APPLICATION OF A DATA MINING MODEL FOR LANDSLIDE HAZARD MAPPING

BISWAJEET PRADHAN[a], SHATTRI MANSOR[b], SARO LEE[c,] MANFRED F. BUCHROITHNER[d]

[a,b]Institute for Advanced Technologies (ITMA), University Putra Malaysia, 43400, UPM, Serdang, Malaysia, Tel: 603-8946-7543 Fax: 603-8656-6061-biswajeet@putra.upm.edu.my (contact author)
[c]Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources (KIGAM) 30, Kajung-Dong, Yusung-Gu, Taejon, Korea, Tel: 82-42-868-3057 Fax: 82-42-861-9714-leesaro@kigam.re.kr
[d]Institute of Cartography Dresden University of Technology 01062, Dresden, GermanyTel.:(0351) 463-37562-Manfred.Buchroithner@mailbox.tu-dresden.de

KEY WORDS: Landslide; Frequency Ratio; Landslide Hazard; Risk Analysis; GIS; Remote Sensing

ABSTRACT:

This paper deals with landslide hazard and risk analysis using Geographic Information System (GIS) and remote sensing data for Cameron Highland, Malaysia. Landslide locations were identified in the study area from interpretation of aerial photographs and field surveys. Topographical/geological data and satellite images were collected and processed using GIS and image processing tools. There are ten landslide inducing parameters which are considered for the landslide hazards. These parameters are topographic slope, aspect, curvature and distance from drainage, all derived from the topographic database; geology and distance from lineament, derived from the geologic database; landuse from Landsat satellite images; soil from the soil database; precipitation amount, derived from the rainfall database; and the vegetation index value from SPOT satellite images. These factors were analyzed using an advanced artificial neural network model to generate the landslide hazard map. Each factor's weight was determined by the back-propagation training method. Then the landslide hazard indices were calculated using the trained back-propagation weights, and finally the landslide hazard map was generated using GIS tools. Landslide locations were used to verify results of the landslide hazard map and the verification results showed 83.45% accuracy. The verification results showed sufficient agreement between the presumptive hazard map and the existing data on landslide areas.

1. INTRODUCTION

Landslide presents a significant constraint to development in many parts of Malaysia. Damages and losses are regularly incurred because, historically, there has been too little consideration of the potential problems in land use planning and slope management. Landslides are mostly occurred in Malaysia mainly due to heavy rainfall. In recent years greater awareness of landslide problems has led to significant changes in the control of development on unstable land, with the Malaysian government and highway authorities stressing the need for local planning authorities to take landslide into account at all stages of the landslide hazard mapping process. So far, few attempts have been made to predict these landslides or preventing the damage caused by them. Through this prediction model, landslide damage could be greatly decreased. Through scientific analysis of landslides, one can assess and predict landslide-susceptible areas, and thus decrease landslide damage through proper preparation. To achieve this aim, landslide hazard analysis techniques have been applied, and verified in the study area using artificial neural network. In addition, landslide-related factors were also assessed.

There have been many studies carried out on landslide hazard evaluation using GIS; for example, Guzzetti et al.1 summarized many landslide hazard evaluation studies. Recently, there have been studies on landslide hazard evaluation using GIS, and many of these studies have applied probabilistic methods 2,3,4,5,6,7,8,9,10,11,12,13,14,15,16. One of the statistical methods available, the logistic regression method, has also been applied to landslide hazard mapping as described in the literature17,18,19,20,21. There is another method of hazard mapping is the geotechnical method and the safety factor method 22,23,24,25,26. There are other new approach to landslide hazard evaluation using GIS, data mining using fuzzy logic, and artificial neural network methods have been applied in various case studies as reported by many researchers 27,28,29,30,31.

Landslide occurrence areas were detected in the Cameron area, Malaysia by interpretation of aerial photographs and field surveys. A landslide map was prepared from aerial photographs, in combination with the GIS, and this were used to evaluate the frequency and distribution of shallow landslides in the area. Topography and lithology databases were constructed and lineament, land cover and vegetation index value extracted from Landsat TM satellite image for the analysis. Then, the calculated and extracted factors were converted to a 10m × 10m grid (ARC/INFO GRID type). Artificial neural network was applied using the database and landslide hazard map was created. Finally, the map was verified and compared using known landslide locations for quantitative verification.

In the study, Geographic Information System (GIS) software, ArcView 3.3, and ArcGIS 9.0 version software packages were used as the basic analysis tools for spatial management and data manipulation.

2. STUDY AREA

The study (Figure 1) area which is part of districts of Cameron Highland seeing a rapid development with land clearing for housing estate, hotel/apartment causing erosion and landslides. Cameron Highland is a district of Pahang state which is one of the 13 states of the Federation of Malaysia. The study area covers an area of 660 square km and is located near the northern central part of peninsular Malaysia. It is bounded to the north by

Kelantan, west by Perak. Annual rainfall is very high averaging between 2,500 mm to 3,000 mm per year. Two pronounced wet seasons from September to December and February to May. Rainfall peaks between November to December and March to May. The geomorphology of the area consists of undulating plateau stretching about 12 km. The geology of the Cameron highland consists of mostly quaternary and Devonian granite. Many landslides have been recorded along stream scouring the sides of the streams.

