Mathematical Geology, Vol. 38, No. 2, February 2006 (© 2006)
DOI: 10.1007/s11004-005-9012-x

The Application of Artificial Neural Networks to Landslide Susceptibility Mapping at Janghung, Korea[1]

Saro Lee,[2] Joo-Hyung Ryu,[3] Moung-Jin Lee,[4] and Joong-Sun Won[4]

The purpose of this study was to develop techniques for landslide susceptibility using artificial neural networks and then to apply these to the selected study area at Janghung in Korea. Landslide locations were identified from interpretation of satellite images and field survey data, and a spatial database of the topography, soil, forest, and land use. Thirteen landslide-related factors were extracted from the spatial database. These factors were then used with an artificial neural network to analyze landslide susceptibility. Each factor's weight was determined by the back-propagation training method. Five different training sets were applied to analyze and verify the effect of training. Then the landslide susceptibility indices were calculated using the back-propagation weights, and susceptibility maps were constructed from Geographic Information System (GIS) data for the five cases. Landslide locations were used to verify results of the landslide susceptibility maps and to compare them. The artificial neural network proved to be an effective tool for analyzing landslide susceptibility.

KEY WORDS: back-propagation; training site; weight; GIS; spatial database.

INTRODUCTION

In Korea, frequent landslides have often resulted in significant damage to people and property, and in Janghung, our chosen study area, landslides have caused much damage. Heavy rainfall was the primary reason for these landslides.

Through scientific analysis of landslides, we hoped to assess and predict landslide-susceptible areas, and potentially decrease landslide damage through proper preparation. To achieve this aim, landslide hazard analysis techniques have

[1]Received 14 January 2003; accepted 9 May 2005; Published online: 20 May 2006.
[2]Geoscience Information Center, Korea Institute of Geoscience & Mineral Resources (KIGAM), 30, Gajung-Dong, Yusung-gu, Daejeon, 305-350, Korea; e-mail: leesaro@kigam.re.kr.
[3]Ocean Climate and Environmental Research Division, Korea Ocean Research and Development Institute (KORDI), Ansan P.O. Box 29, Seoul, 425-600, Korea.
[4]Department of Earth System Science, Yonsei University, 134, Shinchondong, Seoul, 120-749, Korea.

been developed, applied, and verified in the study area using artificial neural networks and the Geographic Information System (GIS). The study area sustained significant landslide damage following heavy rains in 1998, making it a suitable selection for evaluating its susceptibility to landslides (Fig. 1). The site lies between latitudes 37°43′ N and 37°46′ N, longitudes 126°56′ E and 127°01′ E, and covers an area of 40.74 km^2. The study area is in the northwestern part of the Kyonggi gneiss complex, which is composed mainly of gneisses (Lee, 1988).

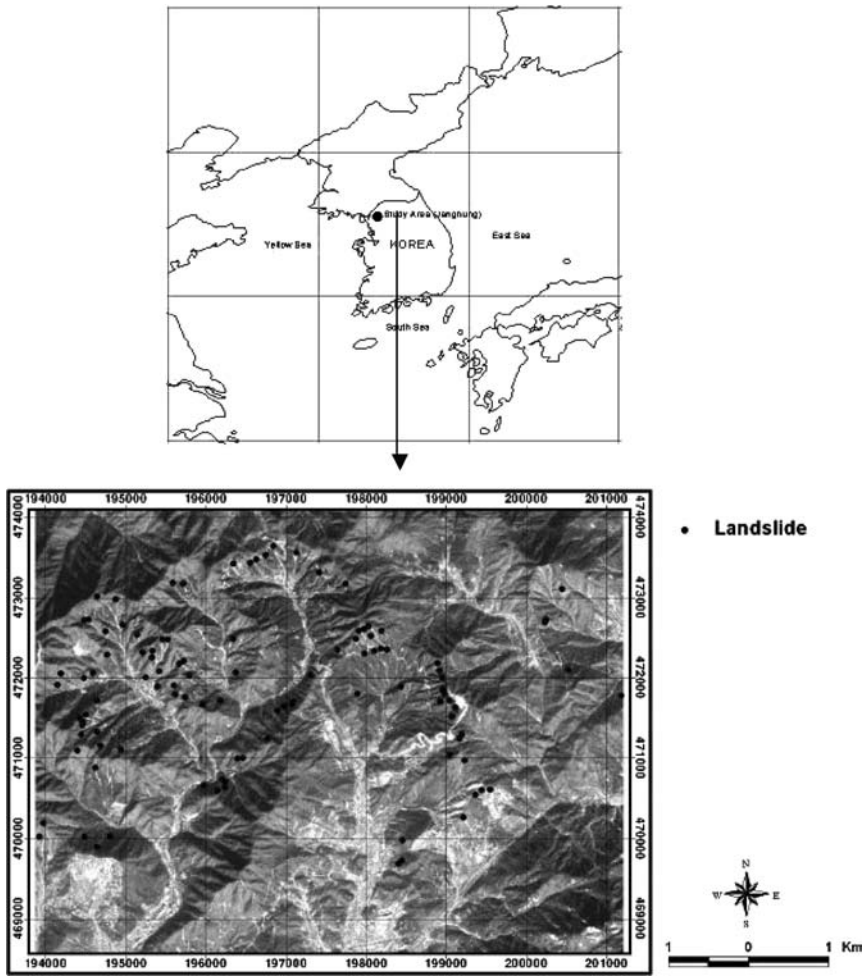

Figure 1. Landslide location with IRS satellite image.

