The Comparative Research of Landslide Susceptibility Mapping Using FR, AHP, LR, ANN

Oh Che Young[1], Cheung Kil Jin[2], Chul Uong Choi[3]

Abstract

The landslides in natural slopes generally occur by complicated problems such as soil properties, topography and geology. FR, AHP, LR, ANN is widely used to solve complicated problems in many research fields. The purpose of this study is the development, application and assessment of FR, AHP, LR, ANN methods for assessing landslide susceptibility in the Republic of Korea. Landslide-related factors such as topographical, hydrological, soil, forest, landcover factor were used in the LSI analysis. Each factors was constructed using 3D Analyst, Spatial Analyst(surface, distance, zonal) tools and Analysis tools(extract, overlay, clip, statics) of ArcGIS9.3. LSI maps were produced from FR, AHP, LR, ANN by Analysis tools(overlay), spatial analyst tools using constructed input data. And then were compared by means of their validations. Respective AUC(area under curve) values of FR, AHP, LR and ANN were 0.794, 0.789, 0.794, and 0.806. All model showed a similar accuracy and the map obtained from ANN is more accurate than the other models

Key Words : FR(Frequency Ratio), AHP(Analytical Hierarchy Process), LR(Logistic Regression), ANN(Artificial Neural Networks), LSI(Landslide Susceptibility Index), landslide

1. Introduction

Patterns of landslide occurrence are different depending on topography, geological conditions and local climate characteristics. Also, various factors and analytical methods are used to analyze the characteristics of occurred landslides. Spatial and temporal studies for landslide occurrence prediction have been carried out developing landslide occurrence models and using geotechnology, remote sensing, GIS and statistics. Prediction of the areas where are vulnerable to or potential landslide occurrence has been studied since the late 1960s, and Newman et al. reported in 1978 that building of landslide prediction map using computer is practically .

Recently many GIS-based studies with probability methods have been reported. ANN(artificial neural network), AHP(analytic hierarchy process), FR(frequency response) and LR(logistic regression) are the Typical statistics-based probability analytical approaches. In this study, real data for the landslide areas in Inje, Gangwon-do were built using GIS, and landslide prediction was estimated and compared applying ANN, AHP, FR and LR methods that had been used in previous studies. In addition, accuracy was verified by comparing inversely real landslide occurrence data with estimated results from each study model. The results of each study model were also compared with landslide hazard map, which is being used in the National Emergency Management Agency.

2. Study area

The AREA of this research is the Republic of Korea located on the Korean Peninsula at the easternmost end

[1] Master, Dept. of Geoinfomatic Engineering, Pukyung National University, 599-1 Daeyeon3-Dong, Nam-Gu, Busan 608-737, Korea(Email : leeieel@nate.com)

[2] Master, Dept. of Geoinfomatic Engineering, Pukyung National University, 599-1 Daeyeon3-Dong, Nam-Gu, Busan 608-737, Korea (Email : rekime@nate.com)

[3] Associate Professor, Dept. of Geoinfomatic Engineering, Pukyung National University, 599-1 Daeyeon3-Dong, Nam-Gu, Busan 608-737, Korea, Korea, Tel : +82-51-629-6655, Tax : +82-51-625-2529 (Corresponding Author, Email : cuchoi@pknu.ac.kr)

of the Asian continent. The site lies between the latitudes 38°05′N and 38°07′N, and the longitudes 128°30′E and 128°35′E. The study area was the upper stream of North Han River of Inje area, Gangwon-do, Korea. As 91% of Inje area is composed of mountains and rivers including Seorak Mt., Bangtae Mt., Soyang Lake and Naerincheon, it is famous for the best nature experience destination in Korea thanks to rafting and various festivals(Fig. 1) The subject landslide of the study was that occurred in the area over Nammyon, Bukmyeon and Girimyeon in Inje-eup, Inje-gun, Gangwon-do that recorded the greatest daily precipitation. Particularly, in this small area of 3640m x 2235m, landslides in about 300 locations occurred accompanied by tremendous damage. Landslide areas in this study site consist of collapsed zone of 162㎡, travelling zone of 2,080㎡ and sediment zone of 1,178㎡, and landslides that are mainly characterized by small-sized debris flow occurred.

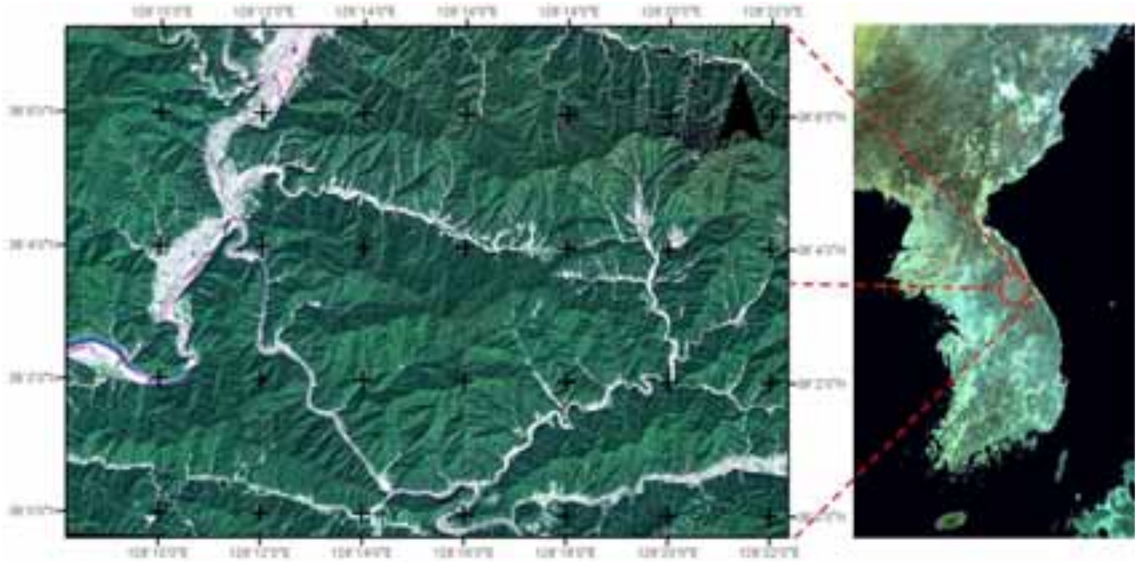

Fig. 1. Study area

3. Data collection and processing

Materials used in research of landslide susceptibility mapping are very diverse(Table 1).