3. ARTIFICIAL NEURAL WOREK

An artificial neural network is a "computational mechanism able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping"[32]. The back-propagation training algorithm is the most frequently used neural network method and is the method used in this study. The back-propagation training algorithm is trained using a set of examples of associated input and output values. The purpose of an artificial neural network is to build a model of the data-generating process, so that the network can generalize and predict outputs from inputs that it has not previously seen. 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. An artificial neural network "learns" by adjusting the weights between the neurons in response to the errors between the actual output values and the target output values. At the end of this training phase, the neural network provides a model that should be able to predict a target value from a given input value.

There are two stages involved in using neural networks for multi-source classification: the training stage, in which the internal weights are adjusted; and the classifying stage. Typically, the back-propagation algorithm trains the network until some targeted minimal error is achieved between the desired and actual output values of the network. Once the training is complete, the network is used as a feed-forward structure to produce a classification for the entire data [33].

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. The arrangement of the nodes is referred to as the network architecture (Figure 2). The receiving node sums the weighted signals from all the nodes that it is connected to in the preceding layer. Formally, the input that a single node receives is weighted according to Equation (1).

$$net_j = \sum_i w_{ij} \cdot o_i \qquad (1)$$

where wij represents the weights between nodes i and j, and oi is the output from node j, given by

$$o_j = f(net_j) \qquad (2)$$

The function *f* is usually a non-linear sigmoid function that is applied to the weighted sum of inputs before the signal propagates to the next layer. One advantage of a sigmoid function is that its derivative can be expressed in terms of the function itself:

$$f'(net_j) = f(net_j)(1 - f(net_j)) \qquad (3)$$

The network used in this study consisted of three layers. The first layer is the input layer, where the nodes were the elements of a feature vector. The second layer is the internal or "hidden" layer. The third layer is the output layer that presents the output data. Each node in the hidden layer is interconnected to nodes in both the preceding and following layers by weighted connections[34].

The error, *E*, for an input training pattern, *t*, is a function of the desired output vector, *d*, and the actual output vector, *o*, given by:

$$E = \frac{1}{2}\sum_k (d_k - o_k) \qquad (4)$$

The error is propagated back through the neural network and is minimized by adjusting the weights between layers. The weight adjustment is expressed as:

$$w_{ij}(n+1) = \eta(\delta_j \cdot o_i) + \alpha \Delta w_{ij} \qquad (5)$$

where η is the learning rate parameter (set to η = 0.01 in this study), δ_j is an index of the rate of change of the error, and α is the momentum parameter (set to α = 0.01 in this study).

The factor δ_j is dependent on the layer type. For example,

for hidden layers, $\delta_j = (\sum \delta_k w_{jk}) f'(net_j)$ (6)

and for output layers,

$$\delta_j = (d_k - o_k) f'(net_k) \qquad (7)$$

This process of feeding forward signals and back-propagating the error is repeated iteratively until the error of the network as a whole is minimized or reaches an acceptable magnitude.

Using the back-propagation training algorithm, the weights of each factor can be determined and may be used for classification of data (input vectors) that the network has not seen before. Zhou[35] described a method for determining the weights using back propagation. From Equation (2), the effect

of an output, o_j, from a hidden layer node, j, on the output, o_k, from an output layer (node k) can be represented by the partial derivative of o_k with respect to o_j as

$$\frac{\partial o_k}{\partial o_j} = f'(net_k) \cdot \frac{\partial (net_k)}{\partial o_j} = f'(net_k) \cdot w_{jk} \tag{8}$$

Equation (8) produces both positive and negative values. If the effect's magnitude is all that is of interest, then the importance (weight) of node j relative to another node $j0$ in the hidden layer may be calculated as the ratio of the absolute values derived from Equation (8):

$$\frac{\left|\partial o_k\right|}{\left|\partial o_j\right|} \Big/ \frac{\left|\partial o_k\right|}{\left|\partial o_{j0}\right|} = \frac{\left|f'(net_k) \cdot w_{jk}\right|}{\left|f'(net_k) \cdot w_{j0k}\right|} = \frac{\left|w_{jk}\right|}{\left|w_{j0k}\right|} \tag{9}$$

We should mention that w_{j0k} is simply another weight in w_{jk} other than w_{ik}.

