International Journal of Physical Sciences Vol. 4 (1), pp. 001-015, January, 2009
Available online at http://www.academicjournals.org/IJPS
ISSN 1992 - 1950 © 2009 Academic Journals

Full Length Research Paper

Landslide risk analysis using artificial neural network model focussing on different training sites

Biswajeet Pradhan[1*] and Saro Lee[2]

[1]Institute for Advanced Technologies (ITMA), University Putra Malaysia, 43400, UPM, Serdang, Malaysia.
[2]Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources (KIGAM) 92 Gwahang-no, Yuseong-gu, Daejeon, Korea.

Accepted 28 November, 2008

This paper presents landslide hazard and risk analysis using remote sensing data, GIS tools and artificial neural network model. Landslide locations were identified in the study area from interpretation of aerial photographs and from field surveys. Topographical and geological data and satellite images were collected, processed, and constructed into a spatial database using GIS and image processing. These factors were used with artificial neural network to analyze landslide hazard. 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 the landslide hazard map was created using GIS tools. Landslide locations were used to verify results of the landslide hazard maps and to compare them. The results of the analysis were verified using the landslide location data and compared with neural network model with all cases. The accuracy observed was 83, 72, 82, 79 and 81% for training sites 1, 2, 3, 4 and 5 respectively. GIS data was used to efficiently analyze the large volume of data, and the artificial neural network proved to be an effective tool for landslide hazard analysis. Further, risk analysis has been performed using DEM, distance from hazard zone, land cover map and damageable objects at risk. DEM was used to delineate the catchments and served as a mask to extract the highest hazard zones of the landslide area. Risk map was produced using map overlying techniques in GIS. This information could be used to estimate the risk to population, property and existing infrastructure like transportation network.

Key words: Landslide, risk, artificial neural network, GIS.

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. To assist the implementation of this policy, the Malaysian Centre for Remote Sensing (MACRES) commissioned one major demonstration projects to assess the potential for landslide and to incorporate this information in the strategic planning 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, we 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

*Corresponding author. E-mail: biswajeet@lycos.com. Tel: 603-8946-7543 Fax: 603-8656-6061.

Figure 1. Location map of the study area.

also assessed. Further, landslide risk analysis was performed using the landslide hazard map and other associated factors.

Risk analysis is a valid technique if and only if it fulfills a series of appropriate criteria. It should state the probability as well as the expected impact, and the latter should be expressed in relation to the size of the population at risk. Varne (1984); Fell (1994); Whitsman (1984); and Christian et al., (1992) have performed stud-ies related to risk analysis. There have been many studies carried out on landslide hazard evaluation using GIS; for example, Guzzetti et al. (1999) 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 models (Akgun and Bulut, 2007; Dahal et al., 2007; Clerici et al., 2006; Cevic and Topal, 2003; Rowbotham and Dudycha, 1998; Jibson et al., 2000; Luzi et al., 2000; Parise and Jibson, 2000; Baeza and Corominas, 2001; Lee and Min, 2001; Temesgen et al., 2001; Clerici et al., 2002; Donati and Turrini, 2002; Lee et al., 2002a; Lee et al., 2002b; Zhou et al., 2002; Lee and Choi, 2003c; Lee et al., 2004b; Lee and Talib, 2005a; Lee and Dan, 2005b; Lee and Lee, 2006a; Pradhan and Lee, 2007; Pradhan et al., 2005). One of the statistical models available, the logistic regression models, has also been applied to landslide hazard mapping (Akgul and Bulut, 2007; Tunusluoglu et al., 2007; Lamelas et al., 2007; Wang and Sasa, 2005; Su¨zen and Doyuran, 2004; Atkinson and Massari, 1998; Dai et al., 2001; Dai and Lee, 2002; Ohlmacher and Davis, 2003; Lee, 2005c; Lee and Sambath, 2006b), as has the geotechnical model and the safety factor model (Gokceoglu et al., 2000; Romeo, 2000; Refice and Capolongo, 2002; Carro et al., 2003; Shou and Wang, 2003; Zhou et al., 2003, Lee, 2007a; Lee and Biwajeet, 2007b). As a new approach to landslide hazard evaluation using GIS, fuzzy logic, and artificial neural network models have been applied (Xie et al., 2004; Ercanoglu and Gokceoglu, 2002; Pistocchi et al., 2002; Lee et al., 2003a; Lee et al., 2003b; Lee et al., 2004a; Tangestani, 2004; Lee et al., 2006c, Lee, 2007).

Landslide occurrence areas were detected in the Penang 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 extracted factors were converted to a 10 × 10 m grid (ARC/INFO GRID type). Artificial neural network was applied using the database and landslide hazard map was created. Further, landslide risk map was computed using DEM, distance from hazard zone and landcover map. Finally, the map was verified and compared using known landslide locations for quantitative verification.

In the study, Geographic Information System (GIS) software, ArcView 3.2, and ArcGIS 9.0 version software packages were used as the basic analysis tools for spatial management and data manipulation.

Study area

Penang Island, which has suffered much landslide damages following heavy rains, was selected as a suitable pilot area to evaluate landslide risk analysis (Figure 1). Penang is one of the 13 states of the Federal territory of Malaysia and is located on the North West coast of the

Table 1. Data layer of study area.

Classification		GIS Data Type		Scale or Resolution	
Spatial Database	Factor	Spatial Database	Factor	Spatial Database	Factor
Landslide	Landslide	ARC/INFO Polygon coverage	ARC/INFO GRID	1:25,000	10m×10m
Topographic Map	Slope	ARC/INFO Line and Point Coverage		1:25,000	
	Aspect				
	Curvature				
Drainage Map	Distance from drainage	ARC/INFO Line Coverage			
Soil Map	Types	ARC/INFO Polygon coverage		1:100,000	
Geology Map	Litho types	ARC/INFO Polygon, Line coverage		1:63,300	
	Distance from lineaments				
Land Cover	Land Cover	ARC/INFO GRID		30 m × 30 m	
NDVI	NDVI	ARC/INFO GRID		10 m x 10 m	
Precipitation	Precipitation	GRID		10 m x 10 m	

Malaysia peninsula. It is bounded to north and east by the state of Kedah, to south by the state of Perak, and to west by the Straits of Malacca and Sumatra (Indonesia). Penang consists of the island of Penang, and a coastal strip on the mainland, known as Province Wellesley. The island covers an area of 285 km^2, and is separated from the mainland by a channel. The study area is located approximately between latitudes 5º15' N to 5º30' N and longitudes 100º10'E to 100º20'E. The landuse in the study area is mainly peat swamp forest, plantation forest, inland forest, scrub, grassland and ex-mining area. The slope of the area ranges from 25º to as much as 87º. The relief of the study area varies between 0 - 420 m above mean sea level. Based on Malaysian Meteorological Department, the temperature of northern part of Penang ranges between 29 to 32ºC and mean relative humidity varies between 65 to 70%. The highest temperature is during April to June while the relative humidity is lowest in June, July and September. The rainfall of about 58.6 to 240 mm per month is recorded in the study area (at the Bayan Lepas weather station provided by the Malaysian Meteorological Services Department). GPS data for landslide locations has been collected for various part of Penang island in the month of July and October 2006. There were about 21 active landslides have been recorded from the field survey and being used to verify the model output. Few of these landslides photos collected during the field survey have been shown in Figure 2. SPOT 5 satellite image of the study area has been shown in Figure 3.