There have been many recent studies of landslide hazard evaluation using GIS (Clerici and others, 2002; Dai and Lee, 2002; Donati and Turrini, 2002; Fernández and others, 1999; Guzzetti and others, 1999; Lee, Choi, and Min, 2002a,b; Lee and Min, 2001; Luzi, Pergalani, and Terlien, 2000; Mandy and others, 2001; Randall, Edwin, and John, 2000; Temesgen, Mohammed, and Korme, 2001; Turrini and Visintainer, 1998). The application of GIS to landslide hazard analysis is a useful and efficient tool. Artificial neural network methods have previously been applied to land use and cover classification using satellite imagery (Benediktsson, Swain, and Ersoy, 1990; Schaale and Furrer, 1995; Serpico and Roli, 1995). In particular, the multilayer perceptron method using the back-propagation algorithm was widely used in a supervised classification algorithm (Atkinson and Tatnall, 1997).

For the landslide susceptibility analysis of this study, the landslide occurrence areas were detected by interpretation of satellite images and field survey data (Figure 1). Maps were constructed in a vector format spatial database for the GIS and these data were used in the application of artificial neural network methods. These included topographic maps, soil maps, forest maps, and 30-m-resolution land use data from Landsat TM satellite images. From the spatial database, 13 factors were calculated and extracted for landslide susceptibility analysis: slope, aspect, curvature, topographic type, soil texture, soil material, soil drainage, soil effective thickness, forest type, timber age, timber diameter, forest density, and land use. Then the landslide susceptibility was analyzed using an artificial neural network program that was partially modified from an original version developed by Hines (1997) in the MATLAB 10.0 software package.

For the artificial neural network application, the locations of the landslide occurrences and the results of landslide susceptibility mapping provided by probability and statistical methods were used to create five training datasets for supervised classification. Landslide susceptibility was mapped using the trained back-propagation neural network. The results predicted by the artificial neural network were converted to grid data, and a landslide susceptibility map was generated using the GIS. Finally, these forecast results were verified using actual landslide locations.

THE 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, 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 referred to as the network architecture (Fig. 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).

$$\text{net}_j = \sum_i w_{ij} o_i \tag{1}$$

where w_{ij} represents the weights between nodes i and j, and o_i is the output from node j, given by

$$o_j = f(\text{net}_j). \tag{2}$$

The function f is usually a nonlinear 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:

$$f'(\text{net}_j) = f(\text{net}_j)(1 - f(\text{net}_j)). \tag{3}$$

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

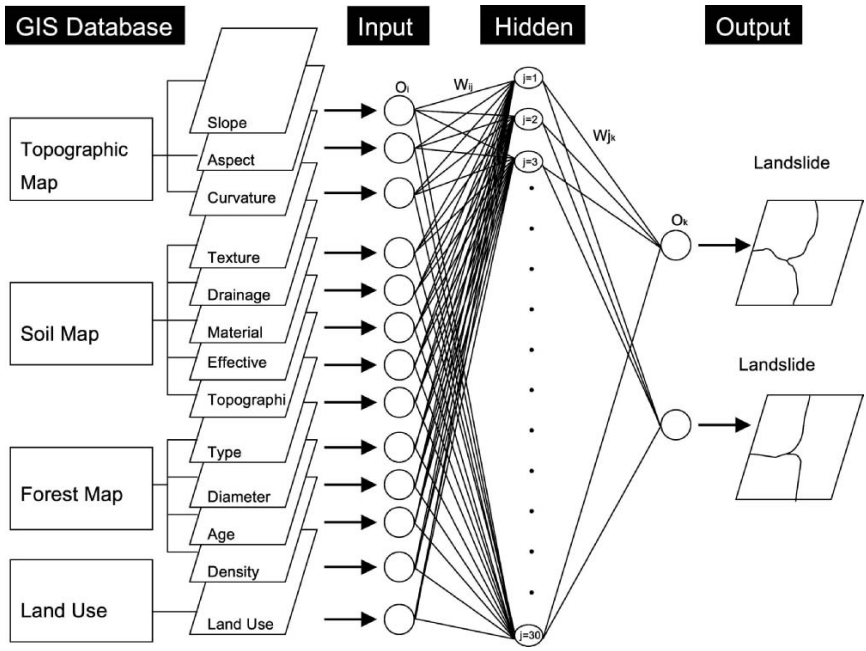

Figure 2. Architecture of artificial neural network.

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 vector, d, and the actual output vector, o, given by

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

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

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

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

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

$$\text{for hidden layers, } \delta_j = \left(\sum \delta_k w_{jk}\right) f'(\text{net}_j) \tag{6}$$

$$\text{and for output layers, } \delta_j = (d_k - o_k) f'(\text{net}_k). \tag{7}$$

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

$$\frac{\partial o_k}{\partial o_j} = f'(\text{net}_k) \times \frac{\partial(\text{net}_k)}{\partial o_j} = f'(\text{net}_k) \times w_{jk}. \tag{8}$$

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):

$$\frac{|\partial o_k|}{|\partial o_j|} \Bigg/ \frac{|\partial o_k|}{|\partial o_j|}\frac{|\partial o_k|}{|\partial o_{j0}|}\frac{|\partial o_k|}{|\partial o_{j0}|} = \frac{|f'(\text{net}_k) w_{jk}|}{|f'(\text{net}_k) w_{j0k}|} = \frac{|w_{jk}|}{|w_{j0k}|}. \tag{9}$$

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.