For a given node in the output layer, the results of Equation (9) show that the relative importance of a node in the hidden layer is proportional to the absolute value of the weight connecting the node to the output layer. When the network consists of output layers with more than one node, then Equation (9) cannot be used to compare the importance of two nodes in the hidden layer.

$$w_{j0k} = \frac{1}{J} \cdot \sum_{j=1}^{J} \left|w_{jk}\right| \tag{10}$$

$$t_{jk} = \frac{\left|w_{jk}\right|}{\frac{1}{J} \cdot \sum_{j=1}^{J} \left|w_{jk}\right|} = \frac{J \cdot \left|w_{jk}\right|}{\sum_{j=1}^{J} \left|w_{jk}\right|} \tag{11}$$

Therefore, with respect to node k, each node in the hidden layer has a value that is greater or smaller than unity, depending on whether it is more or less important, respectively, than an average value. All the nodes in the hidden layer have a total importance with respect to the same node, given by

$$\sum_{j=1}^{J} t_{jk} = J \tag{12}$$

Consequently, the overall importance of node j with respect to all the nodes in the output layer can be calculated by

$$t_j = \frac{1}{K} \cdot \sum_{j=1}^{K} t_{jk} \tag{13}$$

Similarly, with respect to node j in the hidden layer, the normalized importance of node j in the input layer can be defined by

$$s_{ij} = \frac{\left|\omega_{ij}\right|}{\frac{1}{I} \cdot \sum_{i=1}^{I} \left|\omega_{ij}\right|} = \frac{I \cdot \left|\omega_{ij}\right|}{\sum_{i=1}^{I} \left|\omega_{ij}\right|} \tag{14}$$

The overall importance of node i with respect to the hidden layer is

$$s_i = \frac{1}{J} \cdot \sum_{j=1}^{J} s_{ij} \tag{15}$$

Correspondingly, the overall importance of input node i with respect to output node k is given by

$$st_i = \frac{1}{J} \cdot \sum_{j=1}^{J} s_{ij} \cdot t_j \tag{16}$$

4. DATA USING GIS AND REMOTE SENSING

The GIS and remote sensing data used in the present study have been shown in Table 1. Accurate detection of the location of landslides is very important for probabilistic landslide hazard analysis. The application of remote sensing methods, such as aerial photographs and satellite images, are used to obtain significant and cost-effective information on landslides. In this study, 1:25,000–1:50,000-scale aerial photographs were used to detect the landslide locations. These photographs were taken during the period 1981–2000, and the landslide locations were detected by photo interpretation and the locations verified by fieldwork. Recent landslides were observed in aerial photographs from breaks in the forest canopy, bare soil, or other geomorphic characteristics typical of landslide scars, for example, head and side scarps, flow tracks, and soil and debris deposits below a scar. To assemble a database to assess the surface area and number of landslides in each of three study areas, a total of 324 landslides were mapped in a mapped area of 293 km^2.

There were ten factors that were considered for the landslide hazard analysis, and these factors were extracted from the constructed spatial database. These factors were transformed into a vector-type spatial database using the GIS. A digital elevation model (DEM) was created first from the topographic database. Contour and survey base points that had elevation values from the 1:25,000-scale topographic maps were extracted, and a DEM was constructed with a resolution of 10 meter. Using this DEM, the slope angle, slope aspect, and slope curvature were calculated. In the case of the curvature negative curvatures represent concave, zero curvature represent flat and positive curvatures represents convex. The

curvature map was prepared using the avenue routine in ArcView 3.2. In addition, the distance from drainage was calculated using the topographic database. The drainage buffer was calculated in 100 meter intervals. Using the geology database, the types of lithology was extracted, and the distance from lineament were calculated. The lithology map was obtained from a 1:63,300-scale geological map, and the distance from lineament map was calculated in 100 meter intervals. Land cover data was classified using a LANDSAT TM image employing an unsupervised classification method and was verified with field survey. The nine classes identified, such as urban, water, forest, agricultural area, tin mines, rubber and palm oil plantation were extracted for land cover mapping. Finally, the Normalized Difference Vegetation Index (NDVI) map was obtained from SPOT satellite images. The NDVI value was calculated using the formula NDVI = (IR – R) / (IR + R), where IR value is the infrared portion of the electromagnetic spectrum, and R-value is the red portion of the electromagnetic spectrum. The NDVI value denotes areas of vegetation in an image. All the above mentioned landslide inducing factors were converted to a raster grid with 10 meter × 10 meter cells for application of the artificial neural network. The area grid has 2418 rows by 1490 columns with 324 cells of landslide occurrences.

5. LANDSLIDE HAZARD ANALYSIS USING THE ARTIFICIAL NEURAL NETWORK

Before running the artificial neural network program, the training site should be selected. So, the landslide-prone (occurrence) area and the landslide-not-prone area were selected as training sites. Cells from each of the two classes were randomly selected as training cells, with 324 cells denoting areas where landslide not occurred or occurred. First, areas where the landslide was not occurred were classified as “areas not prone to landslide” and areas where landslide was known to exist were assigned to an “areas prone to landslide” training set.