Accurate detection of the landslide location is very important for probabilistic landslide hazard and risk analysis. Remote sensing methods, using aerial photographs and satellite images are employed to obtain significant and cost-effective information on landslides. In this study, 1:10,000 – 1:50,000-scale aerial photographs were used to detect the landslide locations. These photographs were taken during 1981 – 2000. Landslide locations were detected by aerial photo interpretation and further verified by fieldwork. These landslides can be seen in aerial photographs by interpreting breaks in the forest canopy, bare soil, and other typical geomorphic characteristics of landslide scars. A total of 463 landslides were mapped within 285 km^2 to assemble a database to assess the surface area and number of landslides in the study area.

Identification and mapping of a suitable set of instability factors having a relationship with the slope failures require priori knowledge of the main causes of landslides (Guzzetti et al., 1999). These instability factors include surface and bedrock lithology and structure, seismicity, slope steepness and morphology, stream evolution, groundwater conditions, climate, vegetation cover, landuse, and human activity. The availability of thematic data varies widely, depending on the type, scale, and method of data acquisition. To apply the probabilistic model, a spatial database that considers landslide-related factors was designed and constructed. These data are available in Malaysia either as paper or as digital maps. The spatial database constructed is listed in Table 1.

There were ten landslide inducing factors considered in calculating the probability. These factors were transformed into a vector-type spatial database using the GIS. For the DEM creation, 10 meter interval contours and survey base points showing the elevation values were extracted from the 1:25,000-scale topographic maps. Using this DEM, slope angle, slope aspect, and slope curvature were calculated. In addition, the distance from drainage was calculated using the topographic database.

Figure 2. Field photographs illustrating the characteristic and types of landslides in Penang Island.

The drainage buffer was calculated at 100 m intervals and classified into 10 equal area classes. The lithology map is prepared from a 1:63,300-scale geological map and the distance from lineament is calculated based on the Euclidean distance method in ArcView. The lineament buffer was calculated in 100 meter intervals and classified into 10 equal area classes. Landuse map was prepared using Landsat TM image (30 m spatial resolution) using unsupervised classification method and field survey. There were 11 landuse classes identified, such as urban, water, forest, agriculture, and barren area. Finally, the Normalized Difference Vegetation Index (NDVI) map was generated from SPOT 5 (2.5 m spatial resolution) satellite images. The NDVI value was calculated using the formula NDVI = (IR – R) / (IR + R), where IR is the energy reflected in the infrared portion of the electromagnetic spectrum, and R is the energy reflected in the red portion of the electromagnetic spectrum. The NDVI is useful in delineating vegetation.

Artificial neural network

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” (Garrett, 1994). 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 (Paola and Schwengerdt, 1995).

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 refer-red to as the network architecture (Figure 3). 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).

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

WHERE W_{IJ} REPRESENTS THE WEIGHTS BETWEEN NODES *I* AND *J*, AND O_I IS THE OUTPUT FROM NODE *J*, GIVEN BY

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

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:

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

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 (Atkinson and Tatnall, 1997).The error, *E*, for an input training pattern, *t*, is a function of the desired output.

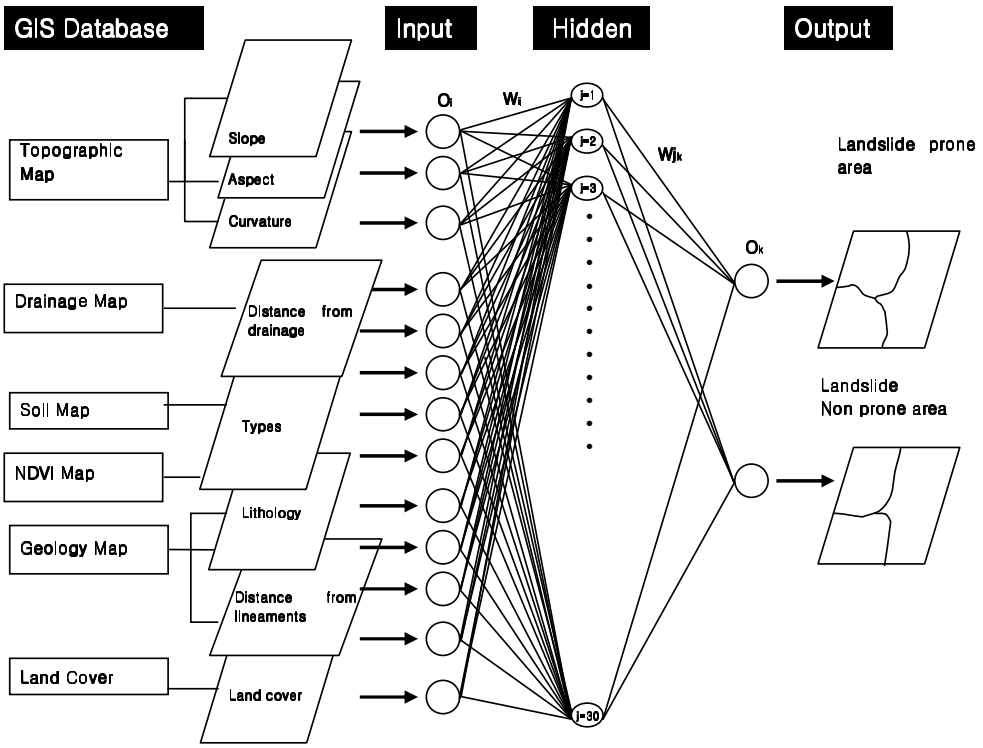

Figure 3. Architecture of neural network.

$$E = \frac{1}{2}\sum_{k}(d_k - o_k). \quad (4)$$

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

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

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,

$$\delta_j = (\sum \delta_k w_{jk}) f'(net_j) \quad (6)$$

and for output layers,

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

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 (1999) 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;

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

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):

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

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.

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

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

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:

$$\sum_{j=1}^{J} t_{jk} = J \quad (12)$$

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

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

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

$$s_{ij} = \frac{\left| \omega_{ij} \right|}{\frac{1}{I} \cdot \sum_{i=1}^{I} \left| \omega_{ij} \right|} = \frac{I \cdot \left| \omega_{ij} \right|}{\sum_{i=1}^{I} \left| \omega_{ij} \right|}. \quad (14)$$

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

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

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

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

Spatial data base using GIS and remote sensing

The data used for the analysis is shown in Table 1. 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 463 landslides were mapped in a mapped area of 285 km^2. The list of all the data layers has been illustrated in Figure 3.

