanisms (No. VG-1196-G-8). Cornell Aeronautical Lab Inc Buffalo NY.
- Shiwakoti, D.R. (2000). An assessment of soil loss and natural hazards in Nepal. *Journal of Nepal Geological Society*, 21, 41–48.
- Sikrikar, S.M., Rimal, L.N., & Jä'äbager, S. (1998). Landslide hazard mapping of Phewa Lake catchment area, Pokhara, central west Nepal. *Journal of Nepal Geological Society*, 18, 335–341.
- Swets, J.A., (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.
- Upreti, B.N. & Dhital, M.R. (1996). Landslide Studies and Management in Nepal. *International Centre for Integrated Mountain Development (ICIMOD) Kathmandu*, Nepal, 87 pp.
- Van Westen, C. J., Rengers, N., & Soeters, R. (2003). Use of geomorphological information in indirect landslide susceptibility assessment. *Natural Hazards*, 30(3), 399-419.
- Yamanaka, H., Yoshida, H. & Arita, K. (1982). Terrace landforms and quaternary deposits around Pokhara valley, central Nepal. *Jour. Nepal Geol.Soc.*, 2 (Special Issue), 113–142