The back-propagation algorithm was then applied to calculate the weights between the input layer and the hidden layer, and between the hidden layer and the output layer, by modifying the number of hidden node and the learning rate. Three-layered feed-forward network was implemented using the MATLAB software package. Here, “feed-forward” denotes that the interconnections between the layers propagate forward to the next layer. The number of hidden layers and the number of nodes in a hidden layer required for a particular classification problem are not easy to deduce. In this study, a 9 x 19 x 2 structure was selected for the network, with input data normalized in the range 0.1 - 0.9. The nominal and interval class group data were converted to continuous values ranging between 0.1 and 0.9. Therefore, the continuous values were not ordinal data, but nominal data, and the numbers denote the classification of the input data.

The learning rate was set to 0.01, and the initial weights were randomly selected to values between 0.1 and 0.3. The weights calculated from 10 test cases were compared to determine whether the variation in the final weights was dependent on the selection of the initial weights. The back-propagation algorithm was used to minimize the error between the predicted output values and the calculated output values. The algorithm propagated the error backwards, and iteratively adjusted the weights. The number of epochs was set to 2,000, and the root mean square error (RMSE) value used for the stopping criterion was set to 0.01. Most of the training data sets met the 0.01 RMSE goal. However, if the RMSE value was not achieved, then the maximum number of iterations was terminated at 2,000 epochs. When the latter case occurred, then the maximum RMSE value was 0.202. The final weights between layers acquired during training of the neural network and the contribution or importance of each of the nine factors used to predict landslide hazard are shown in Table 2.

For easy interpretation, the average values were calculated, and these values were divided by the average of the weights of the some factor that had a minimum value. The land cover value was the minimum value, 0.72, and the slope value was the maximum value, 2.05. Finally, the weights were applied to the entire study area, and the landslide hazard map was created (Figure 3). The values were classified by equal areas and grouped into four classes for visual interpretation. The possibility was classified into four classes (highest 10%, second 10%, third 20% and reminding 60%) based on area for visual and easy interpretation. The minimum value is 0.103 and maximum value is 0.910. The mean value is 0.314 and the standard deviation value is 0.1382.

6. VERIFICATION

The landslide hazard analysis result was verified using known landslide locations. Verification was performed by comparing the known landslide location data with the landslide hazard map. The rate curves were created and its areas of the under curve were calculated for all cases. The rate explains how well the model and factor predict the landslide. So, the area under curve in can assess the prediction accuracy qualitatively. To obtain the relative ranks for each prediction pattern, the calculated index values of all cells in the study area were sorted in descending order. Then the ordered cell values were divided into 100 classes, with accumulated 1% intervals. The rate verification results appear as a line in Figure 4. For example, in the case of all factor used, 90 to 100% (10%) class of the study area where the landslide hazard index had a higher rank could explain 35% of all the landslides. In addition, the 80 to 100% (20%) class of the study area where the landslide hazard index had a higher rank could explain 58% of the landslides. To compare the result quantitative, the areas under the curve were re-calculated as the total area is 1 which means perfect prediction accuracy. So, the area under a curve can be used to assess the prediction accuracy qualitatively. The area ratio was 0.8345 and we could say the prediction accuracy is 83.45%.

7. CONCLUSION AND DISCUSSIONS

Occurrence of landslide makes a significant constraint to development in Malaysia, notably through the inadvertent reactivation of ancient inland landslides. A series of Government funded research projects has provided much background information and identified suitable methods for the use of landslide hazard information in land use planning. However, a number of significant problems remain over the use of this information. In this study, a data mining approach to estimating the susceptible area of landslides using GIS and remote sensing has been presented.

From the application of artificial neural network, the relative importance, weight, between factors was calculated. The slope showed the highest value 2.05, then distance from drainage 1.4 and geology is 1.1. From the result, the slope is most

importance factor because its weight is more than two times weight than the other factors, for landslide hazard mapping. Using the weights, the landslide hazard map was created and verified. The result of verification showed 83.45% prediction accuracy. The verification result is somewhat high value.

Landslide hazard maps are of great help to planners and engineers for choosing suitable locations to implement developments. These results can be used as basic data to assist slope management and land-use planning, but the models used in the study are valid for generalized planning and assessment purposes, although they may be less useful at the site-specific scale where local geological and geographic heterogeneities may prevail. For the model to be more generally applied, more landslide data are needed, as well as application to more regions.

REFERENCE

1. Guzzetti, F., Carrarra, A., Cardinali, M. and Reichenbach, P. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study. Central Italy, *Geomorphology*, 1999, **31**, 181-216.

2. Rowbotham, D. and Dudycha, D. N., GIS modeling of slope stability in Phewa Tal watershed, Nepal, *Geomorphology*, 1998, **26**, 151-170.

3. Jibson, W. R., Edwin, L. H. and John, A. M., A method for producing digital probabilistic seismic landslide hazard maps, *Engineering Geology*, 2000, **58**, 271–289.

4. Luzi, L., Pergalani, F. and Terlien, M. T. J., Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems, *Engineering Geology*, 2000, **58**, 313–336.

5. Parise, M. and Jibson, W. R., A seismic landslide susceptibility rating of geologic units based on analysis of characteristics of landslides triggered by the 17 January, 1994 Northridge, California earthquake, *Engineering Geology*, 2000, **58**, 251–270.