Table 1. Input data

Factor class	Data Name		Type	Scale	Resolution	Creation	Producer
Topographical factor	Digital map	Elevation	Vector→ Grid	1:5,000	5m	2003	Korea National Geographic Information Institute
		Slope	TIN→ Grid				
		Aspect	TIN→ Grid				
Hydrological factor	Calculate from digital map	Drainage distance	Line→ Grid	1:5,000	5m	2003	ArcInfo
		SPI	Grid				
		TWI	Grid				
Soil factor	Detailed soil map	Texture	Vector→ Grid	1:5,000	5m	1998 ~ 2006	Korea National Institute of Agricultural Science and Tech.
		Depth	Vector→ Grid				
Forest factor	4th Forest type map	Type	Vector→ Grid	1:25,000	10m	1998 ~ 2007	Korea Forest Service
		Diameter	Vector→ Grid				
		Age	Vector→ Grid				
		Density	Vector→ Grid				
Land cover factor	Land cover map		Raster	1:25,000	10m	2006	Korea Ministry of Environment

This research collected the relevant data needed for setting up every sort of information related to landslide.

The collected data were constructed in the form of space database using ArcGIS 9.3 and ERDAS 9.2 Program. Built database for the 4th Forest type map and land cover map was converted into AcrInfo GRID file in 10m×10m grid size, while for the other data was in 5m×5m grid size(Fig. 2).

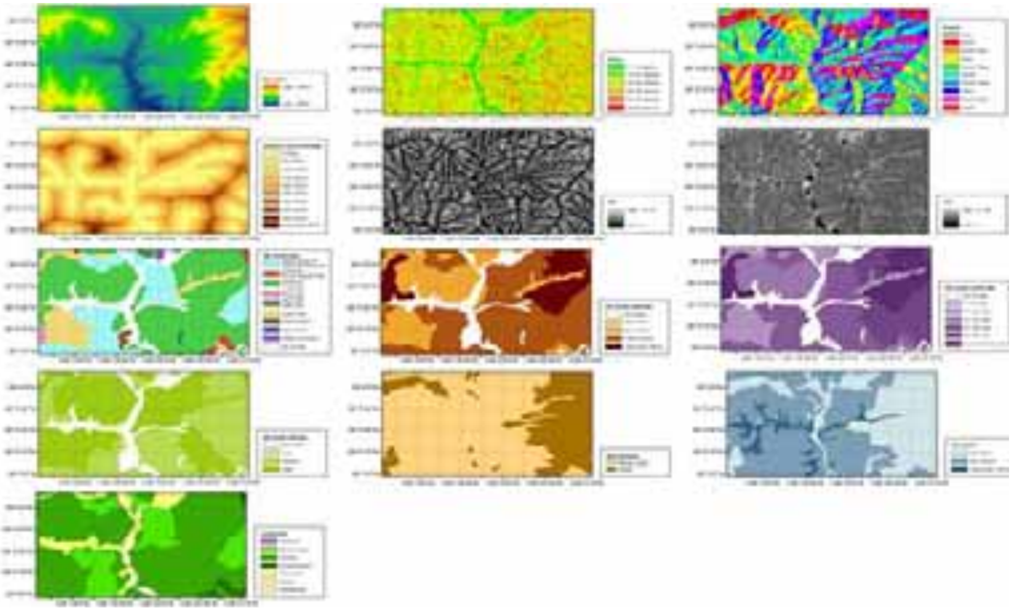

Fig. 2. Parameter maps of study area

Classes, types and scales of the input data of the GIS database built in this study for this study site were as in the Table 1 and the details are as follows :

In addition, in this study, topographic map was used to calculate drainage distance, TWI(topographic wetness index) and SPI(standardized precipitation index). Distance from Drainage was calculated using the Distance function of ArcView 3.3. The TWI , developed by Beven and Kirkby(1979) within the runoff model, has been used to study spatial scale effects on hydrological processes. It is definded as Eq. 1(Yilmaz, 2009). SPI is directly proportional to stream power, which is the time rate of energy expenditure and so is a measure of the erosive power of overland flow(Moore et al., 1991). It is defined as Eq. 2.

$$\mathrm{TWI} = \ln\left(\frac{\alpha}{\tan\beta}\right) \qquad \text{(Eq. 1)}$$

$$\mathrm{SPI} = \alpha \times \tan\beta \qquad \text{(Eq. 2)}$$

Where α is the specific catchment area, β is the slope angle(Yilmaz, 2009).

4. Landslide susceptibility mapping

4.1. Probability methods (FR(Frequency ratio))

That is, to construct a probability model for landslide hazard, it is necessary to assume that landslide occurrence is determined by landslide-related factors, and the future landslides will occur under the same conditions as past landslides. The FR is the ratio of the probability of an occurrence to the probability of a non-occurrence for given attributes(Bonham-Carter, 1994).

