THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS TO LANDSLIDE SUSCEPTIBILITY MAPPING AT JANGHUNG, KOREA

SARO. LEE[1], MOUNG-JIN LEE[2], JOONG-SUN WON[2]

[1]Geoscience Information Center, Korea Institute of Geology & Mineral Resources (KIGAM)
30, Gajung-dong, Yusung-gu, Daejeon, 305-350, Korea
leesaro@kigam.re.kr
[2]Department of Earth System Science, Yonsei University, 134, Shinchondong, Seoul, 120-749, Korea

Abstract: The purpose of this study was to develop landslide susceptibility analysis techniques using artificial neural networks and then to apply these to the selected study area of Janghung in Korea. We aimed to verify the effect of data selection on training sites. 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 was constructed. Thirteen landslide-related factors were extracted from the spatial database. Using these factors, landslide susceptibility was analyzed using an artificial neural network. The weights of each factor were determined by the back-propagation training method. Five different training datasets were applied to analyze and verify the effect of training. Then, the landslide susceptibility indices were calculated using the trained back-propagation weights and susceptibility maps were constructed from Geographic Information System (GIS) data for the five cases. The results of the landslide susceptibility maps were verified and compared using landslide location data. GIS data were used to efficiently analyze the large volume of data, and the artificial neural network proved to be an effective tool to analyze landslide susceptibility.

Keywords: Landslide; susceptibility; artificial neural network; GIS

1. INTRODUCTION

In Korea, frequent landslides has often resulted in significant damage to people and property, and in our chosen study area, Janghung in Korea, much damage has been caused by landslides. The primary reason for these landslides was heavy rainfall, and, as there was little effort to assess or predict the event, damage was extensive. 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 developed, applied, and verified in the study area using artificial neural networks and the Geographic Information System (GIS). The study area sustained much landslide damage following heavy rains in 1998, and so was selected as a suitable case for evaluating its susceptibility to landslides. The site lies between latitudes 37°43' N and 37°46' N, and 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. For the landslide susceptibility analysis of this study, the landslide occurrence areas were detected by interpretation of satellite images and field survey data, and landslide location, topography, soil, and forest databases were used for the analysis. Maps were constructed in a vector format spatial database using the GIS data, and were used in the application of artificial neural network methods. These included topographic maps, soil maps, forest maps, and 30-meter-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 training datasets for supervised classification. Five different training datasets were applied to analyze and verify the effect of the training sites. Using the training datasets, the landslide susceptibility was analyzed using the trained back-propagation of the neural network layers. 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.

2. 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). An artificial neural network 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.

The most frequently used neural network method, and the method used in this study, is the back- propagation–learning algorithm. 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 of their inputs 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. The receiving node sums the weighted signals from all the nodes that it is connected to in the preceding layer.

3. SPATIAL DATABASE USING GIS

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

Landslide occurrence areas were detected in the Janghung area, Korea, by interpretation of the Indian Remote Sensing (IRS) and field survey data. In the study area, several types of landslides were documented, with rainfall-triggered debris flows and shallow soil being the most abundant. Topography, soil, and forest databases were also constructed. 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 using the DEM, the slope, aspect, and curvature were calculated. 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. In the case of geology, 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 x 10 m^2 grid (ARC/INFO grid type), and then 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. The number of cells where landslides occurred was 107.

4. LANDSLIDE SUSCEPTIBILITY ANALYSIS USING THE ARTIFICIAL NEURAL NETWORK

The 13 factors 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 as training pixels, with 107 pixels denoting areas of landslide occurrence. The training sites were processed 10 times to identify any changes in the initial weighting. First, areas where the slope was zero were classified into areas that were not prone to landslides, and areas where landslides were known to exist were assigned to an 'areas that were prone to landslides' training set. Second, after application of the likelihood ratio model, areas with the lowest index values were classified into an 'areas that were not prone to landslides' training dataset, and areas where landslides had occurred were classified into an 'areas that were prone to landslides' dataset. Similarly, after application of the logistic regression model, areas with low index values were classified into an 'areas that were not prone to landslides' training set, and areas where landslides had occurred were classified into an 'areas that were prone to landslides' training set. Fourth, after application of the likelihood ratio model, areas that had the lowest index values were set to 'areas that were not prone to landslides' training sites, and areas that had the highest index values were set to 'areas that were prone to landslides' training sites. Fifth, after application of the logistic regression model, areas with the lowest index values were set to 'areas that were not prone to landslides' training sites and areas with the highest index values were set to 'areas that were prone to landslides' training sites. If the analysis selected more than 107 sites having the same value, then the sites were selected randomly