6. Rautelal, P. and Lakhera, R. C., Landslide risk analysis between Giri and Tons Rivers in Himachal Himalaya (India), *International Journal of Applied Earth Observation and Geoinformation*, 2000**, 2**, 153-160.

7. Baeza, C. and Corominas, J., Assessment of shallow landslide susceptibility by means of multivariate statistical techniques, *Earth surface processes and landforms*, 2001, **26**, 251-1263.

8. Lee, S. and Min, K., Statistical analysis of landslide susceptibility at Yongin, Korea, *Environmental Geology*, 2001, **40**, 1095-1113.

9. Temesgen, B., Mohammed, M. U. and Korme, T., Natural Hazard Assessment Using GIS and Remote Sensing Methods, with Particular Reference to the Landslides in the Wondogenet Area, Ethiopia. *Phys. Chem. Earth (C)*, 2001, **26**, 665-615.

10. Clerici, A., Perego, S., Tellini, C. and Vescovi, P., A procedure for landslide susceptibility zonation by the conditional analysis method, *Geomorphology*, 2002, **48**, 349-364.

11. Donati, L. and Turrini, M. C., An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology, application to an area of the Apennines (Valnerina; Perugia, Italy), *Engineering Geology*, 2002, **63**, 277-289.

12. Lee, S., Chwae, U. and Min, K., Landslide susceptibility mapping by correlation between topography and geological structure: the Janghung area, Korea, *Geomorphology*, 2002a, **46**, 49-162.

13. Lee, S., Choi, J. and Min, K., Landslide susceptibility analysis and verification using the Bayesian probability model, *Environmental Geology*, 2002b, **43**, 120-131.

14. Zhou, C. H., Lee, C. F., Li, J. and Xu, Z. W., On the spatial relationship between landslides and causative factors on Lantau Island, Hong Kong, *Geomorphology*, 2002, **43**, 197-207.

15. Lee, S. and Choi, U., Development of GIS-based geological hazard information system and its application for landslide analysis in Korea, *Geoscience Journal*, 2003c, **7**, 243-252.

16. Lee, S., Choi, J. and Min, K., Probabilistic Landslide Hazard Mapping using GIS and Remote Sensing Data at Boun, Korea, *International Journal of Remote Sensing*, 2004b, **25**, 2037-2052.

17. Atkinson, P. M. and Massari, R., Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy, *Computer & Geosciences*, 1998, **24**, 373-385.

18. Dai, F. C., Lee, C. F., Li, J. and Xu, Z. W., Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong, *Environmental Geology*, 2001, **40**, 381-391.

19. Dai, F. C. and Lee, C. F., Landslide characteristics and slope instability modeling using GIS, Lantau Island. Hong Kong, *Geomorphology*, 2002, **42**, 213-228.

20. Ohlmacher, G.. C. and Davis, J. C., Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansa, USA, *Engineering Geology*, 2003, **69**, 331-343.

21. Remondo, J., Gonzalez, A., Diaz de Teran, J. R., Cendrero, A., Fabbri, A. G.. and Chung, F. C-J., Validation of landslide susceptibility maps: examples and applications from a case study in Northern Spain, *Natural Hazards*, 2003, **30**, 437-449.

22. Gokceoglu, C., Sonmez, H. and Ercanoglu, M., Discontinuity controlled probabilistic slope failure risk maps of the Altindag (settlement) region in Turkey, *Engineering Geology*, 2000, **55**, 277–296.

23. Romeo, R., Seismically induced landslide displacements: a predictive model, *Engineering Geology*, 2000, **58**, 337-351.

24. Carro, M., De Amicis, M., Luzi, L. and Marzorati, S., The application of predictive modeling techniques to landslides induced by earthquakes, the case study of the 26 September 1997 Umbria-Marche earthquake(Italy), *Engineering Geology*, 2003, **69**, 139-159.

25. Shou, K. J. and Wang, C. F., Analysis of the Chiufengershan landslide triggered by the 1999 Chi-Chi earthquake in Taiwan, *Engineering Geology*, 2003, **68**, 237-250.

26. Zhou, G.., Esaki, T., Mitani, Y., Xie, M. and Mori, J., Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach, *Engineering Geology*, 2003, **68**, 373-386.

27. Ercanoglu, M. and Gokceoglu, C., Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkye) by fuzzy approach, *Environmental Geology*, 2002, **41**, 720-730.

28. Pistocchi, A., Luzi, L. and Napolitano, P., The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods, *Environmental Geology*, 2002, **41**, 765–775.

29. Lee, S., Ryu, J. H., Min, K. and Won, J. S., Landslide susceptibility analysis using GIS and artificial neural network, *Earth Surface Processes and Landforms*, 2003a, **27**, 1361-1376.

30. Lee, S., Ryu, J. H., Lee, M. J. and Won, J. S., Landslide susceptibility analysis using artificial neural network at Boun, Korea, *Environmental Geology*, 2003b, **44**, 820-833.