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

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

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

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

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

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

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

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

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

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

SPATIAL DATABASE USING GIS

Instability factors for landslides include lithology and geological structure, bedding altitude, seismicity, slope steepness and morphology, stream evolution, groundwater conditions, climate, vegetation cover, land use, and human activity. There is a wide availability of thematic data depending on the type, scale, and

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	1:5,000	10 m×10 m
Topographic Map	Slope	ARC/INFO Line and Point Coverage		1:5,000	
	Aspect				
	Curvature				
Soil Map	Texture	ARC/INFO Polygon coverage		1:25,000	
	Drainge				
	Material				
	Effetive Thickness				
	Topographic Type				
Forest Map	Type	ARC/INFO Polygon coverage		1:25,000	
	Diameter				
	Age				
	Density				
Land Use	Land Use	ARC/INFO GRID		30 m × 30 m	

method of data acquisition. To apply the artificial neural network, a spatial database was created that took landslide-related factors such as topography, soil, forest, and land use into consideration (Table 1).

Landslide occurrence areas were detected in the Janghung area, Korea, from both Indian Remote Sensing (IRS) and field survey data. In the study area, rainfall-triggered debris flows and shallow soil slides are the most abundant. Maps relevant to landslide occurrence were constructed in a vector format spatial database using the GIS ARC/INFO software package. These included 1:25,000 scale topographic maps, 1:50,000 scale soil maps, and 1:25,000 scale forest maps. Contour and survey base points that had an elevation value read from a topographic map were extracted, and a digital elevation model (DEM) was constructed. The DEM has a 10 m resolution and was used to calculate the slope, aspect, and curvature. Soil texture, parent material, drainage, effective thickness, and topographic type were extracted from the soil database. Forest type, timber age, timber diameter, and timber density were extracted from forest maps. Land use was classified from Landsat TM satellite imagery. There were only two types of lithology in the study area, and therefore, the geology was excluded in this study.

Both the calculated and extracted factors were converted to form a 10 × 10 m^2 grid (ARC/INFO grid type), and then were converted to ASCII data for use

with the artificial neural network program. The dimensions of the study area grid were 555 rows by 734 columns, and so the total number of cells was 407,370. Landslides occurred in 107 cells.

LIKELIHOOD RATIO AND LOGISTIC REGRESSION

Likelihood ratio approaches are based on observed relationships between the distribution of landslides and each landslide-related factor and are used to reveal the correlation between landslide locations and factors in the study area. The likelihood ratio is the ratio of occurrence probability to nonoccurrence probability for specific attributes. So if the ratio is greater than 1, the greater the relationship between a landslide location and the specific factor's attribute; and if the ratio is less than 1, the lower the relationship between a landslide location and the specific factor's attribute. The likelihood ratio is calculated from analyzing the between the location of the relation landslides location and the considered factors. Therefore, the likelihood ratios of each factor's type or range were calculated from their relationship with landslide events. The 14 landslide-related factors are slope, aspect, curvature and geomorphologic unit, texture, drainage, material, thickness of soil, forest type, stand age, stand diameter, stand density, lithology, and land use. To calculate the Landslide Susceptibility Index (LSI), each factor's likelihood ratio values were summed as in Equation (17).

$$\mathrm{LSI} = \mathrm{Fr}_1 + \mathrm{Fr}_2 + \cdots + \mathrm{Fr}_n \tag{17}$$

Fr is the rating of each factor's type or range. The landslide susceptibility value represents the relative susceptibility to landslide occurrence. The greater the value, the higher the susceptibility to landslide occurrence and the lower the value, the lower the susceptibility to landslide occurrence. The landslide susceptibility map of the training area was made using the LSI values and was used for interpretation.

Logistic regression, was used because it is useful when the observed outcome is restricted to two variables based on values of a set of predictor variables. The advantage of logistic regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types and do not necessarily have normal distributions. In the case of multiple regression analysis, the factors must be numerical. In the case of a similar statistical method, discriminant analysis, the variables must have a normal distribution. In the present situation, the dependent variable is a binary variable representing presence or absence. Where the dependent variable is binary, the logistic link function is applicable (Atkinson and Massari, 1998). For this study, the dependent variable must be input as either 0 or 1, so the method applies well to landslide occurrence possibility analysis. Quantitatively, the relationship between the occurrence and its dependency on

several variables can be expressed as

$$p = \exp(z)/(1 + \exp(z)) \tag{18}$$

where p is the probability of an event occurring.

In the present situation, p is the estimated probability of landslide occurrence. The probability varies from 0 to 1 on an S-shaped curve and z is the linear combination. It follows that logistic regression involves fitting an equation of the following form to the data:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \tag{19}$$

b_0 is the model's intercept; b_i are the slope coefficients of the logistic regression model; and x_i are independent variables.