Then, the back-propagation algorithm was 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 layers 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 forwards 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 x 30 x 2 model 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. The denotation of the numbers used was not considered in the calculations using the neural network program, but the number used to distinguish the classes of each factor was used in the calculations using the artificial neural network program. 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 2,000, 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 EMSE value was not archived, then the maximum number of iterations was terminated at 2,000 epochs. When the latter case occurred, then the maximum RMSE value was <0.2.

The calculations were repeated 10 times to see if the randomly extracted samples represented each class well. However, there were no large differences. 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. Finally, weights were applied to the entire study area, and the landslide susceptibility index value was then calculated.

5. LANDSLIDE SUSCEPTIBILITY FORECAST MAPPING AND VERIFICATION

The calculated landslide susceptibility index values were converted into an ARC/INFO grid. Then, a landslide susceptibility map was created. In these maps, the index range was classified into 10 classes, each having an equal area for easy visual interpretation.

Verification was performed by comparison of existing landslide data and the landslide susceptibility analysis results of the study area. Comparison results of the five test cases (using estimation methods) are shown in Figure 1. The line graphs, the success rates in Figure 1, 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 1 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.

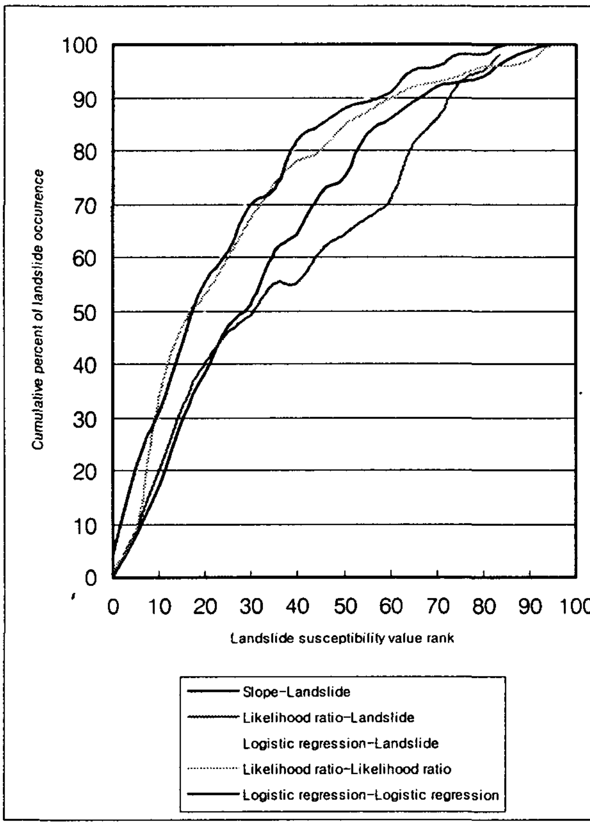

Fig. 1. Illustration of cumulative frequency diagram showing landslide susceptibility index rank (x-axis) occurring in cumulative percent of landslide occurrence (y-axis).

6. 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 using logistic regression–logistic regression and logistic regression–landslide methods were better than the other three estimation methods used, with the results from the likelihood ratio–landslide method being the worst.
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, and there are many advantages in using techniques that are data-driven. 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

C.F. Chung, A.G.Fabbri, 1999, Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing*, 65:1389-1399.

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

J.W.Hines, 1997, Fuzzy and neural approaches in engineering, John Wiley and Sons, Inc. New York: 210pp.

J.D.Paola and R.A.Schowengerd, 1995, A review and analysis of back propagation neural networks for classification of remotely sensed multi-spectral imagery. *International Journal of Remote Sensing*, 16(16):3033-3058.