CROSS-VALIDATION OF ARTIFICIAL NEURAL NETWORK FOR LANDSLIDE SUSCEPTIBILITY ANALYSIS: A CASE STUDY OF 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 aim of this study is to cross-validate of spatial probability model, artificial neural network at Boun, Korea, using a Geographic Information System (GIS). Landslide locations were identified in the Boun, Janghung and Youngin areas from interpretation of aerial photographs, field surveys, and maps of the topography, soil type, forest cover and land use were constructed to spatial data-sets. The factors that influence landslide occurrence, such as slope, aspect and curvature of topography, were calculated from the topographic database. Topographic type, texture, material, drainage and effective soil thickness were extracted from the soil database, and type, diameter, age and density of forest were extracted from the forest database. Lithology was extracted from the geological database, and land use was classified from the Landsat TM image satellite image. Landslide susceptibility was analyzed using the landslide-occurrence factors by artificial neural network model. For the validation and cross-validation, the result of the analysis was applied to each study areas. The validation and cross-validate results showed satisfactory agreement between the susceptibility map and the existing data on landslide locations.

Keywords: Landslide; Susceptibility; GIS; Artificial neural network; Validation, Korea

1. INTRODUCTION

In Korea, frequent landslides often result in significant damage to people and property, the most recent having occurred in 1991, 1996, 1998, 1999 and 2002. In the study area, Boun, Janghung and Yongin in Korea, much damage was caused on these occasions. The reason for the 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. In order to achieve this, landslide hazard analysis techniques were validated in the study area using artificial neural network model.

For the study, first, the study areas, Bound, Janghung and Youngin in Korea were selected. Then landslide occurrence areas were detected and topography, soil, forest, and land use databases were constructed to spatial data-sets. Using the detected landslide locations and the constructed spatial data sets, an artificial neural network model was applied and landslide susceptibility map was made. Then, the susceptibility map was validated using existing landslide location.

The first study area, Boun, lies between the latitudes 36 °25' 21" N and 36° 30' 00" N, and longitudes 127° 39' 36" E and 127° 45' 00" E, and covers an area of 68.43km^2. The bedrock geology of the study area consists mainly of biotite granite. The landslides occurred where the maximum daily rainfall is 407 mm. The second study area, Janghung 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. In the study area, the landslides occurred where the maximum daily rainfall is 208.5 mm. The third study area, the Yongin, lies between the latitudes 37.14° N and 37.19° N, and longitudes 127.11° E and 127.23° E, and covers an area of 66 km^2. The bedrock geology of the study area consists mainly of granite and gneiss. The landslides occurred where the maximum daily rainfall exceeded 114 mm, with a maximum hourly rainfall of 40mm.

2. SPATIAL DATA SETS AND METHODOLOGY

Identification and mapping of a suitable set of instability factors (thematic mapping) bearing a relationship with slope failures requires an a priori knowledge of the main causes of landslides (Guzzetti and others 1999). These instability factors include surface and bedrock lithology and structure, bedding altitude, seismicity, slope steepness and morphology, stream evolution, groundwater conditions, climate, vegetation cover, land-use, and human activity. The availability of thematic data varies largely, depending on the type, scale, and method of data acquisition. A digitised map of landslide boundaries was produced, and these digital data were input to the GIS. A vector-to-raster conversion was undertaken to provide a raster data of landslide areas. Maps relevant to landslide occurrence were constructed in vector-type spatial data sets using the ARC/INFO GIS software package. These included 1:5000-scale topographic maps, 1:25000 or 1:50,000-scale soil maps, and 1:25000-scale forest maps. In the Janghung, 1:50,000-scale soil map was used because there is no published 1:25:000-scale soil map. A land-use map was extracted from Landsat TM satellite images having a resolution of 30 m. Contour and survey base points that had an elevation value read from the topographic map were extracted, and a Digital Elevation Model (DEM) was constructed. Using the DEM, the slope, aspect and curvature were calculated. The topographic type, texture, drainage, material, and thickness were acquired from a soil map. The type, diameter, age and density were obtained from forest maps, and land use data was classified according to LANDSAT TM satellite images. In the study areas, the data sets were was divided into a grid with 10 m × 10 m cells. The Boun data set was composed of 555 rows by 734 columns, so the total cell number is 407,370 and the cell number where landslides occurred is 107. The Janghung data set was composed of 555 rows by 734 columns, so the total cell number is 407,370 and the cell number where landslides occurred is 107. The Youngin data set was composed of 555 rows by 734 columns, so the total cell number is 407,370 and the cell number where landslides occurred is 107.

For the landslide-hazard analysis, first, the study area was selected. Then landslide occurrence areas were detected in the Boun area, Korea by interpretation of aerial photographs and field surveys. A map of recent landslides was developed from 1:20,000 scale aerial photographs, in combination with the GIS, and this was used to evaluate the frequency and distribution of shallow landslides in the area. The factors such as altitude, slope, aspect and curvature from the topographic database, soil texture, material, drainage, effective thickness, and topography from the soil database, forest type, forest diameter, and forest density from the forest map, and land use data from Landsat TM image were used. Using the detected landslide locations and the constructed spatial database, a landslide analysis method, with artificial neural network, was applied and validated. To achieve this, the calculated and extracted factors were converted to a 10 m × 10 m grid (ARC/INFO GRID type), and then converted to ASCII data for use with the artificial neural network program. The analysis results were converted to grid data using GIS.