The relationship between landslide-occurrence locations and landslide-related factors was calculated using the logistic regression method. The grids were converted to ASCII files for use in the statistical package that calculated the correlation between a landslide and each factor. After the calculation, coefficients of each factor's range or type and a constant were calculated; results are shown in Table 2. In addition, logistic regression equations were created using Equation (19). Finally, the probability that predicts the possibility of a landslide-occurrence for the training area was calculated using the spatial database, the coefficients from Table 2, and Equations (18) and (20):

$$\begin{aligned} z = {} & (0.035 \times \text{SLOPE}) + (-0.030 \times \text{CURVA}) + \text{ASPECT}_b + \text{TOPO}_b \\ & + \text{TEXTURE}_b + \text{DRAIN}_b + \text{SMATER}_b + \text{THICK}_b + \text{SANG}_b + \text{KUNG}_b \\ & + \text{YUNG}_b + \text{FMILDO}_b + \text{GEOL}_b + \text{LANDUSE}_b - 15.931 \end{aligned} \tag{20}$$

where SLOPE is the slope value; CURVA is the curvature value; ASPECT_b, TOPO_b, TEXTURE_b, DRAIN_b, SMATER_b, THICK_b, SANG_b, KUNG_b, YUNG_b, FMILDO_b, GEOL_b, LANDUSE_b are the logistic regression coefficient values listed in Table 2; and z is a parameter. A landslide susceptibility map was produced using these equations.

LANDSLIDE SUSCEPTIBILITY ANALYSIS USING THE ARTIFICIAL NEURAL NETWORK

The 13 factors listed in Table 1 were used as the input data. The landslide-prone (occurrence) locations and the locations that were not prone to landslides were selected as training sites. Pixels from each of the two classes were selected

Table 2. Likelihood and Logistic Regression Value

Factor	Class	No. of pixels in domain[a]	% of domain	No. of landslide[b]	% of land-slide	Ratio[c]	Logistic regression coefficients
Topographic slope (degree)	0–5	28168	6.91	0	0.00	0.00	0.0627
	6–10	41174	10.11	2	1.87	0.18	
	11–15	62663	15.38	7	6.54	0.43	
	16–20	69186	16.98	10	9.35	0.55	
	21–25	73028	17.93	21	19.63	1.09	
	26–30	61574	15.12	24	22.43	1.48	
	31–35	41987	10.31	23	21.50	2.09	
	36–40	20484	5.03	13	12.15	2.42	
	41–65	9106	2.24	7	6.54	2.93	
Topographic aspect	Flat	8969	2.20	0	0.00	0.00	−3.0738
	N	41223	10.12	5	4.67	0.46	0.3059
	NE	45094	11.07	7	6.54	0.59	0.2686
	E	59568	14.62	13	12.15	0.83	0.6529
	SE	64645	15.87	27	25.23	1.59	1.2583
	S	54475	13.37	21	19.63	1.47	1.1670
	SW	48558	11.92	20	18.69	1.57	1.2639
	W	44083	10.82	8	7.48	0.69	0.0810
	NW	40755	10.00	6	5.61	0.56	0.0000
Topographic curvature	Concave (−)	142808	35.06	41	38.32	1.09	−0.0148
	Flat (0)	121917	29.93	30	28.04	0.94	
	Convex (+)	142645	35.02	36	33.64	0.96	
Topographic type	Mountainous and hilly valley alluvium	38074	9.35	0	0.00	0.00	0.0000
	Piedmont slope area	14551	3.57	4	3.74	1.05	0.0000
	Hilly area	30429	7.47	4	3.74	0.50	−0.1326
	Mountainous area	248942	61.11	90	84.11	1.38	0.0000
	Low hilly area	75374	18.50	9	8.41	0.45	0.0000
Soil texture	Alluvial soil and humic gley soil	38074	9.35	0	0.00	0.00	—
	Regosols	13939	3.42	4	3.74	1.09	—
	Lithosols	282434	69.33	94	87.85	1.27	—

Table 2. Continued

Factor	Class	No. of pixels in domain[a]	% of domain	No. of landslide[b]	% of landslide	Ratio[c]	Logistic regression coefficients
	Red-yellow podzolic soils and lithosols	16479	4.05	0	0.00	0.00	—
	Regosols and red-yellow podzolic soils	612	0.15	0	0.00	0.00	—
	Rock	55832	13.71	9	8.41	0.61	—
Soil material	Valley alluvium	38074	9.35	0	0.00	0.00	0.0000
	Alluvial and colluvial	13939	3.42	4	3.74	1.09	0.0000
	Acidic rocks residuum	295850	72.62	94	87.85	1.21	7.9967
	Acidic rocks colluvium and sediment	612	0.15	0	0.00	0.00	7.0035
	Limestone residuum	3063	0.75	0	0.00	0.00	0.0000
	Rock	55832	13.71	9	8.41	0.61	0.0000
Soil drainage	Moderately well drained or poorly drained	13939	3.42	4	3.74	1.09	0.0000
	Moderately well drained or somewhat poorly drained	7001	1.72	0	0.00	0.00	0.0000
	Moderately well drained or well drained	31073	7.63	0	0.00	0.00	0.1093
	Well drained	17091	4.20	0	0.00	0.00	0.0000
	Excessively drained	282434	69.33	94	87.85	1.27	0.1267
	Rock	55832	13.71	9	8.41	0.61	0.0614