There were ten factors considered for the analysis, and the factors were extracted from the constructed spatial database. The factors were transformed into a vector-type spatial database using the GIS, and landslide-related factors were extracted using the database. 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 m intervals. Using the geology database, the 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 m intervals. Land cover data was classified using a LANDSAT TM image employing an unsupervised classification method and field survey. The 9 classes identified, such as urban, water, forest, agricultural area, tin mines, rubber and palm oil plantation were extracted for land cover mapping. A soil map was prepared based on soil texture map retrieved from Malaysian agricultural department. The source data was in a hard copy format that was converted to digital data for further processing in GIS. Finally, the Normalized Difference Vegetation Index (NDVI) map was obtained from SPOT satellite images. The NDVI va-

Table 2. Average weights of each factor

Factors/Methods	Average values for (Case 1, Case 2, Case 3, Case 4 and Case 5)
Slope	0.930
Aspect	0.382
Curvature	0.186
Distance from Drainage	0.071
Distance from lineament	0.274
Geology	0.489
Land cover	0.326
NDVI	0.214
Soil	0.266
Precipitation	0.147

lue 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. Historical rainfall data was collected from the rain gauge station in the study area. The data collected over last 20 years were interpolated using IDW method.

Landslide hazard and risk 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 463 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 as-signed 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 gro--up 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 0.051. The final weights between layers acquired during training of the neural network and the contribution or importance of each of the 9 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 distance from drainage value was the minimum value, 0.071, and the slope value was the maximum value, 0.930. Finally, the weights were applied to the entire study area, and the landslide hazard map was created (Figure 5). 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.

Risk maps demarcate the areas under potential consequences where consequences can be those affecting human life, having economic effects or causing environmental changes for instance. A particular surface area subject to the same hazard can face a variety of consequences, depending on land use. Landside risk map was prepared through the overly of land use maps on hazard map explicitly with a map reflecting property values. It is not only possible that the same danger (hazard) produces different risks dependent on the land use but the reverse, namely, different dangers (hazards) affecting the same area can happen. Risk map will reflect not only landslide hazards but also avalanches affecting the same area. In landslide hazard map a particular surface area subjected to the same hazard can face a variety of consequences, depending on landuse. The flow chart for the risk analysis is shown in Figure 4. Three main input parameters taken for the risk model calculation are DEM, hazard zones map and land use map. DEM was used to delineate the catchments and served as a mask to extract the highest hazard zones of the landslide area. Three types of land use are selected, i.e. agricultural, urban and settlement, were overlaid with the extracted hazard zone to produce the different categories of risk area. Risk map

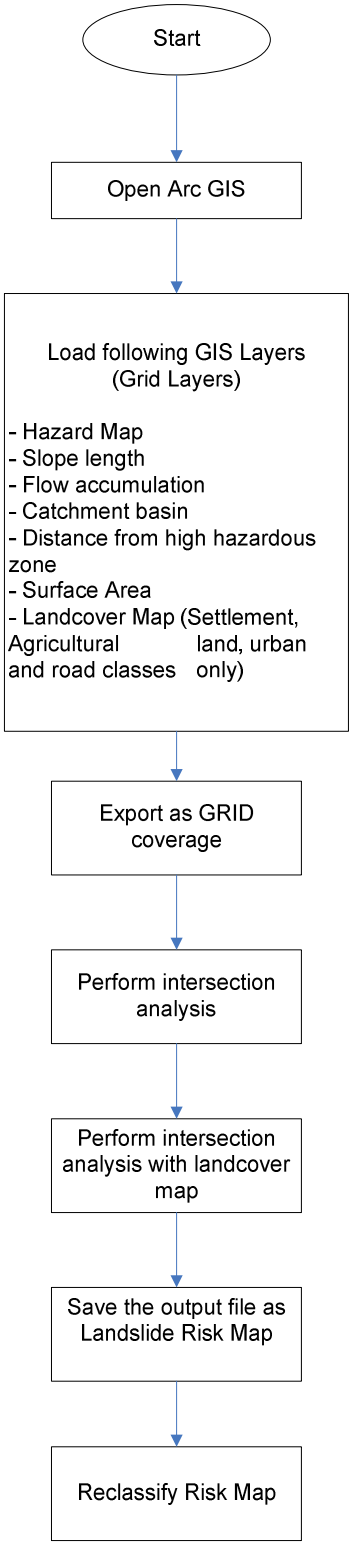

Figure 4. Flow chart of the risk analysis.

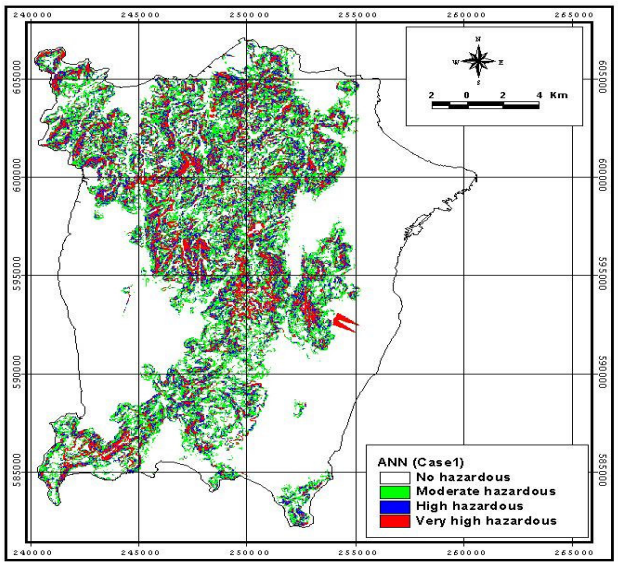

Figure 5 (a) - (e): Landslide hazard mapping using artificial neural network. **(a) Case 1:** Use of landslide location as prone training site and slope is 0 as non-prone training site.

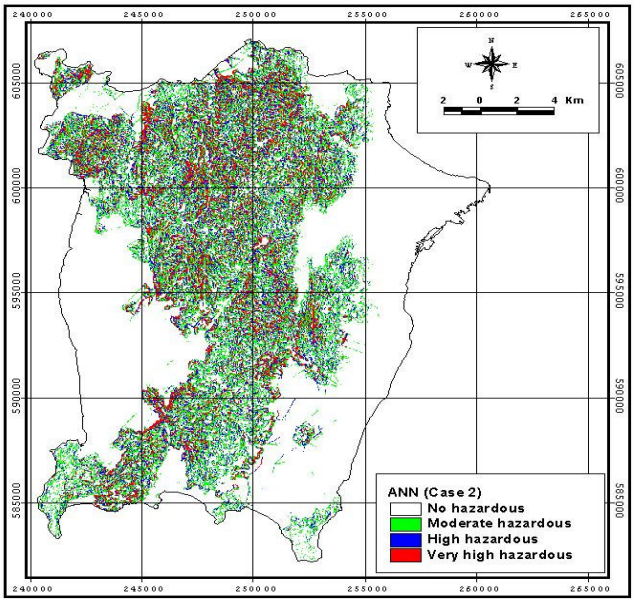

Figure 5 (a) - (e). Landslide hazard mapping using artificial neural network. **(b) Case 2:** Use of landslide location as prone training site and result from likelihood ratio as non- prone training site.

have been classified qualitatively into 4 different zones with respect to the percentage of areas such as follows: No risk area - 60%; Moderate risk area - 20%; High risk area - 10% and Very high risk area - 10%. Figure 6 shows the landslide risk map of the study area. It has been observed that, many settlements have been built up on high landslide risk areas where the probability of occurrence of landslide is very high. Those high risk areas need to be brought to the notice of the public so that people can realize the possibility of future landslides.