31. Lee, S., Ryu, J. H., Won, J. S. and Park, H. J., Determination and application of the weights for landslide susceptibility mapping using an artificial neural network, *Engineering Geology*, 2004a, **71**, 289-302.

32. Garrett, J., Where & why artificial neural networks are applicable in civil engineering, *Journal of Computing Civil Engineering*, 1994, **8**, 129-130.

33. Paola, J. D. and Schwengerdt, R. A., A review of backpropagation neural networks for classification of remotely sensed multi-spectral imagery, *International Journal of Remote Sensing*, 1995, **16**, 3033-3058.

34. Atkinson, P. M. and Tatnall, A. R. L., Neural networks in remote sensing. *International Journal of Remote Sensing*, 1997, **18**, 699-709.

35. Zhou, W., Verification of the nonparametric characteristics of back propagation neural networks for image classification, *IEEE Transaction Geosciences Remote Sensing*, 1999, **38**, 771–779.

Classification	Sub-Classification	GIS Data Type	Scale
Geological Hazard	Landslide	Point coverage	1:50,000
Basic Map	Topographic Map	Line and Point coverage	1:50,000
	Geological Map	Polygon coverage	1:63,300
	Land Cover	GRID	30meter×30meter
	Vegetation Index (NDVI)	GRID	10meter×10meter

Table 1. Data layer of study area

Factor	Weight	Normalized Weight
Slope (unit: degree)	0.205	1
Aspect	0.079	0.047
Curvature (unit: unit less)	0.085	0.099
Distance from drainage (unit: meter)	0.141	0.514
Geology	0.117	0.334
Distance from lineament (unit: meter)	0.093	0.155
Soil	0.105	0.246
Land cover	0.072	0.01
NDVI	0.105	0.236

Table 2. Weights of each factor estimated by neural network considered in this study.

Legend of Figures

Figure 1 Landslide location map with hill shaded map of study area
Figure 2 Architecture of neural network
Figure 3 Landslide hazard maps based on Landslide hazard map based on Artificial Neural Network model
Figure 4 Cumulative frequency diagram showing landslide hazard index rank occurring in cumulative percent of landslide occurrence