In this study, GIS (Geographic Information System) software, ArcView 3.2 and ARC/INFO 8.1 NT version, and statistical software, SPSS 10.0 was used as the basic analysis tool for spatial management and data manipulation.

3. ARTIFICIAL NEURAL NETWORKS AND WEIGHT DETERMINATION

An artificial neural network is a "computational mechanism able to acquire, represent, and compute a mapping from multivariate space of information to another given a set of data representing that mapping" (Garrett 1994). An artificial neural network is trained by the use of 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 is the back propagation-learning algorithm. This is a learning algorithm of a multi-layered neural network, which consists of an input layer, hidden layers, and an output layer. The S-shaped sigmoid function is commonly used as the transfer function. Artificial neural network "learns" by adjusting the weights between the neurons in response to the errors between actual output values and target output values. At the end of this training phase, the neural network represents a model, which should be able to predict a target value given an input value.

There are three stages involved in using neural networks for multi-source classification; the training stage, the determinating weight stage, and the classification stage. The back propagation algorithm trains the network, typically, until some targeted minimal error is achieved between the desired and actual output values of the network. Once training is complete, the network is used as a feed-forward structure to produce a classification for the entire database (Paola and Schwengerdt 1995). In this study, only inter-layer weights were used in the training stage.

4. LANDSLIDE SUSCEPTIBILITY ANALYSIS USING ARTIFICIAL NEURAL NETWORK

The weight between layers was acquired by training the neural network, which calculated reversely, and the contribution or importance of each factor was calculated. Therefore, the contribution or importance of each factor, weight, was determined. A GIS spatial database was used as input data and landslide locations were used as training sites. Among the artificial neural network methods the back propagation method was used. The program developed by Hines (1997) using MATLAB was partially modified for the landslide analysis. It was modified in the input and output parts for the use of GIS data.

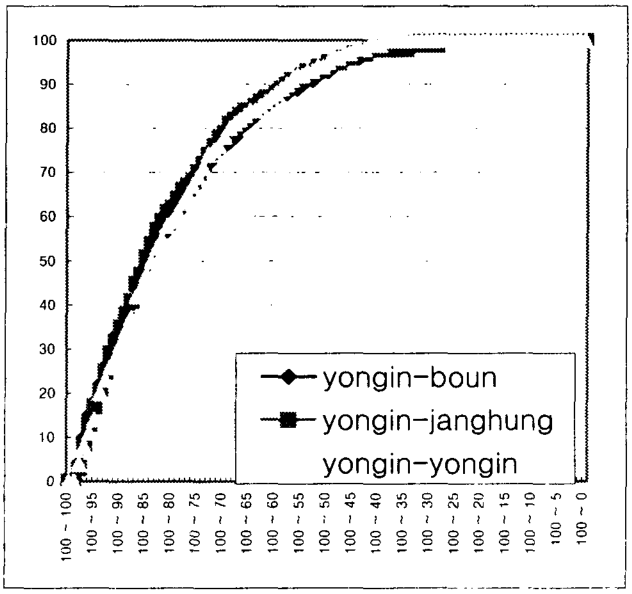
For analysis of landslide susceptibility, the training sites were selected from the landslide-related factors and the back propagation algorithm was applied to calculate weights between the input layer and the hidden layer, and between the hidden layer and the output layer, by modifying the number of hidden layers and the learning rate. A three-layered feed forward network was implemented in MATLAB on the basis of the framework provided by Hines (1997). Feed-forward means that all 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, the 13 factors were used for calculating the weights. Therefore, the structure 13 (input layer) × 30 (hidden layer) × 2 (output layer) used were selected for the network with input data normalized to the range 0.1 to 0.9.

The weights were applied to the each study areas and the landslide susceptibility index (LSI) values were calculated. The calculated index values were converted into an

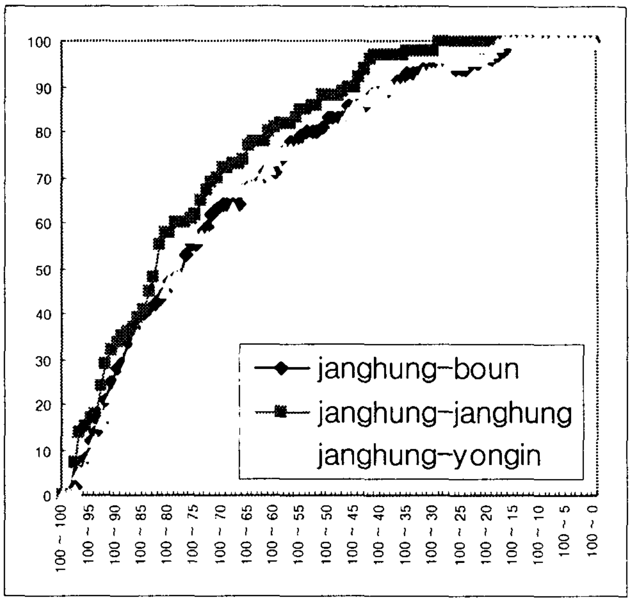
ARC/INFO GRID using the GIS. Then the landslide susceptibility map was created using the GRID data. The computed LSI value index was classified into equal areas and grouped into five classes for visual and easy interpretation.