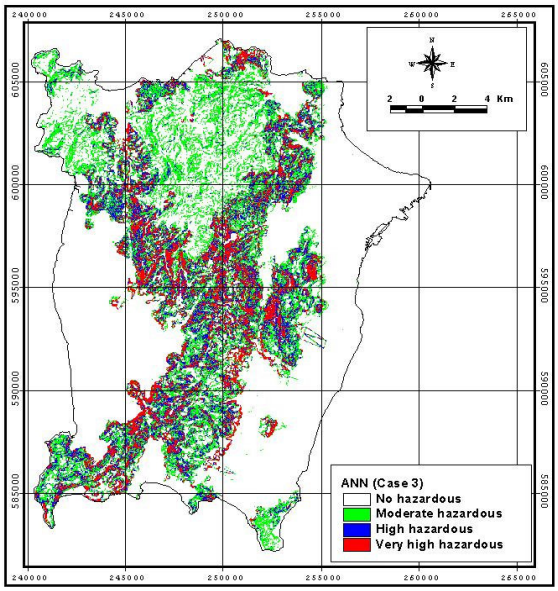

Figure 5 (a) - (e). Landslide hazard mapping using artificial neural network. **(c) Case 3:** Use of landslide location as prone training site and result from logistic regression as non- prone training site.

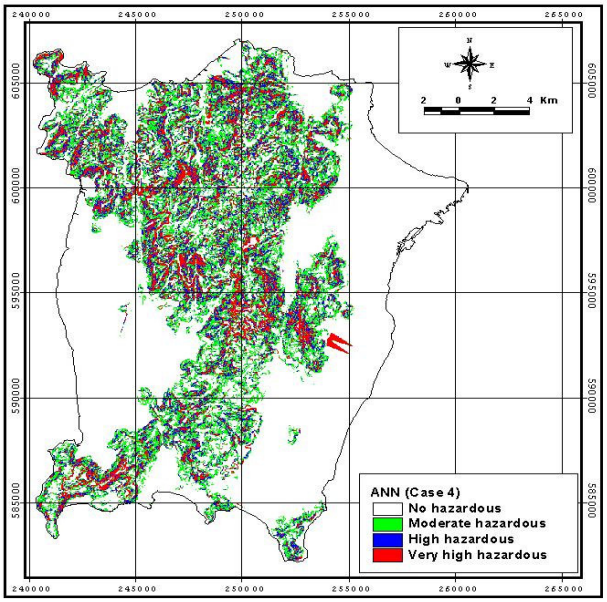

Figure 5 (a) - (e). Landslide hazard mapping using artificial neural network. **(d) Case 4:** Use of result from likelihood ratio as prone training site and result from likelihood ratio as non-prone training site.

This could save their property and life.

Verification

The landslide hazard analysis result was verified using

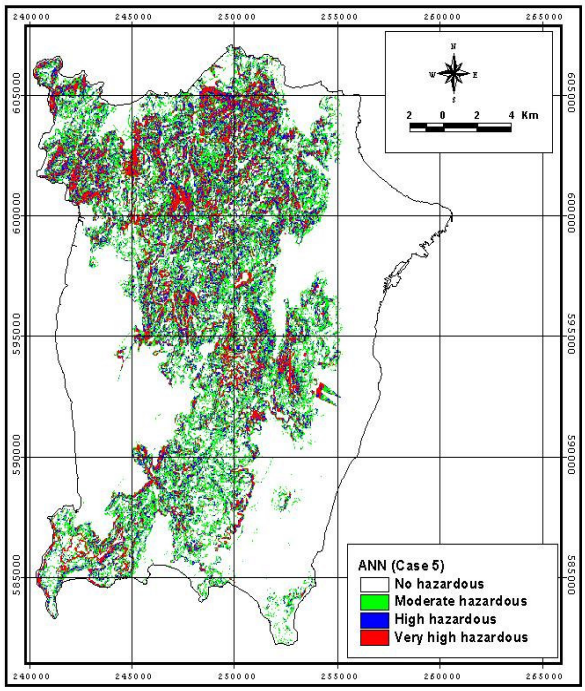

Figure 5 (a) - (e). Landslide hazard mapping using artificial neural network. **(e) Case 5:** Use of result from logistic regression as prone training site of result from logistic regression as non-prone training site.

known landslide locations. Global Positioning System data for landslide locations has been collected for various part of Penang island in the month of July and October 2005. 21 active landslides have been recorded and used to verify the model output. The result shows that there is a fair agreement between the prediction accuracy and the occurrence of landslide in a particular area. 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 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 7. 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 accu-