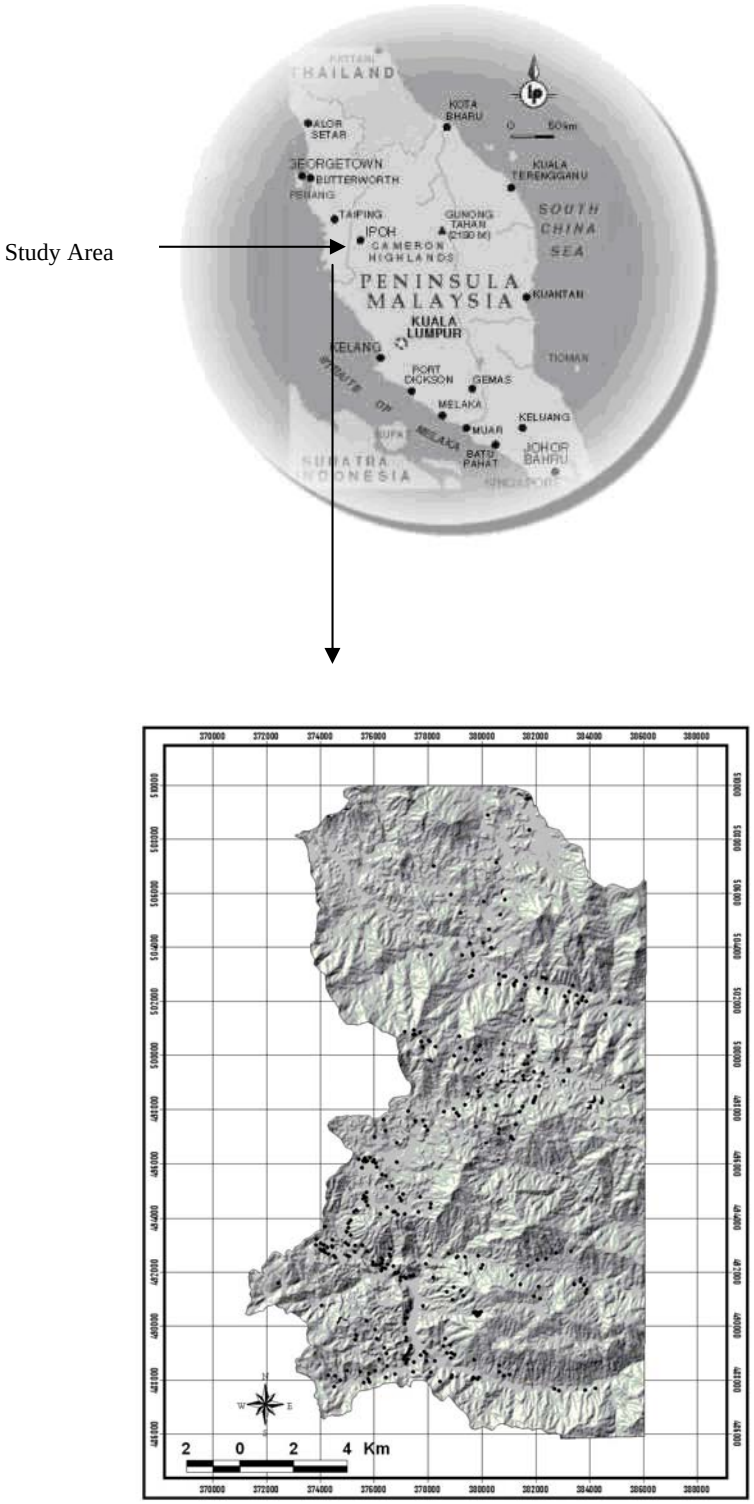

Figure 1. Landslide location map with hillshaded map of study area.

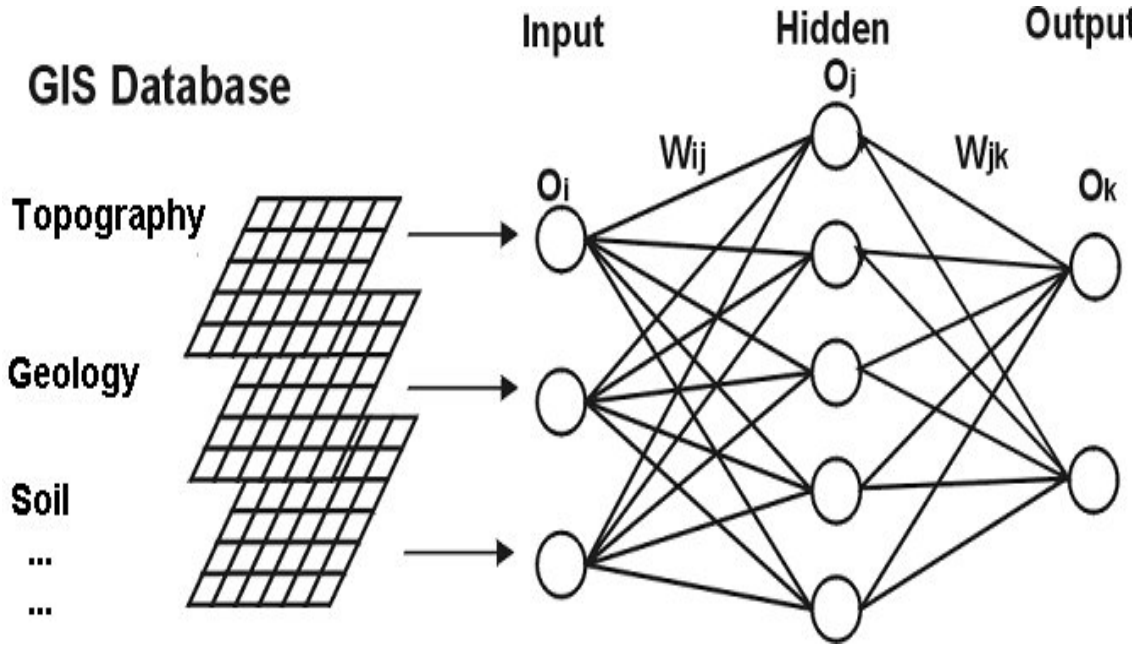

Figure 2. Architecture of neural network.