Therefore, the FR of each factor's type or range were calculated from their relationship with landslide events as shown in Table 2. If this ratio is greater than 1, the stronger the relationship between landslide occurrence and the given factor's attribute(Lee and Talib, 2005).

Using the probability model, the FR were calculated for 13 factors(Table 1). The 13 factors were converted in the form of grid to calculate the LSI(Landslide Susceptibility Index). Using GIS software, the grids were overlaid with the geographic coverage for study area. During the overlay, the LSI is calculated by adding all the grade value of each factor with each factor given the fixed weight of 1 by using Eq.3. The LSI represents the relative hazard to landslide occurrence. So the greater the value, the higher the hazard to landslide occurrence and the lower the value, the lower the hazard to landslide occurrence(Lee and Pradhan, 2007; Lee and Dan, 2005).

$$\mathrm{LSI} = \sum \mathrm{Fr} \quad (\mathrm{Fr} : \text{Rating of each factor's type or range}) \quad \text{(Eq. 3)}$$

Table 2 shows the relationship between landslide occurrence and each factor. First, Topographical factors, such as elevation, slope, aspect were used. In the case of the relationship between landslide occurrence and elevation, landslide generally occur at the elevation range 600~800m. The ratio greater than 1 are distributed at the topographic elevations between 600 and 800m. For slope angle above 15° except for in a range 45~50°, the ratio was >1, which indicates a high probability of landslide occurrence. Generally, shear stress is related to the slope. As the slope angle increases, the shear stress on the slope material increases. Analyses showed that frequency ratio increased with increasing of slope degree. Slope aspect analyses showed that landslide were most abundant in eastern, southern, southeastern, southwestern slopes.

In the case of the relationship between landslide occurrence and the distance from drainage, as the distance from drainage increases, the landslide ratio generally increase. At a distance of >450m, the ratio was 0, indicating zero probability. And the ratio was the highest (1.5) when TWI was 13~14, and was still high (>1) even when TWI was in the range of 5~7. The ration was >1 except when SPI was in the range 0~0.4, 3.6~4.0 and 5.6~6.4, and exhibited the highest when SPI was in the range of 6.8~7.2. In the case of the relationship between landslide occurrence and the soil factors, the ratio was >1 for the sandy loam and the soil depth in the range of 50~100cm. Relationship between forest type, age-class, girth-class and density was analyzed as factors related to landslide and forest type.

The ratio was high in conifer and pine tree forests in regard of forest type. In regard of age-class, it was high for 21-year-old or older except for the age range of 41~50. In regard of girth-class, the ratio was high for ≥18cm in diameter. Also, in regard of forest density, it was high in low-density areas. In the case of the relationship between landslide occurrence and the landcover, landslide occurrence values were higher in mixed forest and conifer(Table 2).

Table 2. FR factor and AHP rating to landslide

Classes	Landslide[a)] (%)	Domain[b)] (%)	FR[c)]	AHP scale	Classes	landslide (%)	Domain (%)	FR	AHP scale
Elevation					Aspect				
450~500	1.3	5.8	0.2	1	N	6.7	11.1	0.6	8
500~550	9.9	17.3	0.6	2	NE	6.5	9.6	0.7	7
550~600	15.3	24.3	0.6	3	E	13	11.7	1.1	4
600~650	26.6	22.8	1.2	4	SE	20.6	14.3	1.4	1
650~700	16.6	15.3	1.1	5	S	19.1	15.3	1.2	3
700~750	16	7.5	2.1	6	SW	11.9	9.1	1.3	2
750~800	13.6	5.2	2.6	7	W	7.8	8.4	0.9	5
800~	0.7	1.9	0.4	8	NW	8.4	10.3	0.8	6
SLOPE					SPI				
0~5	5.8	10.3	0.6	1	0~0.4	18.6	20.7	0.9	1
5~10	0	2.1	0	1	0.4~0.8	12.2	12.1	1	1
10~15	1.6	6.6	0.2	2	0.8~1.2	12.8	13.4	1	2
15~20	14.4	15	1	2	1.2~1.6	14	14.6	1	2
20~25	22.7	20.8	1.1	3	1.6~2.0	13.5	13.5	1	3
25~30	19.7	18.6	1.1	3	2.0~2.4	10.8	9.8	1.1	3
30~35	18.4	12.7	1.5	4	2.4~2.8	6.3	6	1.1	4
35~40	9.4	7.5	1.3	4	2.8~3.2	4.8	3.3	1.5	4
40~45	5.1	3.8	1.4	5	3.2~3.6	2	1.9	1	5
45~50	1.5	1.7	0.9	5	3.6~4.0	0.5	1.1	0.4	5

50~55	0.7	0.7	1.1	6	4~4.4	1.4	0.8	1.9	6
55~60	0.4	0.3	1.6	6	4.4~4.8	1.1	0.6	2	6
60~70	0.1	0.1	1.7	7	4.8~5.2	0.8	0.4	2	7
70~	0.1	0	2.8	7	5.2~5.6	0.3	0.3	1	7
TWI					5.6~6.4	0.1	0.4	0.4	8
3~4	0.5	4.6	0.1	1	6.4~6.8	0.3	0.2	1.3	8
4~5	0	0	0	1	6.8~7.2	0.3	0.1	2.4	8
5~6	12.6	10.6	1.2	2	7.2~	0.1	0.1	1.3	8
6~7	73.5	65.9	1.1	3	Forest type				
7~8	10.1	12.9	0.8	4	1(Conifer)	31	24	1.3	1
8~9	1.8	2.8	0.6	5	2(Other)	0	0	0	8
9~10	0.7	1.4	0.5	6	4(Pine tree)	62	49	1.3	2.5
10~11	0.5	0.7	0.7	7	5(Agriculture)	0	0	0	7
11~12	0.2	0.4	0.5	8	6(Broadleaf)	5	13	0.4	4
12~13	0	0.3	0	8	12(No tree)	1	14	0.1	5.5
13~14	0.2	0.1	1.5	8	Forest diameter(cm)				
14~15	0	0.1	0	8	0(Other)	1.2	13.8	0.1	1
15~16	0	0.1	0	8	2(6~18)	32.1	37.7	0.9	4
16~17	0	0.1	0	8	3(18~28)	52.7	40.4	1.3	6
					4(upper28)	14	8	1.7	8

a) Landslide(%) : Percent of landslide, b) Domain(%) : Percent of domain, c) FR : landslide(%) /domain(%)