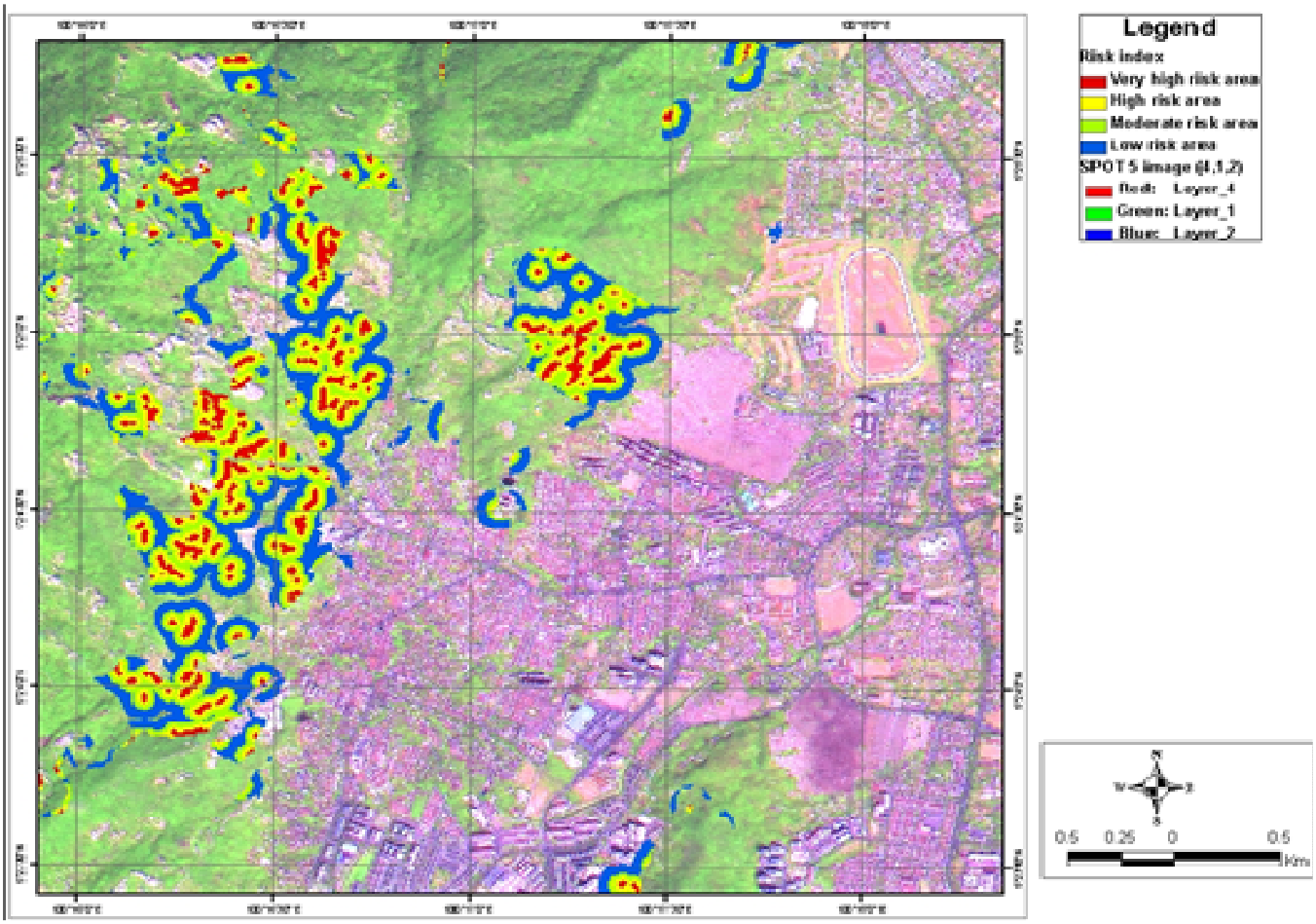

Figure 6. Landslide risk map for the study area. **(a)** Landslide risk map for the north eastern part of the study area.

racy qualitatively.

Verification results show that in the training site 1 (case 1) where slope equal to “zero” used for hazard map, the area ratio was 0.8345 and the prediction accuracy was 83.45%. In the training site 2 (case 2) for high frequency ratio values, the area ratio was 0.7271 and the prediction accuracy was 72.71. In the training site 3 (case 3) for low frequency ratio values, the area ratio was 0.8200 and the prediction accuracy was 82.00%. In the training site 4 (case 4) for high logistic regression values, the area ratio was 0.7904 and the prediction accuracy was 79.04%. In the training site 5 (case 5) for low logistic regression values, the area ratio was 0.8101 and the prediction accuracy was 81.01%. So from the prediction accuracy graphs (Figure 7), it is quite evident that, training site 1 shows the best prediction accuracy of 83.45%, where as training site 2 shows the least prediction accuracy of 72.71% with difference is about 10%. Therefore, selecting training site is very important.

In the analysis, weight has an impact on the accuracy of the landslide hazard calculation, as the results with weighting were more suitable for landslides hazard mapping, and the weights distinguished the LHIs more widely.

DISCUSSIONS AND CONCLUSION

Landslides present a significant constraint to develop ment 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 is presented. The relationship between landslide occurrence and slope (Table 2) shows that steeper slopes have greater landslide probabilities. Below a slope of 15°, the frequency ratio was 0.20, which indicates a very low probability of landslide occurrence. For slopes above 16°, the ratio was >1, which indicates a high probability of landslide occurrence. As the slope angle increases, then the shear stress in the soil or other unconsolidated material generally increases. Gentle slopes are expected to have a low frequency of landslides because of the generally lower shear stresses associated with low gradients. Steep natural slopes re-