Table 2. Continued

Factor	Class	No. of pixels in domain[a]	% of domain	No. of landslide[b]	% of landslide	Ratio[c]	Logistic regression coeffi-cients
Soil effective thickness	Thick (100–150 cm)	390891	4.05	0	0.00	0.00	—
	Thick or medium thick (50–150 cm)	7001	11.20	0	0.00	0.00	—
	Medium thick (50–100 cm)	400369	69.33	4	3.74	0.33	—
	Thin (20–50 cm)	282434	13.71	94	87.85	1.27	—
	Rock	55832	4.05	9	8.41	0.61	—
Forest type	Nonforest	34540	8.48	0	0.00	0.00	−6.9880
	Pine	644	0.16	0	0.00	0.00	−8.3891
	Korea nut pine	84274	20.69	52	48.60	2.35	0.3458
	Broad leaf tree	164640	40.42	42	39.25	0.97	−0.6022
	Needle and broad leaf tree	73800	18.12	11	10.28	0.57	−0.6178
	Field	9168	2.25	0	0.00	0.00	−8.2868
	Cultivated land	18650	4.58	0	0.00	0.00	−8.7499
	Water	10	0.00	0	0.00	0.00	−0.5065
	Larch	13239	3.25	0	0.00	0.00	−7.7656
	Rigida pine	8405	2.06	2	1.87	0.91	0.0000
Timber diameter	Nonforest	62357	15.31	0	0.00	0.00	0.7496
	Medium diameter (more than 50% 16 × 28 cm)	6222	1.53	0	0.00	0.00	0.0000
	Small diameter timber (more than 50% 6×16 cm)	279803	68.69	74	69.159	1.01	0.0569

Table 2. Continued

Factor	Class	No. of pixels in domain[a]	% of domain	No. of landslide[b]	% of landslide	Ratio[c]	Logistic regression coeffi-cients
	Very small diameter (more than 50% below 6 cm)	58988	14.48	33	30.841	2.13	8.2322
Timber age	Nonforest	62357	15.31	107	0.00	0.00	0.0000
	1st age (more than 50% 1×10 years old)	58989	14.48	74	30.84	2.13	−8.4756
	2nd age (more than 50% 11×20 years old)	75649	18.57	82	23.36	1.26	−0.0253
	3rd age (more than 50% 21×30 years old)	204153	50.11	58	45.79	0.91	0.0000
	4th age (more than 50% 31×40 years old)	5738	1.41	107	0.00	0.00	0.0564
	5th age (more than 50% 41×50 years old)	484	0.12	107	0.00	0.00	0.0000
Forest density	Nonforest	121346	29.79	33	30.84	1.04	0.0000
	Loose (forest area less than 50%)	4592	1.13	0	0.00	0.00	−7.3839
	Moderate (forest area 51–70%)	156171	38.34	27	25.23	0.66	−0.2093
	Dense (forest area more than 71%)	125261	30.75	47	43.93	1.43	0.0000
Land use	Water	819	0.20	0	0.00	0.00	0.0000
	Urban	8295	2.04	0	0.00	0.00	−5.8363
	Forest	289320	71.02	87	81.31	1.14	−5.2396
	Grass	14133	3.47	1	0.93	0.27	0.7000
	Agriculture	43842	10.76	13	12.15	1.13	−6.4240
	Barren	50961	12.51	6	5.61	0.45	0.7197

[a]Total pixels in study area is 407,370.
[b]Total number of landslide occurrences is 107.
[c]% landslide /% domain.

as training pixels, with 107 pixels denoting areas where landslides occurred. The training sites were processed 10 times to identify any changes that might occur due to the assignment of random initial weights. First, areas where the slope was zero were classified as "areas not prone to landslides," and areas where landslides were known to exist were assigned to an "areas prone to landslides" training set (Case 1). Second, after application of the likelihood ratio model, areas with the lowest index values were classified into an "areas not prone to landslides" training dataset, and areas where landslides had occurred were classified into an "areas prone to landslides" dataset (Case 2). Third, after application of the logistic regression model, areas with low index values were classified into an "areas not prone to landslides" training set, and areas where landslides had occurred were classified into an "areas prone to landslides" training set (Case 3). Forth, after application of likelihood ratio model, areas with low index values were classified into an "areas not prone to landslides" training set, and after application of likelihood ratio model, areas with high index value were classified into an "areas prone to landslides" training set (Case 4). Fifth, after application of the logistic regression model, areas with low index values were classified into an "areas not prone to landslides" training set, and after application of logistic regression model, areas with high index value were classified into an "areas prone to landslides" training set (Case 5). If the analysis selected more than 107 sites with the same value, then the sites were selected randomly.

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. A three-layered feed-forward network was implemented using the MATLAB software package based on the framework provided by Hines (1997). 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 $13 \times 30 \times 2$ 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 results show that the initial weights did not have an influence on the final weight under the conditions used. 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 2000, and the

root mean square error (RMSE) value used for the stopping criterion was set to 0.1. Most of the training data sets met the 0.1 RMSE goal. However, if the RMSE value was not achieved, then the maximum number of iterations was terminated at 2000 epochs. When the latter case occurred, then the maximum RMSE value was <0.2.

The final weights between layers acquired during training of the neural network and the contribution or importance of each of the 13 factors used to predict landslide susceptibility are shown in Table 3. The results were not the same, as the initial weights were assigned random values. Therefore, in this study, the calculations were repeated 10 times, to allow the results to achieve similar values. The standard deviation of the results was in the range 0–0.014, and therefore, the random sampling did not have a large effect on the results. Average values were calculated for easy interpretation, and these values were divided by the minimum value weighting. 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. In the Case 1, the timber age value was the minimum value, 1.00, and the aspect was the maximum value, 1.716. In the Case 2, the curvature value was the minimum value, 1.00, and the slope was the maximum value, 1.540. In the other cases, the slope showed the maximum value. Finally, the weights were applied to the entire study area, and the landslide susceptibility index value was then calculated.

LANDSLIDE SUSCEPTIBILITY FORECAST MAPPING AND VERIFICATION

The calculated landslide susceptibility index values computed using back-propagation was converted into an ARC/INFO grid. Then a landslide susceptibility map was created. The final landslide susceptibility maps are shown in Figure 3. In these maps, the index range was classified into 10 classes, each having an equal area for easy visual interpretation.