Table 2. continued

Classes	Landslide[a)] (%)	Domain[b)] (%)	FR[c)]	AHP scale	Classes	landslide (%)	domain (%)	FR	AHP scale
Soil texture					Forest year(year)				
1(sandy loam)	97.4	89	1.1	1	0(other)	1.2	13.8	0.1	1
2(loam)	2.6	11	0.2	8	2(11~20)	2.3	8.2	0.3	4
Soil depth(cm)					3(21~30)	29.8	29.5	1	5
2(20~50)	4	21.8	0.2	1	4(31~40)	48.9	35.1	1.4	6
3(50~100)	93.5	66	1.4	4.5	5(41~50)	3.8	5.4	0.7	7
4(100~)	2.5	12.2	0.2	8	6(51~60)	14	8	1.7	8
Drain distance(m)					Forest density				
0~50	10.1	22.8	0.4	1	0(Other)	1.2	13.9	0.1	1
50~100	14.8	18.1	0.8	2	1(Low)	14.1	2.8	5	3.3
100~150	16.6	15.5	1.1	3	2(Midium)	29.3	21	1.4	5.6
150~150	10.5	14.2	0.7	4	3(High)	55.4	62.4	0.9	8
200~250	14.2	11.9	1.2	5	Landcover				
250~300	20.9	9.3	2.3	6	4(Wetland)	0	0.5	0	8
300~350	7.5	5.3	1.4	7	7(Mixed forest)	28.2	22.3	1.3	1
350~400	3.1	2.1	1.5	8	8(Conifer)	67.8	66	1	2.7
400~450	2.4	0.8	3.1	8	13(Crop land)	3.5	9	0.4	4.4
450~	0	0.2	0	8	14(Paddy)	0.5	2.2	0.2	5.2

a) Landslide(%) : Percent of landslide, b) Domain(%) : Percent of domain, c) FR : landslide(%) /domain(%)

5.2. AHP(Analytical Hierarchy Process) using Logistic Regression Weight

That is, AHP is a decision support system designed to seek optimum decision making for a complex circumstance through hierarchical structure, which is comprised of targets to be attained, various criteria for decision making and alternatives to be selected(Yang et al, 2006).

In the construction of a pair-wise comparison matrix, preference is denoted by a vector of weights following an AHP scale of relative importance ranging from 1 to 9(Bantayan and Bishop, 1998). So the pair-wise matrix A among the decision variable of $A_1, A_2, A_3, \cdots A_n$ utilizing the interval scale is the $n \times n$ matrix. The a_{ij}, the element of A has same meaning with Eq. 4.

$$a_{ij} = \frac{\text{priorities of factor } A_i}{\text{priorities of factor } A_j} = \frac{W_i}{W_j} \quad (i, j = 1, 2, \cdots, n) \qquad \text{(Eq. 4)}$$

The square matrix A of $n \times n$ of such meaning can be expressed by the matrix composed by the weighted ratio(Yoon, 2002). The relative importance of factors is possible to determine the degree of consistency that has been used in developing the judgments(Dai et al., 2001).

In AHP, an index of consistency, known as the CR(Consistency Ratio), is used to indicate the probability that the matrix judgments were randomly generated(Satty, 1977)

$$CR = CI/RI \qquad \text{(Eq. 5)}$$

Where RI is the average of the resulting consistency index depending on the order of the matrix given by Satty(1977) and CI is the consistency index and can be expressed as

$$CI = (\lambda_{max} - n)/(n-1) \qquad \text{(Eq. 6)}$$

Where λ_{max} is the largest or principal eigenvalue of the matrix and n is the order of the matrix(Saaty, 1977). As reported by Satty(1990), if the ratio of CI to that from random matrices is significantly small(carefully specified to be about 10% or less), we accept the estimate of w. Otherwise, we attempt to improve consistency(Saaty, 1990)

Usually, the AHP was calculated in relative weights among the factors by performing the questionnaire (Ex : How much do you think that A is more important than B?) to the specialist with provision of the questionnaire in 5 steps or 7 steps. The results used for making the pair-wise comparison matrix and used mainly for the interpretation of liberal art and social data(Saaty, 1977; Yoon, 2002).

The AHP method is adopted for the spatial decision making using the GIS recently, however, it has the problem of scale rating decision method of metric data in standardization of factor measurements, the problem of not to be used as a nominal data which is non-metric data and problem of acquiring the objectivity in calculation of relative weights with participation of a few specialists only.

However, this study calculated the pair-wise comparison matrix of AHP utilizing “B” which is the factor wise weight calculated through logistic regression in SPSS 13 after the standardization of factor measurements. The objectivity has been increased differently with conventional studies through this and the weighted value for the 13 factors is shown in Table 3.