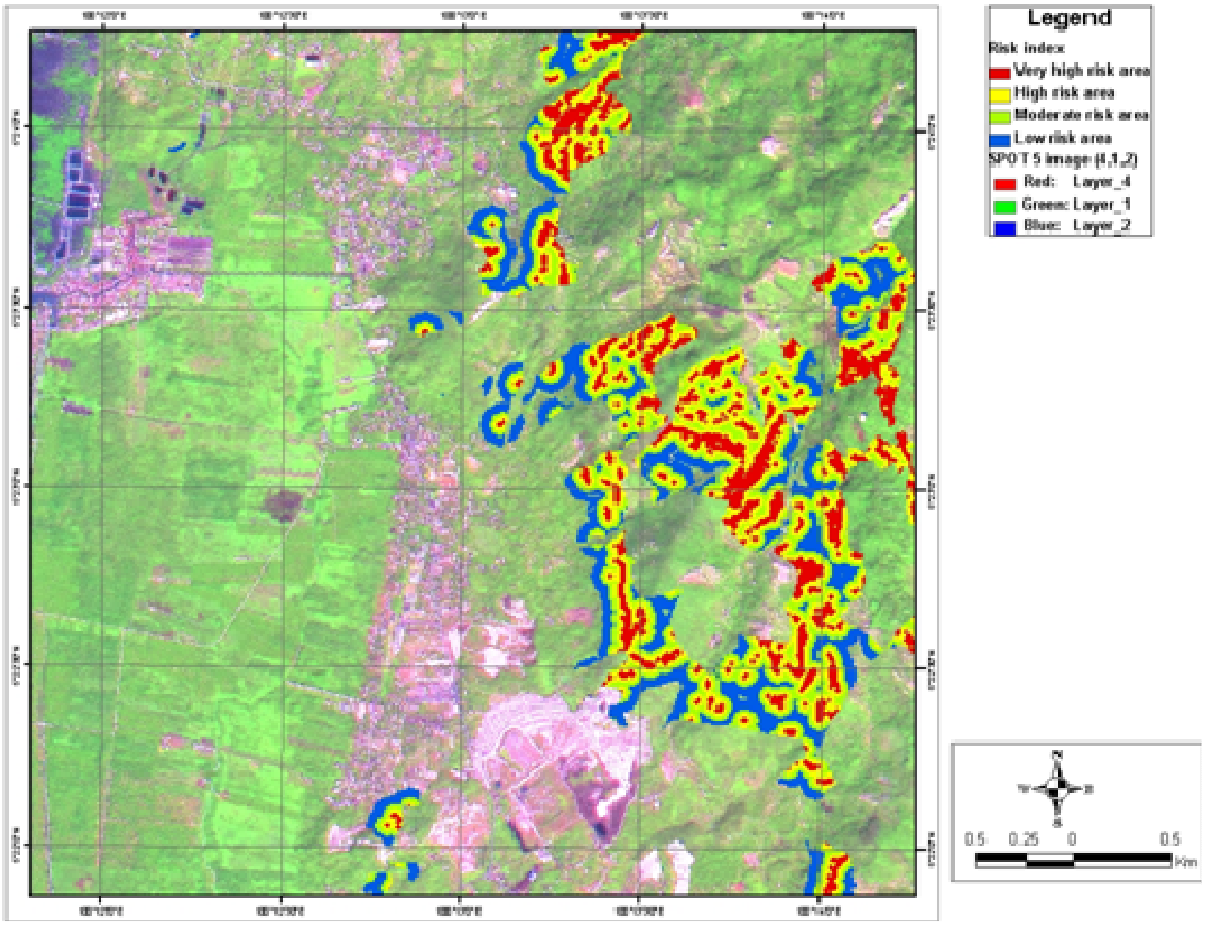

Figure 6. Landslide risk map for the study area. **(b)** Landslide risk map for the north western part of the study area.

sultting from outcropping bedrock, however, may not be susceptible to shallow landslides. In the case of the aspect (Table 2), landslides were most abundant on south-facing and northeast-facing slopes. The frequency of landslides was lowest on east-facing, west-facing, and northwest-facing slopes, except in flat areas. The curvature values represent the morphology of the topography. A positive curvature indicates that the surface is upwardly convex at that pixel. A negative curvature indicates that the surface is upwardly concave at that pixel. A value of zero indicates that the surface is flat. As shown in Table 2, the more positive or negative the curvature value the higher the probability of landslide occurrence. Flat areas had a low frequency ratio of 0.20. Concave areas had a frequency ratio of 0.41. The reason for this is that following heavy rainfall; a concave slope contains more water and retains this water for a longer period which could lead to failure of slope triggering landslide. Convex areas had a frequency ratio of 3.07. The reason for this is that a convex rounded hilltop slope could expose to repeated dilation and contraction of loose debris on an inclined surface that might induce a creeping or mudslide due to heavy rainfall. Analysis was carried out to assess the influence of drainage lines on landslide occurrence. For this purpose, the proximity of landslide to drainage line was identified by buffering (Table 2). It can be seen that as the distance from a drainage line increases, the landslide frequency generally decreases. At a distance of <250 m, the ratio was >1, indicating a high probability of landslide occurrence, and at distances >251 m, the ratio was <1, indicating very less probability. This can be attributed to the fact that terrain modification caused by gully erosion may influence the initiation of landslides. However, at a distance of <50m, the frequency ratio is 0.81 which is due to the less number of previously occurred landslides.

For geological factors (Table 2), it was found that in the case of the lithology, the frequency ratio was higher (1.3) in igneous rock areas, and was lower (0.02) in alluvium areas. In the case of the distance from lineament, the closer the distance was to lineament, the greater was the landslide-occurrence probability. For distances of <1000 m, the ratio was >1, indicating a high probability of landslide occurrence, and for distances of >1000 m, the ratio was <1, indicating a low probability. As the distance from lineament decreases, the fracture of the rock increases,

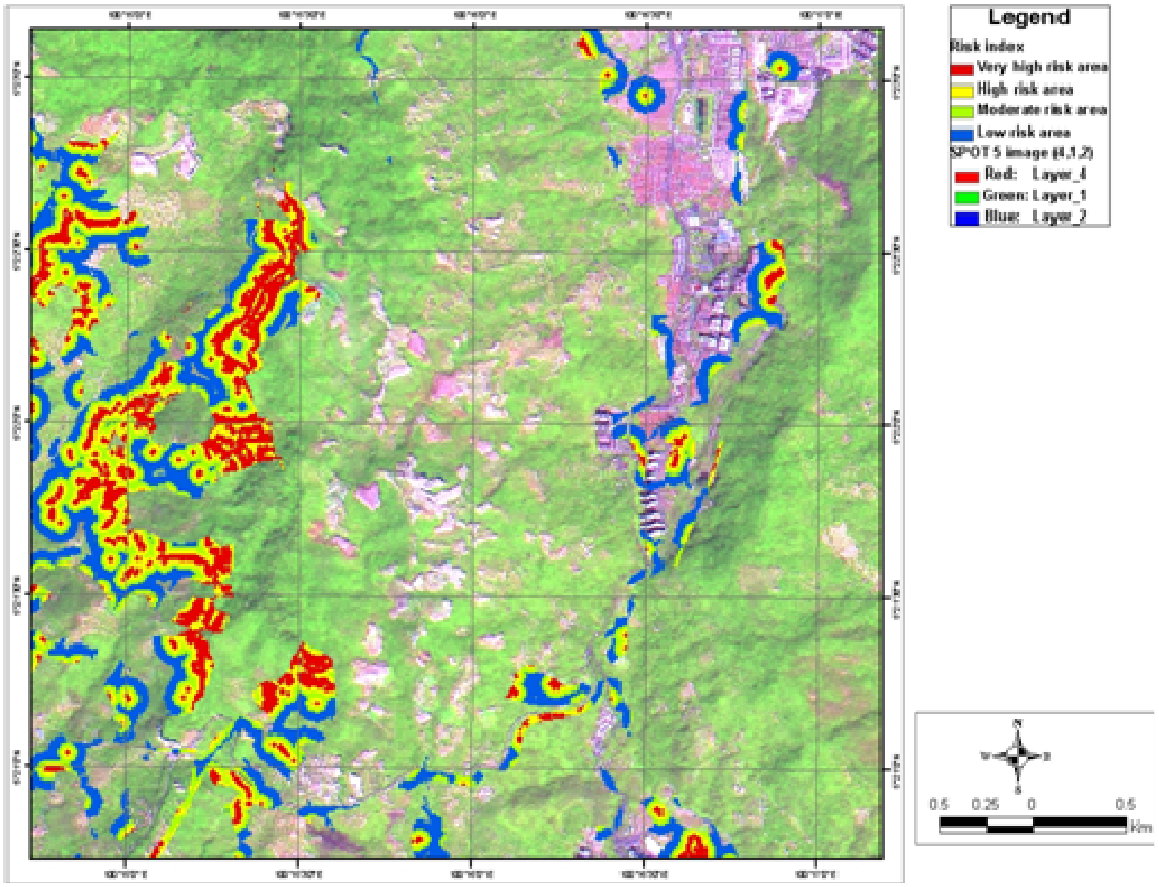

Figure 6. Landslide risk map for the study area. **(c)** Landslide risk map for the southern part of the study area.

and the degree of weathering increases resulting in greater chances of landslides. In the case of landuse (Table 2), the landslide-occurrence values were higher in tin mine and cultivated land areas, and lower in grass, coconut and oil palm plantation areas. In the case of the vegetation index (Table 2), for NDVI values below 37, the frequency ratio was <1, which indicates a low landslide-occurrence probability, and for NDVI values above 37, the frequency ratio was >1, indicating a high landslide-occurrence probability. However, for NDVI value between 41 and 43 the frequency ratio was 0.96 which is almost equal to 1. This result means that the landslide probability increases with the vegetation index value. This could be due to more vegetation is seen along structurally weaker zones. For soil class (Table 2), it was found that the frequency ratio was higher for Rengambukit temiang association (2.07) and steep land (1.41), and was lower (0.0 – 0.22) in other series. This result indicates that the landslide probability increases with the steep land. In the case of rainfall precipitation (Table 2), for precipitation amount below 2707mm, the frequency ratio was <1, which indicates a low landslide-occurrence probability, and for precipitation amount above 2707mm, the frequency ratio was >1, indicating a high landslide-occurrence probability. However, for precipitation value between 2719 and 2742 mm, the frequency ratio was 0.96 and 0.89 which is almost equal to 1. This could be due to the prolong spell of rainfall during monsoon season that could led to the high landslide-occurrence probability.