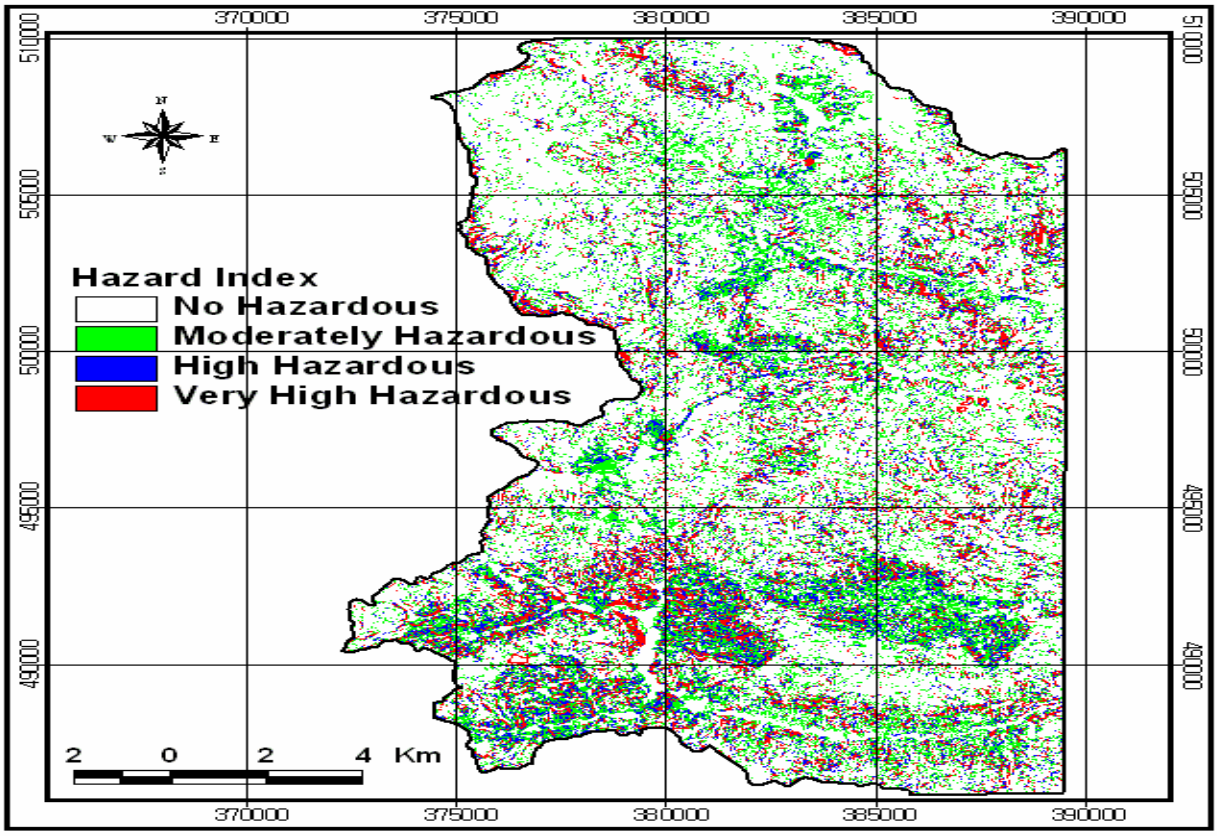

Figure 3. Landslide hazard map based on Artificial Neural Network model

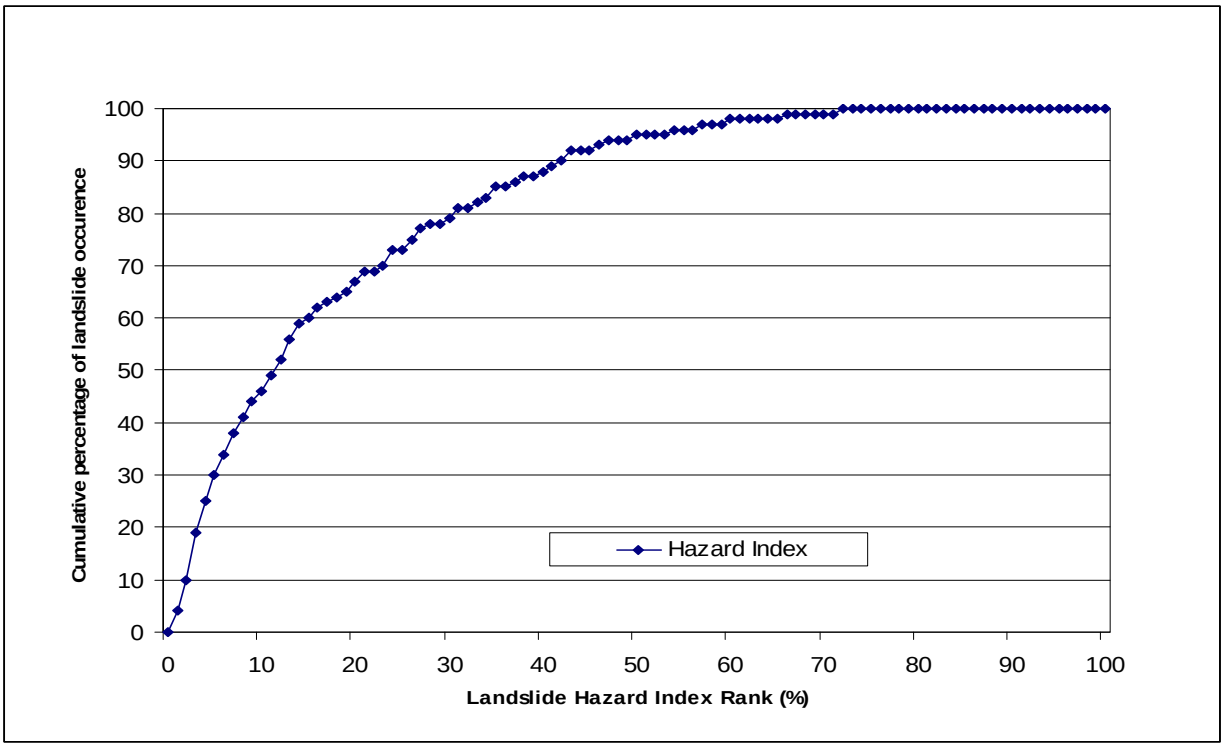

Figure 4. Cumulative frequency diagram showing landslide hazard index rank occurring in cumulative percent of landslide occurrence.