Table 3. Pair-wise comparison matrix for assessing the weights of factors

	1	2	3	4	5	6	7	8	9	10	11	12	13	Coeffi-cient	Wei-ght
1	1.0	2.4	-9.9	2.0	-10.6	7.4	-0.5	-1.3	0.7	-3.3	-0.5	-38.6	1.6	0.3	-1.2
2	0.4	1.0	-4.1	0.8	-4.4	3.1	-0.2	-0.6	0.3	-1.4	-0.2	-16.0	0.7	0.1	-0.5
3	-0.1	-0.2	1.0	-0.2	1.1	-0.7	0.1	0.1	-0.1	0.3	0.1	3.9	-0.2	0.0	0.1
4	0.5	1.2	-5.0	1.0	-5.4	3.7	-0.3	-0.7	0.3	-1.7	-0.3	-19.6	0.8	0.2	-0.6
5	-0.1	-0.2	0.9	-0.2	1.0	-0.7	0.0	0.1	-0.1	0.3	0.1	3.7	-0.2	0.0	0.1
6	0.1	0.3	-1.3	0.3	-1.4	1.0	-0.1	-0.2	0.1	-0.4	-0.1	-5.2	0.2	0.0	-0.2
7	-1.9	-4.6	18.8	-3.8	20.1	-14.0	1.0	2.5	-1.3	6.3	1.0	73.4	-3.0	-0.7	2.2
8	-0.7	-1.8	7.4	-1.5	7.9	-5.5	0.4	1.0	-0.5	2.5	0.4	28.9	-1.2	-0.3	0.9
9	1.5	3.7	-15.0	3.0	-16.0	11.2	-0.8	-2.0	1.0	-5.0	-0.8	-58.5	2.4	0.5	-1.8
10	-0.3	-0.7	3.0	-0.6	3.2	-2.2	0.2	0.4	-0.2	1.0	0.2	11.7	-0.5	-0.1	0.4
11	-1.9	-4.6	18.6	-3.7	19.9	-13.9	1.0	2.5	-1.2	6.2	1.0	72.9	-3.0	-0.7	2.2
12	0.0	-0.1	0.3	-0.1	0.3	-0.2	0.0	0.0	0.0	0.1	0.0	1.0	0.0	0.0	0.0
13	0.6	1.5	-6.2	1.2	-6.6	4.6	-0.3	-0.8	0.4	-2.1	-0.3	-24.2	1.0	0.2	-0.7

1 : elevation, 2 : slope, 3 : aspect, 4 : drain distance, 5 : TWI, 6 : SPI, 7 : forest type, 8 : tree diameter, 9 : forest year
10 : forest density, 11 : soil texture, 12 : soil depth, 13 : landcover, Consistency ratio : 0.0000

The conventional calculation method calculated from pair-wise comparison matrix is the integer type (1, 2, 3, 4, 5, 6, 7) or fraction type (1/1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7), but this study calculates in float type. The conventional weights are float type of unsigned 0~1 but the weight in the study calculated in signed float type. n addition, the characteristic of data was calculated in positive correlation but this study calculated the positive and negative correlation.

4.3. Logistic regression model

Logistic regression model, which was developed by McFadden(1973), employs the use of independent variables to create a mathematical formula that the probability that an event occurs on any given parcel of land. The key to logistic regression is that the dependent variable is dichotomous. The independent variables in this model are predictors of the dependent variable and can be measured on a nominal, ordinal, interval or ratio scale. The relationship between the dependent variable and independent variables is nonlinear(Yesilnacar and Topal, 2005).

The assumed results are shown in Table 4. The value of Sig is less than 0.05 as against all variables except C(aspec), E(TWI), F(SPI), J(broadleaf), P(soildepth), R(mixed forest), U(paddy), so these variables appeared significant at a 5% significance level. In short, These variables except the above mentioned 7 variables are found to have a statistically significant effect on landslide.

Table 4. Binary logistic model result

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
A: elevation	0.0082	0.0	61.1	1.0	0.0	1.0	1.0	1.0
B: slope	0.0217	0.0	20.1	1.0	0.0	1.0	1.0	1.0
C: aspect	0.0002	0.0	0.2	1.0	0.7	1.0	1.0	1.0
D: drain_dist	0.0031	0.0	29.9	1.0	0.0	1.0	1.0	1.0
E: twi	0.0637	0.0	1.6	1.0	0.2	1.1	1.0	1.2
F: spi	0.0244	0.0	0.4	1.0	0.5	1.0	0.9	1.1
G: f4type	·	·	53.7	4.0	0.0	·	·	·
H: f4type(1)	6.4349	1.2	28.7	1.0	0.0	623.2	59.1	6571.1
I: f4type(2)	5.6841	1.2	24.0	1.0	0.0	294.1	30.3	2856.9
J: f4type(3)	-15.9575	40193.0	0.0	1.0	1.0	0.0	0.0	.
K: f4type(4)	5.0071	1.2	16.8	1.0	0.0	149.5	13.7	1634.9
L: f4diameter	-0.8492	0.4	5.0	1.0	0.0	0.4	0.2	0.9
M: f4yaer	0.6257	0.2	7.8	1.0	0.0	1.9	1.2	2.9
N: f4dencity	-0.4151	0.1	9.7	1.0	0.0	0.7	0.5	0.9
O: soilteture(1)	4.7086	0.5	75.7	1.0	0.0	110.9	38.4	320.4
P: soildepth	-0.0218	0.1	0.0	1.0	0.9	1.0	0.7	1.3
Q: landcover	·	·	33.8	4.0	0.0	·	·	·
R: landcover(1)	-20.4063	11977.1	0.0	1.0	1.0	0.0	0.0	·.
S: landcover(2)	-2.0739	0.6	11.9	1.0	0.0	0.1	0.0	0.4
T: landcover(3)	-2.2004	0.6	15.4	1.0	0.0	0.1	0.0	0.3
U: landcover(4)	-1.0388	0.6	3.1	1.0	0.1	0.4	0.1	1.1
Constant	-14.0552	1.4	102.5	1.0	0.0	0.0	·	·

* Variable(s) entered on step 1: elevation, slope, aspect, drain_dist, twi, spi, f4type, f4diameter, f4yaer, f4dencity, soilteture, soildepth, landcover.