Verification was performed by comparing existing landslide data with the landslide susceptibility analysis results of the study area. Comparison results of the five test cases (using estimation methods) are shown in Figure 4. The line graphs illustrate how well the estimators performed with respect to the landslides used in constructing the estimators (Chung and Fabbri, 1999). To obtain relative ranks for each prediction pattern, the calculated index values of all the cells in the study area were sorted in descending order. Then, the ordered cell values were plotted along the *x*-axis, with the accumulated intervals plotted along the *y*-axis. For example, the 10% class in Figure 4 contains a 30% success rate for the study area using a logistic regression–logistic regression method on areas not prone to landslides and areas that are prone to landslides, and the 20% class occupies 55% of the study area.

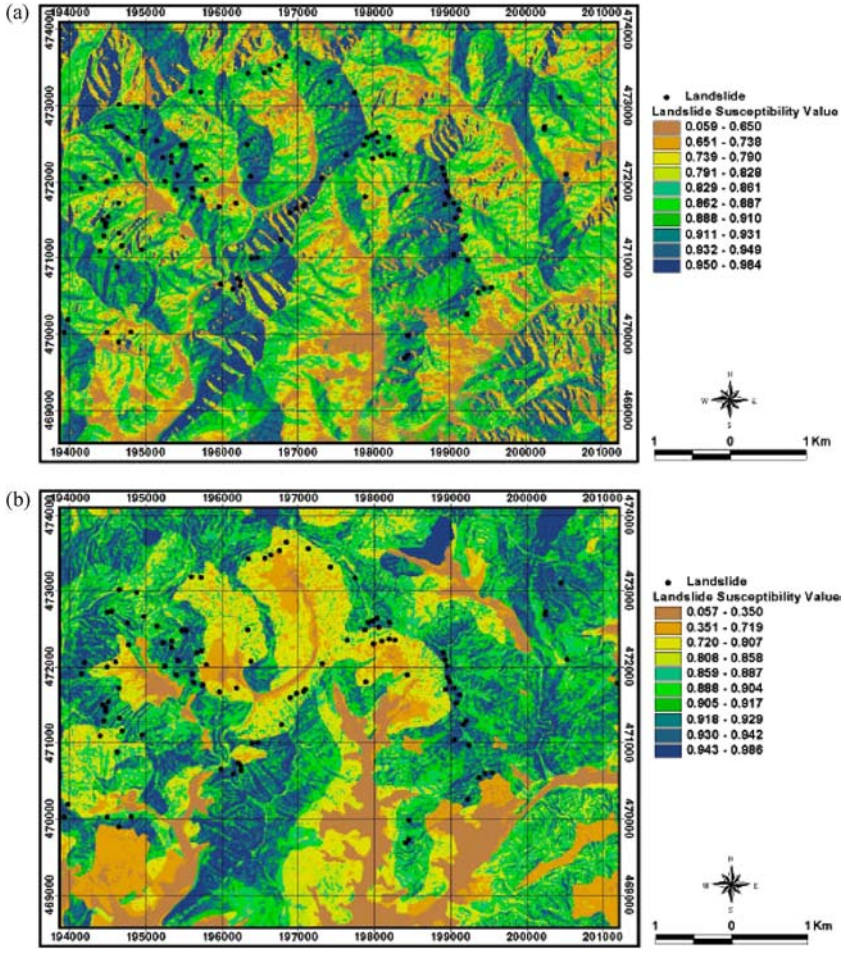

Figure 3. Landslide susceptibility mapping using artificial neural network. (a) Case 1: Use of landslide location as prone training site and slope is 0 as nonprone training site. (b) Case 2: Use of landslide location as prone training site and result from likelihood ratio as nonprone training site. (c) Case 3: Use of landslide location as prone training site and result from logistic regression as nonprone training site. (d) Case 4: Use of result from likelihood ratio as prone training site and result from likelihood ratio as nonprone training site. (e) Case 5: Use of result from logistic regression as prone training site of result from logistic regression as nonprone training site.