An artificial neural network approach has been used to estimate areas susceptible to landslides using a spatial database for a Penang Island. Five estimation methods were used for comparison purposes. The results using result from logistic regression as prone training site and result from logistic regression as non-prone training site (Case 5) and result from likelihood ratio as prone training site and result from likelihood ratio as non-prone training site (Case 4) were better than the other three estimation cases, with the results using of landslide location as prone training site and result from likelihood ratio as non-prone training site being the worst.. From the application of artificial neural network, the relative importance, weight, between factors was calculated. From the result, the slope is most importance factor which is more than 2 time among the other factors, for landslide hazard mapping. The results of average normalized value for all the five cases of training sites shows slope has the highest weight values of 0.930. Geology plays an important role

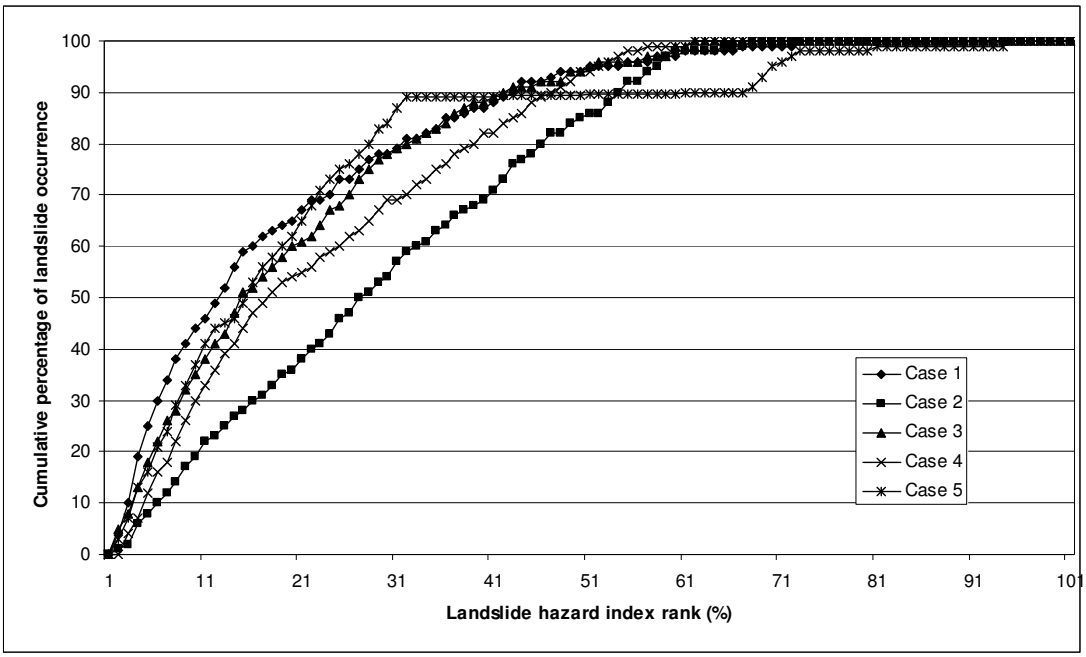

Figure 7. Success rate curve of each case.

giving a weight value of 0.489; then aspect 0.382. Distance from drainage has the lowest weight age value 0.071.

The back-propagation training algorithm presents difficulties when trying to follow the internal processes of the procedure. The method also involves a long execution time, has a heavy computing load, and there is the need to convert the database to another format. However, landslide susceptibility can be analyzed qualitatively. In addition to using a multi-faceted approach to a solution, they enable the extraction of reliable results for a complex problem, and for continuous and discrete data processing.

Risk analysis was performed for the study area. The landslide hazard map was overlaid on settlement map to produce the landslide risk map. These results can be used as basic data to assist slope management and land-use planning. The methods used in the study are also valid for generalized planning and assessment purposes, although they may be less useful on 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.

Decision making under uncertainty is closely related to risk analysis. Landslide risk map will help for decision making for planners. These decisions are usually in the form of technical countermeasures, regulatory management or combinations of the two. Classic examples of regulatory management are zoning maps which, for instance, exclude some areas from habitation. Regulatory management is often quite intricate in prescribing different permit procedures which may include detailed evaluations and additional exploration or even go so far to prescribing particular slope designs (slope grades e.g.). The latter is actually a combination of regulatory and technical management. Technical mitigating measures range from a variety of stabilizing measures to protective measures such as rock fall galleries to warning devices. One of the most important steps of developing a hazard mitigation plan is assessing risks, or estimating potential losses to the people and properties within the landslide prone area.

ACKNOWLEDGMENT

Authors would like to thank Jasmi Abdul Talib from Malaysian Centre for Remote Sensing (MACRES) for providing landslide location data used in the analysis. This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Ministry of Science and Technology of Korea. Thanks are due to Alexander Von Humboldt Foundation for awarding AvH fellowship for awarding guest scientist position at Dresden University of Technology, Germany.

REFERENCE

Akgun A, Dag S, Bulut F (2007). Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. Environmental Geology Online first.

Akgun A, Bulut F (2007). GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. Environmental Geology Online first.