4.4. Artificial neural network(ANN)

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).

The S-shaped sigmoid function is commonly used as the transfer function. An artificial neural network "learns" by adjusting the weights between the neurons in response to the errors between actual output values and target output values(Lee et al., 2003). The back-propagation algorithm randomly selects the initial weights, then compares the calculated output for a given observation with the expected output for that observation. The difference between the expected and calculated output values across all observations is summarized using the mean squared error(Rumelhart and Williams, 1986).

$$E = \| T - t \|^2 \qquad \text{(Eq. 11)}$$

where T is a target value, t is a true value, $\| \;\|$ is a vector's 2-norm. This process of feeding forward signals and back-propagating the errors is repeated iteratively(in some cases, many thousands of times) until the error stabilizes at a low level(Pijanowski et. al., 2002).

MatLab 7.0 was used for training and testing the neural networks. A three-layer feed-forward network that consists of an input layer(13 neurons), one hidden layer(25 neurons) and one output layer was used as a network structure of 13-25-1. In this study, 2983 training samples(Landslide occurrence locations : 1,368, No landslide 1,615) were used. The input data normalized to the 0.1-0.9, learning rate was set to be 0.01, and the initial weights were randomly selected. The number of epochs were set to 2,500 and RMSE(Root Mean Square Error) goal for stopping criterion was set to 0.01.

5. Validation and Accuracy analysis of model

5.1 AUC(Area Under the Curve) Validation in model

In this study, the ROC curve was used to assess the performance of FR, AHP, LR, ANN in modeling LSI map. For this purpose, the landslide susceptibility analysis result was tested using known landslide locations. Testing was performed by comparing the known landslide location data with the LSI map.

Table 5. AUC value in randomly sample area

	Area	Std. Error[a)]	Asymptotic Sig.[b)]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
FR	0.794	0.008	0.0	0.779	0.810
AHP	0.789	0.008	0.0	0.773	0.805
LR	0.794	0.008	0.0	0.779	0.810
ANN	0.806	0.008	0.0	0.771	0.821

a) Under the nonparametric assumption, b) Null hypothesis : true area = 0.5

The AUC depicted in Fig. 8. It is explains how well the model and factor predict the landslide. The AUC value of FR, AHP, LR, ANN were 0.794, 0.789, 0.794, 0.806 respectively. All the models were found to be acceptable, because their AUC(area under curve) values were in the range of 0.794~0.806. All model showed a similar accuracy and the map obtained from ANN model looks like more accurate than the other models. Also, the case of less than 0.05 in asymptotic sig. value can be regarded as useful. The case of more than 0.5 in lower bound and upper bound in asymptotic 95% confidence interval can be regarded as useful. As it is shown in Table 5, all 4 models satisfy the both of two cases.

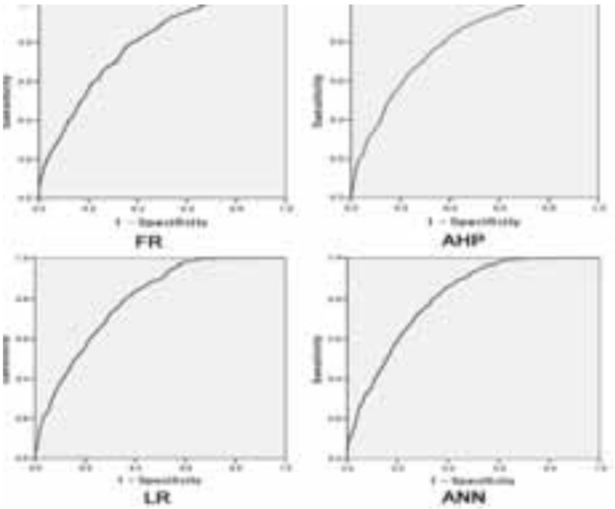

Fig. 8. AUC representing quality of the models used

5.2. Model real Accuracy in research area

The LSI maps produced by FR, AHP, LR and ANN were validated using AUC values. As a result they were proved to be effective. For this reason AUC values were recalculated over the whole study areas to evaluate quantitatively the prediction accuracy. The AUC values recalculated in this way are depicted in Fig. 9. The AUC values of FR, AHP, LR and ANN were 0.715, 0.712, 0.705 and 0.757, respectively. AUC values of all the models ranging from 0.705 to 0.757 were slightly lower than the accuracy in the models, 0.794~0.806. All model showed a similar accuracy and the map obtained from ANN model also looks like more accurate than the other models. Also, the case of less than 0.05 in asymptotic sig. value can be regarded as useful. The case of more than 0.5 in lower bound and upper bound in asymptotic 95% confidence interval can be regarded as useful. As it is shown in Table 6, all 4 models satisfy the both of two cases.

Also, LSI maps built in this study were compared with the landslide hazard maps provided by the Korea Forest Service through landslide spatial information distribution system. The landslide hazard maps were built for the whole area of the Republic of Korea over the years 2004 to 2005 in 1:25,000 scale. The landslide hazard map is comprised of the following 4 grades; 1st grade (serious: very high), 2nd grade (warning: high), 3rd grade (watchful: moderate), 4th grade (concerned: low).

