

## Research Article

## Open Access

Lee Saro, Jeon Seong Woo, Oh Kwan-Young, and Lee Moun-Jin\*

# The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea

DOI 10.1515/geo-2016-0010

Received October 22, 2014; accepted April 25, 2015

**Abstract:** The aim of this study is to predict landslide susceptibility caused using the spatial analysis by the application of a statistical methodology based on the GIS. Logistic regression models along with artificial neural network were applied and validated to analyze landslide susceptibility in Inje, Korea. Landslide occurrence area in the study were identified based on interpretations of optical remote sensing data (Aerial photographs) followed by field surveys. A spatial database considering forest, geophysical, soil and topographic data, was built on the study area using the Geographical Information System (GIS). These factors were analysed using artificial neural network (ANN) and logistic regression models to generate a landslide susceptibility map. The study validates the landslide susceptibility map by comparing them with landslide occurrence areas. The locations of landslide occurrence were divided randomly into a training set (50%) and a test set (50%). A training set analyse the landslide susceptibility map using the artificial network along with logistic regression models, and a test set was retained to validate the prediction map. The validation results revealed that the artificial neural network model (with an accuracy of 80.10%) was better at predicting landslides than the logistic regression model (with an accuracy of 77.05%). Of the weights used in the artificial neural network model, 'slope' yielded the highest weight value (1.330), and 'aspect' yielded the lowest value (1.000). This research applied two statistical analysis methods in a GIS and compared their results. Based on the findings, we were able to derive a more effective method for analyzing landslide susceptibility.

**Keywords:** Landslide; GIS; artificial neural network; logistic regression; susceptibility map; Korea

**Lee Saro:** Geological Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM) 124 Gwahang-no, Yuseong-gu, Daejeon, 34132, Korea

**Lee Saro:** Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon, 02841 Korea

© 2016 L. Saro *et al.*, published by De Gruyter Open.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License. The article is published with open access at [www.degruyter.com](http://www.degruyter.com).

## 1 Introduction

Landslide largely occurs in hilly or mountainous areas. In most cases of landslide, spots with shattered zones caused by crustal movements are affected by frequent concentrated torrential heavy rainfall. In case of Korea, the heavy rainy season, usually between June to September, is the period when landslides occur. Korea recently experiences frequent landslide. Quite recently it is noted that landslide in 2001, 2002, 2006 and 2010 increased landslide frequency. As one of major natural geological hazards, landslides cause significant damages in people and property. Due to landslides caused by heavy rainfall, much damage was caused in Inje. As there was little effort to predict them and to assess the consequences of such events, the damage was extensive. Through scientific analysis identify and assess landslide-susceptible areas and, by taking appropriate preparatory measures.

Using GIS as the basic analysis tool to map landslide susceptibility can be effective for manipulation of spatial data and management, together with particular equitable models for the analysis. Recently, Research on landslide susceptibility evaluation using GIS has been conducted, and many of the studies used probabilistic models. One of the statistical models available, an artificial neural network and logistic regression models, have also been applied to landslide susceptibility mapping. There are other methods for susceptibility mapping, such as the safety factor models and the geotechnical model.

Several methods have been suggested to assess landslide susceptibility and increasingly these use various

**Jeon Seong Woo:** Division of Environmental Science & Ecological Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea

**Oh Kwan-Young:** Department of Geoinformatics, University of Seoul, Siripdae-gil 13, Dongdaemun-gu, Seoul, 02504, Korea

**Oh Kwan-Young:** Korea Environment Institute, 11F, Bldg B, 370 Sicheong-daero, Sejong, 30147, Korea

**\*Corresponding Author: Lee Moun-Jin:** Korea Environment Institute, 10F, Bldg B, 370 Sicheong-daero, Sejong, 30147, Korea, E-mail: [leemj@kei.re.kr](mailto:leemj@kei.re.kr)

other models along with geographic information systems (GIS). Previous studies have applied probabilistic models including an AHP: Analytic Hierarchy Process, Artificial Neural Network, Dempster-Shapfer theory of evidence, fuzzy logic and Monte Carlo methods [1–11] among statistical models, the logistic regression model has also been applied to landslide susceptibility mapping [12–27]. More sophisticated assessments have involved weight of evidence approaches and frequency ratio [25–38]. Research on rainfall probability calculation has primarily been limited to improving the rainfall probability predictions accuracy and to studies targeting water resources [39–43]. Recently, analysis of landslide is used in the high-resolution airborne laser scanning (LiDAR) and soil moisture content changes [44–47].

The study area, Deokjeok-ri around Inje, Kangwon-do was selected where landslide counts were highest in the damaged areas, where landslide occurred on July 14, 2006. The areas of landslide occurrence were detected in the study area by interpretation aerial photographs and field surveys (Figure 1 and Figure 2). A flowchart outlining the methodology is shown in Figure 3. For applying and validating landslide susceptibility models, forest, land use, soil and topography spatial databases were built on the analysis. From the databases, 15 factors were selected. Using the calculated factors and detected landslide locations, two landslide analysis models, an artificial neural networks and logistic regression, were implemented. Using the calculated factors and detected landslide locations, two landslide analysis models, an artificial neural networks and logistic regression, were implemented. For the application, artificial and statistical neural network program and GIS program were used. Lastly, validation for quantitative, prediction curve methods such as success and prediction rates were used to verify the results.