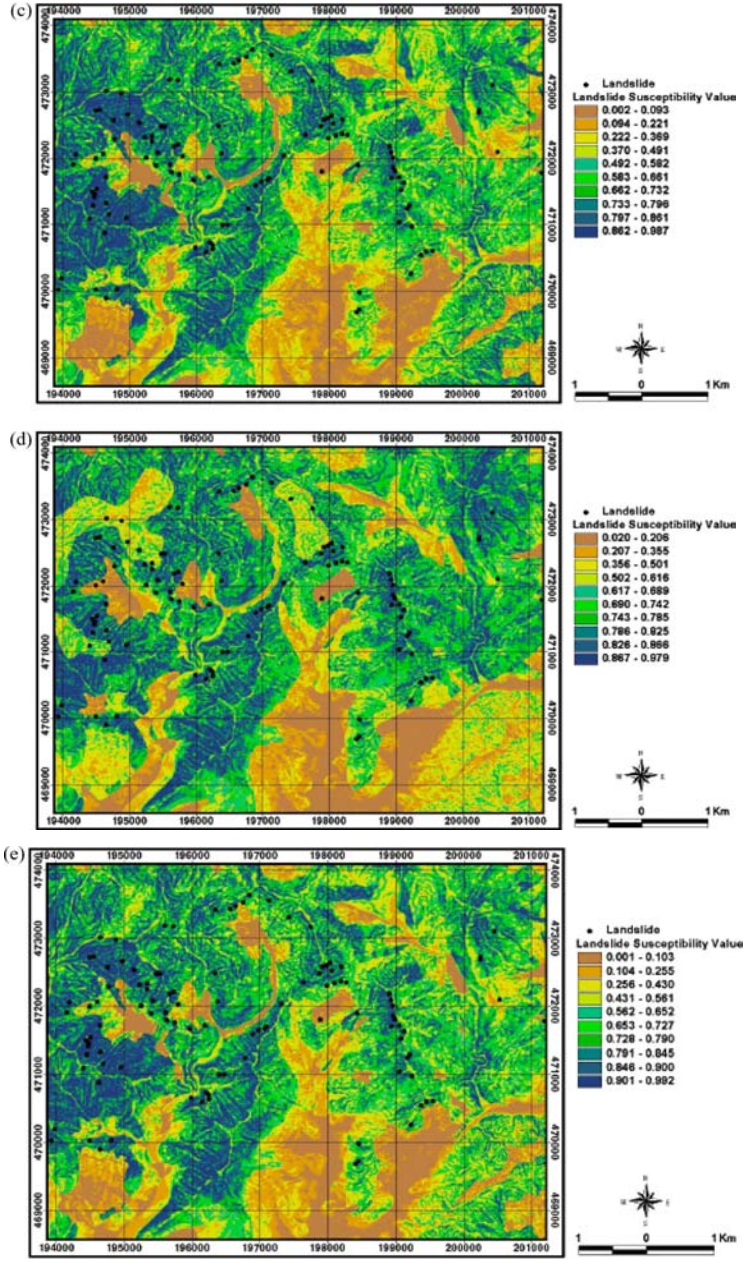

Figure 3. Continued.

Table 3. Weights of Each Factor for Each Selection of Training Site

	Case 1			Case 2			Case 3			Case 4			Case 5		
Factors	Average	Standard deviation	Weight	Average	Standard deviation	Weight	Average	Standard deviation	Weight	Average	Standard deviation	Weight	Average	Standard deviation	Weight
Slope	0.106	0.009	1.554	0.105	0.007	1.540	0.123	0.007	1.916	0.109	0.011	1.635	0.116	0.010	1.765
Aspect	0.117	0.007	1.716	0.071	0.008	1.035	0.071	0.007	1.116	0.069	0.010	1.031	0.070	0.010	1.074
Curvature	0.071	0.005	1.046	0.068	0.008	1.000	0.066	0.010	1.035	0.067	0.005	1.000	0.066	0.008	1.003
Topographic type	0.069	0.010	1.008	0.068	0.011	1.000	0.072	0.012	1.119	0.071	0.010	1.068	0.081	0.009	1.231
Soil Texture	0.069	0.006	1.018	0.073	0.009	1.073	0.064	0.009	1.000	0.073	0.010	1.092	0.071	0.009	1.089
Soil drainage	0.072	0.010	1.053	0.087	0.007	1.279	0.082	0.011	1.277	0.101	0.010	1.515	0.079	0.010	1.207
Soil effective thickness	0.071	0.009	1.049	0.069	0.006	1.013	0.067	0.010	1.050	0.073	0.008	1.096	0.066	0.008	1.003
Soil Material	0.068	0.009	1.006	0.070	0.009	1.017	0.067	0.012	1.042	0.070	0.013	1.042	0.071	0.010	1.082
Forest type	0.073	0.009	1.070	0.073	0.005	1.064	0.086	0.007	1.337	0.068	0.006	1.015	0.085	0.011	1.299
Timber diameter	0.074	0.010	1.093	0.102	0.012	1.499	0.087	0.009	1.358	0.079	0.014	1.189	0.086	0.014	1.313
Forest density	0.073	0.011	1.068	0.071	0.011	1.036	0.073	0.011	1.137	0.076	0.010	1.134	0.066	0.011	1.000
Timber age	0.068	0.009	1.000	0.075	0.012	1.100	0.075	0.013	1.178	0.073	0.011	1.092	0.076	0.008	1.157
Land use	0.069	0.008	1.018	0.067	0.011	0.978	0.068	0.009	1.055	0.071	0.007	1.070	0.067	0.012	1.016

CONCLUSIONS

Landslides are one of the most hazardous natural disasters, not only in Korea, 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. An artificial neural network approach has been used to estimate areas susceptible to landslides using a spatial database for a selected study area in Janghung, Korea. Five estimation methods were used for comparison purposes. The results in case 5 with logistic regression on prone training sites and with logistic regression on nonprone training sites (Fig. 3(e)) and result from Case 4 when likelihood ratio on prone training sites and with likelihood ratio on nonprone training sites (Fig. 3d) were better than the other three estimation cases (Figs. 3(a)–(c)). The results using landslide location as prone training site and result from likelihood ratio as nonprone training site being the worst (Fig. 3(b)).