Table 6. AUC value in research area

	Area	Std. Error[a)]	Asymptotic Sig.[b)]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
FR	0.715	0.007	0.0	0.702	0.729
AHP	0.712	0.007	0.0	0.699	0.725
LR	0.705	0.006	0.0	0.693	0.718
ANN	0.757	0.006	0.0	0.746	0.768
Landside hazard map[c)]	0.471	0.008	0.0	0.456	0.485

a) Under the nonparametric assumption, b) Null hypothesis : true area = 0.5, c) Made by Korea Forest Service

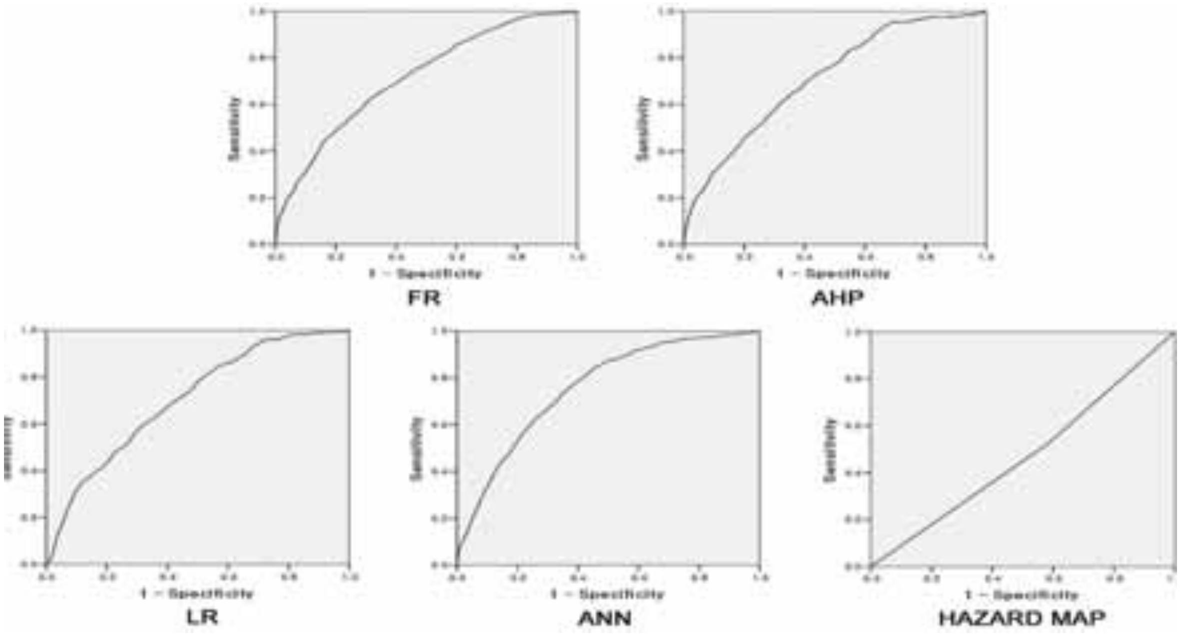

Fig. 9. AUC representing quality of the models used

5.3. Maximum matching rate analysis in landslide zone

In this study, to validate the accuracy of LSI maps built by FR, AHP, LR and ANN approaches and the landslide hazard map built by the Korea Forest Service, net accumulative values and inverse accumulative values were calculated as landslide grades of the study area. Crossing points were analyzed by cross applying net accumulative values and inverse accumulative values. Crossing points are mostly related to the landslide occurrence or non-occurrence. Areas that have higher values than those of average points were set as landslide occurrence areas (event : 1), while those that have lower values as non-occurrence areas (event : 0). Then, each model that has previously calculated landslide occurrence event values, landslide hazard maps and LSI maps built in this study were analyzed using matrix analysis of ERDAS 9.3.

Analysis results showed that maximum matching rates of FR, AHP, LR, ANN and LSI maps were similar in general, showing ≥66.34% on the average. In regard of matching rate of each model, ANN, LR, FR and AHP showed 69.77%, 65.51%, 65.50% and 64.57%. On the other hand, landslide hazard map showed 48.77% of matching rate, which was lower than that of other models(Table 7).

Table 7. Maximum matching rate

		Prediction			
		0[a)]	1[b)]	0	1
LOG : 65.51%	0	96286	50603	64.93	34.13
	1	536	860	0.36	0.58
FR : 65.50%	0	96198	50652	64.89	34.17
	1	488	907	0.33	0.61
AHP : 64.57%	0	94859	52030	63.97	35.09
	1	501	895	0.34	0.60
ANN : 69.77%	0	114306	49462	69.17	29.93
	1	487	995	0.29	0.60
HAZARD MAP[c)] : 48.77%	0	71549	92195	48.26	62.19
	1	734	748	0.50	0.50

a) no Landslide, b) yes Landslide, c) made by Korea forest service

5.4 Correlation analysis each model

To see how much similarity exists between LSI maps built using FR, AHP, LR and ANN methods, Pearson's correlation analysis was conducted. Correlation coefficient is between -1 and 1. If the coefficient is close to -1 or 1, it means that there is a strong correlation between 2 models. On the other hand, if the coefficient is close to 0, it means that there is little or no correlation between 2 models. General interpretation criteria of the correlation coefficient based on its absolute value is as follows: 0~0.2, little correlation; 0.3~0.6, correlation exists; ≥0.7, strong correlation.

Analysis results showed that there was a high correlation between LR and ANN methods exhibiting the correlation coefficient of 0.829. Correlation coefficients between AHP and ANN methods, AHP and LR methods, FR and LR methods and FR and ANN methods were 0.821, 0.812, 0.763 and 0.693, respectively. The lowest coefficient, 0.619 was found between AHP and FR methods, showing that there is the lowest correlation between them(Table 8).