5. VALIDATION OF LANDSLIDE SUSCEPTIBILITY MAPPING

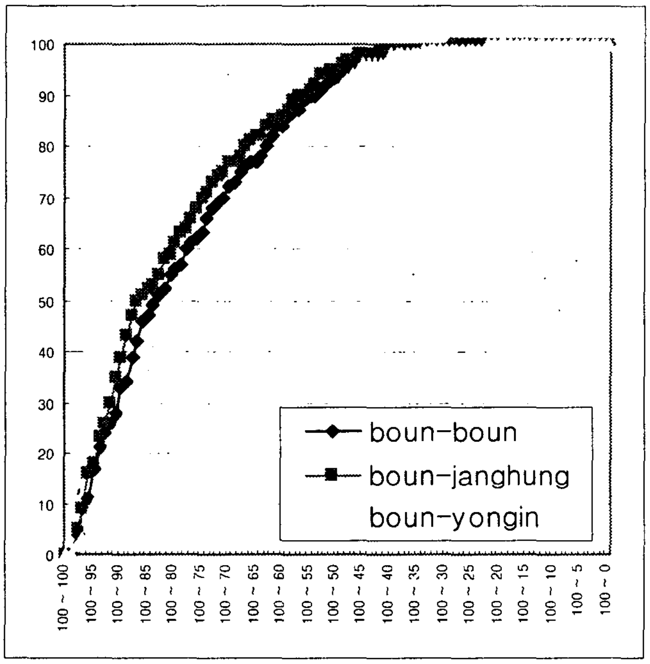
The landslide susceptibility analysis result validated using the landslide locations for the same study areas and cross-validated using the landslide locations of the others study areas. The validation method was performed by comparison of existing landslide data and landslide susceptibility analysis results for each study areas. The comparison results are shown in Fig. 1 as a line graph, with logistic multiple regression method at the case of success rate and prediction rate. The success rates in Fig. 1 illustrate how well the estimators perform with respect to the left side landslides used in constructing those estimators. The prediction rates in the Fig. 1, on the other hand, are used as measurements of how well the probability model and its estimators predict the distribution of future landslides. Therefore, strictly speaking, the success rate is not a suitable validation method. However, the success rate validation method needs information about the properties of analysis method, and checks the landslide susceptibility analysis calculation for major errors. It also needs to be tested against the prediction rate validation method.
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. The above procedure also was adapted for the each study areas by comparing the classes obtained with the distribution on the each study areas. In Fig. 1, the success rate validation results are divided into classes of accumulated area ratio % according to the landslide susceptibility index value.



(a)



(b)



(c)

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

6. DISCUSSION AND CONCLUSION

Landslides are among the most hazardous natural disasters. Government and research institutions worldwide have attempted for years to assess the landslide hazard and risk and to show its spatial distribution. In this study, a validation of probabilistic approach to estimating the susceptible area of landslides using GIS is presented. For the landslide susceptibility analysis, landslide location was detected using aerial photographs and a landslide-related database was constructed for the study area of

Boun, Janghung and Youngin, Korea.
For the landslide susceptibility analysis, artificial neural network model was applied and validated for the study area of Yongin, Korea, using the spatial data-sets. Using the 13 factors, likelihood ratio model was applied to analyze the landslide hazard. Then, the results were validated by calculating the correlation observed between landslide occurrence location and the predicted occurrences. Generally, the validation results showed satisfactory agreement between the susceptibility map and the existing data on landslide location.
In comparison between success rate and prediction rate, success rate showed the better accuracy than prediction rate for all cases. In the Janghung case for success rate show the best accuracy among the 3 cases in success rate. Among the 6 cases, Yongin rate for Janghung showed the best accuracy and Young rate for Janghung showed the worst accuracy.
In this study, only the susceptibility analysis was performed, because the small area studied did not allow us to determine the distribution of rainfall. However, if data on factors causing the landslides, such as rainfall, earthquake shaking, or slope cutting, exist, then the possibility analysis could also be done. In particular, if the data could be combined with a hydrological model, a more accurate analysis could be done. If the factors relevant to vulnerability of buildings and other property were available, risk analysis could also be done.

REFERENCES

J.Garrett, 1994, Where and why artificial neural networks are applicable in civil engineering, *Journal of Computing Civil Engineering*, 8:129-130.
F.Guzzetti, A.Carrarra, M.Cardinali and P.Reichenbach, 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, *Geomorphology*, 31:181-216.
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.