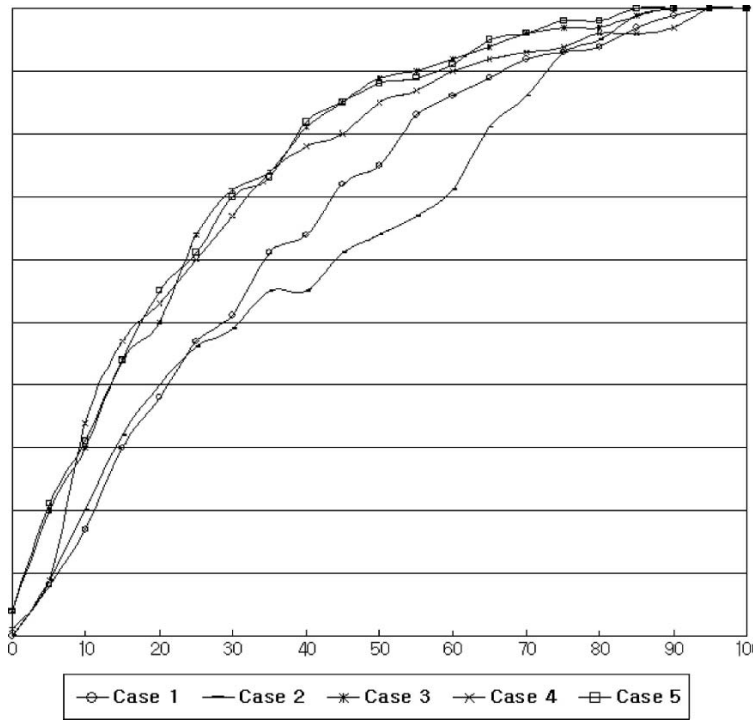

Figure 4. Success rate curve of each case.

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.

REFERENCES

Atkinson, P. M., and Tatnall, A. R. L., 1997, Introduction to neural networks in remote sensing: Int. J. Remote Sensing, v. 18, no. 4, p. 699–709.

Benediktsson, J. A., Swain, P. H., and Ersoy, O. K., 1990, Multisource data classification and feature extraction with neural networks: IEEE Trans. Geosci. Remote Sensing, v. 28, p. 540–552.

Chung, C. F., and Fabbri, A. G., 1999, Probabilistic prediction models for landslide hazard mapping: Photogrammetr. Eng. Remote Sensing, v. 65, no.12, p. 1389–1399.

Clerici, A., Perego, S., Tellini, C., and Vescovi, P., 2002, A procedure for landslide susceptibility zonation by the conditional analysis method: Geomorphology, v. 48 p. 349–364.

Dai, F. C., and Lee, C. F., 2002, Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong: Geomorphology, v. 42 p. 213–228.

Donati, L., and Turrini, M. C., 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): Eng. Geol., v. 63 p. 277–289.

Fernández, C. I., Castillo, T. F., Hamdouni, R. E., and Montero, J. C., 1999, Verification of landslide susceptibility mapping: A case study: Earth Surf. Process. Landforms, v. 24, p. 537–544.

Garrett, J., 1994. Where and why artificial neural networks are applicable in civil engineering: J. Comp. Civil Eng., v. 8, no. 2, p. 129–130.

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

Hines, J. W., 1997, Fuzzy and neural approaches in engineering. Wiley, New York, p. 210.

Lee, D. 1988. Geology of Korea. Kyohak-Sa Inc, Seoul, 514.

Lee, D., Xu, L. J., 2001, Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong: Environ. Geol., v. 40, p. 381—391.

Lee, S., Choi, J., and Min, K., 2002a, Landslide susceptibility analysis and verification using the Bayesian probability model: Environ. Geol., v. 43, p. 121–130.

Lee, S., Chwae, U., and Min, K., 2002b, Landslide susceptibility mapping by correlation between topography and geological structure: the Janghung area, Korea: Geomorphology, v. 46, p. 149–162.

Lee, S., and Min, K., 2001, Statistical analysis of landslide susceptibility at Yongin, Korea: Environ. Geol., v. 40, p. 1095–1113.

Luzi, L., Pergalani, F., and Terlien, M. T. J., 2000, Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems: Eng. Geol., v. 58, p. 313–336.

Mandy, L. G., Andrew, M. W., Richard, A., and Stephan, G. C., 2001, Assessing landslide potential using GIS, soil wetness modeling and topographic attributes, Payette River, Idaho: Geomorphology, v. 37, p. 149–165.

Paola, J. D., and Schowengerdt, R. A., 1995, A review and analysis of backpropagation neural networks for classification of remotely sensed multi-spectral imagery: Int. J. Remote Sensing, v. 16, no. 16, p. 3033–3058.

Randall, W. J., Edwin, L. H., and John, A. M., 2000, A method for producing digital probabilistic seismic landslide hazard maps: Eng. Geol., v. 58, p. 271–289.

Schaale, M., and Furrer, R., 1995, Land surface classification by neural networks: Int. J. Remote Sensing, v. 16, no. 16, p. 3003–3031.

Serpico, S. B., and Roli, F., 1995, Classification of multisensor remote-sensing images by structured neural network: IEEE Trans. Geosci. Remote Sensing, v. 33, p. 562–578.

Temesgen, B., Mohammed, M. U., and Korme, T., 2001, Natural hazard assessment using GIS and remote sensing methods, with particular reference to the landslides in the Wondogenet Area, Ethiopia: Physics and Chemistry of the Earth, Part C: Solar, Terrestrial Planet. Sci., v. 26, p. 665–675.

Turrini, M. C., and Visintainer, P., 1998, Proposal of a method to define areas of landslide hazard and application to an area of the Dolomites, Italy: Eng. Geol., v. 50, p. 255–265.

Zhou, W., 1999, Verification of the nonparametric characteristics of backpropagation neural networks for image classification: IEEE Trans. Geosci. Remote Sensing, v. 37, p. 771–779.