Table 9. The compare of correlation coefficient

	LOG	AHP	FR	ANN
LOG	1	0.812	0.763	0.829
AHP	0.812	1	0.619	0.821
FR	0.763	0.619	1	0.693
ANN	0.829	0.821	0.693	1

6. Conclusion

Landslides occur frequently in the Republic of Korea due to its slopes scattered in the mountain areas and torrential rain in summer season. The Korean government is always busy trying to recover from the damage after the event, although the landslides cause casualties and property damage every year. Landslides can be prevented, however, if the landslide occurrence possibility is evaluated and predicted and appropriate measures are taken beforehand (Lee et al, 2000).

Accordingly, in this study, landslide susceptibility was analyzed for Inje area located in Gangwon-do to predict landslide occurrence possibility using FR, AHP, LR and ANN methods, and LSI(Landslide Susceptibility Index) maps were built. For these purposes, landslide occurrence-related factors, such as topographical factors(elevation, slope and aspect), hydrological factors(drainage distance, SPI and TWI), soil factors(texture and depth), forest factors(type, diameter, age and density) and land cover factors, were considered as input data.

AUC analysis was conducted for the LSI maps built by each model using SPSS 13.0 to analyze the accuracy in the models. Analysis results show that

Validation with AUC values resulted that as AUC was 0.7~0.8 which are normal values, AUC values were recalculated for the whole study area to evaluate the prediction accuracy quantitatively. As a result, the AUC values for FR, AHP, LR and ANN methods were 0.715, 0.712, 0.705 and 0.757, respectively. All the models showed AUC values from 0.705 to 0.757, which are slightly lower than the accuracy in the models, i.e., 0.794~0.806. The AUC values of the landslide hazard maps for the study area provided by the Korea Forest Service was also very low, i.e., 0.471, meaning low accuracy.

In addition, the LSI values of landslide occurrence areas and non-occurrence areas for the study area were cross applied calculating net accumulative values and inverse accumulative values. Through this, accuracy of LSI maps built in this study and landslide hazard maps was analyzed. As a result, maximum matching rates of FR, AHP, LR, ANN and LSI maps were similar. Maximum matching rate of ANN, LR, FR and AHP showed 69.77%, 65.51%, 65.50% and 64.57%. On the other hand, landslide hazard map showed 48.77% of matching rate, which was lower than that of other models.

Pearson's correlation analysis was conducted to see the correlation between each model. Analysis results showed that there was a high correlation between LSI maps using LR and ANN methods exhibiting the highest correlation coefficient of 0.829. The lowest coefficient, 0.619 was found between AHP and FR methods.

As examined above, the LSI map models built using FR, AHP, LR and ANN methods are not supposed to be significantly different as they have similar accuracy and high correlation. However, each model has the following advantages and disadvantages. FR method can simply and rapidly be applied, whereas LR method needs data conversion to be read by the statistical software program and has a limitation in calculating in the program when the data is massive. Also, AHP method generally needs questionnaire survey to calculate relative weights and involves use of separate software program, called Idrisi GIS. In this study, calculation time could be saved by calculating weights conducting logistic regression with statistical software program without questionnaire survey. ANN method is time consuming and requires a high computer capacity in actual application, because internal calculation process cannot easily understood and calculation is massive (Lee et al., 2000). For this reason, it is recommended to select an appropriate method that meets purpose of the study under the judgment of the researchers.

In addition, as the landslide hazard maps established by the Korea Forest Service showed a low accuracy, further revision and complementary works are needed to be used as useful data. The LSI maps built in this study are supposed to be used as useful data to establish effective forest management and national territory management system.

7. References

Bantayan, N.C. and Bishop I.D., 1998, Linking objective and subjective modeling for landuse decision-making. Landscape and Urban Planning, 43, 35-48.

Beven, K.J., Kirkby, M. J., 1979, A phtsically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43-69.

Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists, modeling with GIS. Pergamon Press, Oxford, 398 p.

Dai, F.C., Lee, C.F. and Zhang, X.H., 2001, GIS-based geo-environmental evaluation for urban land-use planning : a case study. Engineering Geology, 61, 257-271.

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

Lee, S. and Dan N.T., 2005, Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam : focus on the relationship between tectonic fractures and landslide. Environmental Geology, 48, 778-787.

Lee, S. and Pradhan, B., 2007, Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides, 4, 33-41.

Lee, S. and Talilb, J.A., 2005, Probabilistic landslide susceptibility and factor effect analysis. Environmental Geology, 47, 982-990.

Lee, S., Kim, Y.J. and Min, K.D., 2000, Development of spatial landslide information system and application of spatial landslide information. The Journal of GIS association of Korea, 8, 141-153.

McFadden, D., 1973, Conditional logit analysis of quantitative choice behavior, in P. Zarembka, ed., Frontiers in Econometrics, New York : Academic Press, 105-142 p.

Moore, I.D.,Grayson,R.B.,Ladson,A.R.,1991. Digital terrain modeling a review of hydro-hydrological, geomorphological, and biological applications. HydrologicalProcesses5,3–30.

Satty, T.L., 1977, A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15, 234-281.

Satty, T.L., 1990, How to make a decision : the analytic hierarchy process. International Journal of Services Sciences, 48, 9-26.

Yang, I.T., Chun, K.S. and Park, J.H., 2006, The effect of landslide factor and determination of landslide vulnerable area using GIS and AHP. The Korean Society for Geo-Spatial Information System, 14, 3-12.

Yesilnacar, E. and Topal, T., 2005, Landslide susceptibility mapping : a comparison of logistic regression and neural networks methods in a medium scale study Hendek region(Turkey). Engineering Geology, 79, 251-266.

Yilmaz, I., 2009, Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison : A case study from Kat landslides(Tokat-Turkey). Computers & Geosciences, 35, 1125-1138.

Yoon, J.M., 2002, A study on the urban growth using fuzzy-AHP and cellular automata, Korea. Ph. D. thesis, Pusan national university, Pusan, 21 p.