Atkinson PM, Massari R (1998). Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Comp. Geosci. 24 (4): 373-385.
Baeza C, Corominas J (2001). Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth surface processes and landforms 26: 1251-1263.
Carro M, De Amicis M, Luzi L, Marzorati S (2003). The application of predictive modeling techniques to landslides induced by earthquakes: the case study of the 26 September, 1997. Umbria-Marche earthquake(Italy). Engr. Geol. 69: 139-159.
Cevic E, Topal T (2003). GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). Environ. Geol. 44: 949–962.
Christian JT, Ladd CC, Baecher GB (1992). Reliability and probability in stability. In: Stability and Performance of Slope and Embankments - II; Geotechnical Special Publication No. 3; Proc. ASCE Specialty Conference, Berkeley, pp. 1071-1111.
Clerici A, Perego S, Tellini C, Vescovi P (2002). A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphology 48: 349–364.
Clerici A, Perego S, Tellini C, Vescovi P (2006). A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: the Baganza valley case study (Italian Northern Apennines). Environ. Geol. 50: 941–961.
Dahal RK, Hasegawa S, Nonomura S, Yamanaka M, Masuda T, Nishino K (2007). GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. Environ. Geol. Online first.
Dai FC, Lee CF (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology 42: 213– 228.
Dai FC, Lee CF, Li J, Xu ZW (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ. Geol. 40: 381-391.
Donati L, Turrini MC (2002). 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). Engr. Geol. 63: 277– 289.
Ercanoglu M, Gokceoglu Z (2002). Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkye) by fuzzy approach. Environ. Geol. 41: 720-730.
Fell R1(994). Landslide risk assessment and acceptable risk. Canadian Geotechnical J. 31: 261-272.
Gokceoglu C, Sonmez H, Ercanoglu M (2000). Discontinuity controlled probabilistic slope failure risk maps of the Altindag (settlement) region in Turkey. Engr. Geol. 55: 277–296.
Guzzetti F, Carrarra A, Cardinali M, Reichenbach P (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31: 181-216.
Jibson WR, Edwin LH, John AM (2000). A method for producing digital probabilistic seismic landslide hazard maps. Engr. Geol. 58: 271–289.
Lamelas MT, Marinoni O, Hoppe A, Riva J (2007). Doline probability map using logistic regression and GIS technology in the central Ebro Basin (Spain). Environ. Geol. Online first.
Lee S, Min K (2001). Statistical analysis of landslide susceptibility at Youngin, Korea. Environ. Geol. 40: 1095-1113.
Lee S, Choi J, Min K (2002a). Landslide susceptibility analysis and verification using the Bayesian probability model, Environ. Geol. 43: 120-131.
Lee S, Chwae U, Min K (2002b). Landslide susceptibility mapping by correlation between topography and geological structure: the Janghung area, Korea. Geomorphology 46: 149-162.
Lee S, Ryu JH, Min K, Won JS (2003a). Landslide Susceptibility Analysis using GIS and Artificial neural network. Earth Surface Processes and Landforms 27: 1361-1376.
Lee S, Ryu JH, Lee MJ, Won JS (2003b). Landslide susceptibility analysis using artificial neural network at Boeun, Korea. Environ. Geol. 44: 820-833.
Lee S, Choi U (2003c). Development of GIS-based geological hazard information system and its application for landslide analysis in Korea. Geosci. J. 7: 243-252.
Lee S, Ryu JH, Won JS, Park HJ (2004a). Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Engr. Geol. 71: 289-302.
Lee S, Choi J, Min K (2004b). Probabilistic Landslide Hazard Mapping using GIS and Remote Sensing Data at Boeun, Korea. Intl. J. Remote Sensing 25: 2037-2052.
Lee S, Talib JA (2005a). Probabilistic landslide susceptibility and factor effect analysis. Environ. Geol. 47: 982-990.
Lee S, Dan NT (2005b). Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. Environ. Geol. 48, 778–787
Lee S (2005c). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. Intl. J. Remote Sensing 26: 1477-1491.
Lee S, Lee MJ (2006a). Detecting landslide location using KOMPSAT 1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea. Adv. Space Res. 38: 2261–2271.
Lee S, Sambath T (2006b). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environ. Geol. 50: 847–855.
Lee S (2007a). Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea. Earth Surface Processes and Landforms, Published online.
Lee S, Pradhan B (2007b). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4: 33-41.
Lee S (2007). Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. Environ. Geol. 52, 615-623.
Luzi L, Pergalani F, Terlien MTJ (2000). Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems. Engr. Geol. 58: 313–336.
Ohlmacher GC, Davis JC (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansa, USA. Engr. Geol. 2157: 1-13.
Park YW, Kim KL, Yeo WK (1993). A study on Yongin-Ansung landslides in 1991 (in Korean). Korean Geotech. J. 9: 103-116.
Parise M, Jibson WR (2000). A seismic landslide susceptibility rating of geologic units based on analysis of characteristics of landslides triggered by the 17 January, 1994 Northridge, California earthquake. Engr. Geol. 58: 251–270.
Pistocchi A, Luzi L, Napolitano P (2002). The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. Environ. Geol. 41: 765–775.
Pradhan B, Lee S (2007). Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model; Earth Science Frontiers 14 (6): 143- 152.
Pradhan B, Singh RP, Buchroithner MF (2005). Estimation of stress and its use in evaluation of landslide prone regions using remote sensing data; Adv. Space Res. 37: 698- 709
Refice A, Capolongo D (2002). Probabilistic modeling of uncertainties in earthquake-induced landslide hazard assessment. Comp. Geosci. 28: 735-749.
Romeo R (2000). Seismically induced landslide displacements: a predictive model. Engr. Geol. 58: 337-351.
Rowbotham D, Dudycha DN (1998). GIS modeling of slope stability in Phewa Tal watershed, Nepal. Geomorphology 26: 151-170.
Shou KJ, Wang CF (2003). Analysis of the Chiufengershan landslide triggered by the 1999 Chi-Chi earthquake in Taiwan. Engr. Geol. 68: 237-250.
Su¨zen ML, Doyuran V (2004). A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environ. Geol. 45: 665–679.
Tangestani MH (2004). Landslide susceptibility mapping using the fuzzy gamma approach in a GIS, Kakan catchment area, southwest Iran. Aust. J. Earth Sci. 51: 439–450.
Temesgen B, Mohammed MU, Korme T (2001). Natural Hazard Assessment Using GIS and Remote Sensing Methods, with Particular Reference to the Landslides in the Wondogenet Area, Ethiopia. Phys. Chem. Earth (C) 26: 665-615.
Tunusluoglu MC, Gokceoglu C, Nefeslioglu HA, Sonmez H (2007). Extraction of potential debris source areas by logistic regression technique: a case study from Barla, Besparmak and Kapi mountains

(NW Taurids, Turkey). Environmental Geology Online first.
Varne DJ (1984). Landslide hazard zonation: A review of principles and practice. Natural Hazards 3: 63.
Wang HB, Sassa K (2005). Comparative evaluation of landslide susceptibility in Minamata area, Japan. Environ. Geol. 47: 956–966.
Whitsman RV (1984). Evaluating calculated risk in geotechnical engineering. ASCE J. Geotechnical Engr. 110: 145-188.
Xie M, Esaki T, Cai M (2004). A time-space based approach for mapping rainfall-induced shallow landslide hazard. Earth surface processes and landforms. 26: 1251-1263.
Zhou G, Esaki T, Mitani Y, Xie M, Mori J (2003). Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach. Engr. Geol. 68: 373-386.
Zhou CH, Lee CF, Li J, Xu ZW (2002). On the spatial relationship between landslides and causative factors on Lantau Island, Hong Kong. Geomorphology 43: 197-207.