## 2 Spatial data sets

To analyses probabilistic landslide susceptibility, accurate detection the area of landslide occurrence is very important. This research has analyzed aerial photographs, one kind of optical remote sensing data, for landslide occurring location. In addition, the result has been validated through field survey; because field survey is must necessary for remote sensing data. In order to obtain significant and cost-effective landslide information, the application of remote sensing methods including aerial photographs and field surveys is used. In this study, 1:10,000–1:50,000-scale aerial photographs, which were taken in 2009, were ex-

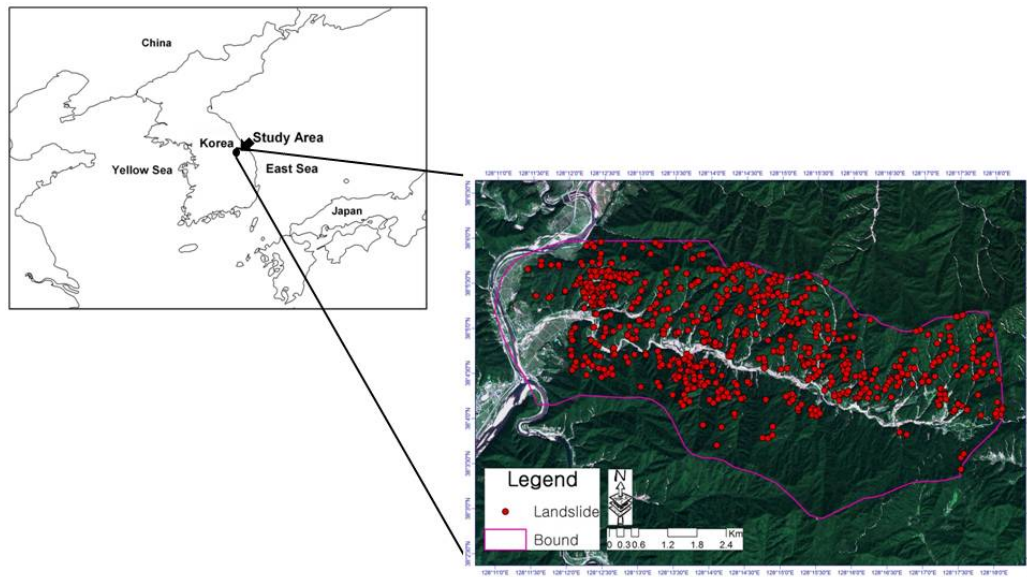
amined to identify landslide occurrence areas, which were verifiable by fieldwork. Recent landslides were detected in the aerial photographs from breaks in the bare soil, forest canopy, or geomorphic characteristics typical of landslide scars, such as flow tracks, head and side scarps, and soil and debris deposits below a scar. To assemble database, it is to assess the surface area and number of landslides in the study area, 693 landslides were mapped in the 37km<sup>2</sup> study area (Figure 1). The research area is not large area, so only one weather station is located on this area. The research supposes that effect of groundwater and rainfall amount of landslide occurring period is same on the study area. Thus, the research has concentrated on indexes for locational difference than effects of rainfall and groundwater.

Maps of landslide occurrence were built on vector format spatial database through the ArcGIS software. The map included 1:50,000 scale geological map, 1:25,000 scale soil maps, and 1:25,000 scale forest maps. Contour and survey base points elevation value from topographic map were a Digital Elevation Model (DEM). The DEM with a 10 m resolution, was used to aspect, curvature, slope, stream power index (SPI) and topographic wetness index (TWI). The DEM data are utilized in replace of the altitude. Soil database includes soil drainage, material, texture, thickness and topography. Forest database include timber age, density, diameter and type. The geology was extracted from the geological database. The spatial database constructed in this study is shown in Table 1.

To calculate the probability of landslides, 15 factors relevant to landslides were considered. Aspect, curvature, slope steepness, SPI and TWI, soil drainage, soil material, soil texture, soil thickness, timber age, timber density, timber diameter, timber type and geology were extracted from the spatial database as factors contributing to potential landslides. Also, Altitude data is not used because it is not according to the landslide high correlation.

When locating landslide points in the study areas during the 2006 rainy season, aerial photos before and after landslide are first compared, and then field surveys are conducted to assess the accuracy of the detected locations. The aerial photogrammetry before landslide is that issued by the National Geographic Information Institute (NGII) (1/20,000; taken on 2005.04.04), and that after landslide is that presented by Daum portal ([www.daum.net](http://www.daum.net)) (50cm; taken on 2008.03.04). As seen in Table 1, the maps of landslide distribution, topography, geology, forest and soil are collected as data for analyzing landslide susceptibility in Inje, the study area.

As for topographic map, the DEM is prepared from data transformation and interpolation of the digital topo-



**Figure 1:** Study area map and landslide location map.



**Figure 2:** Photographs of landslide in the study area (Above : Near road in the field).

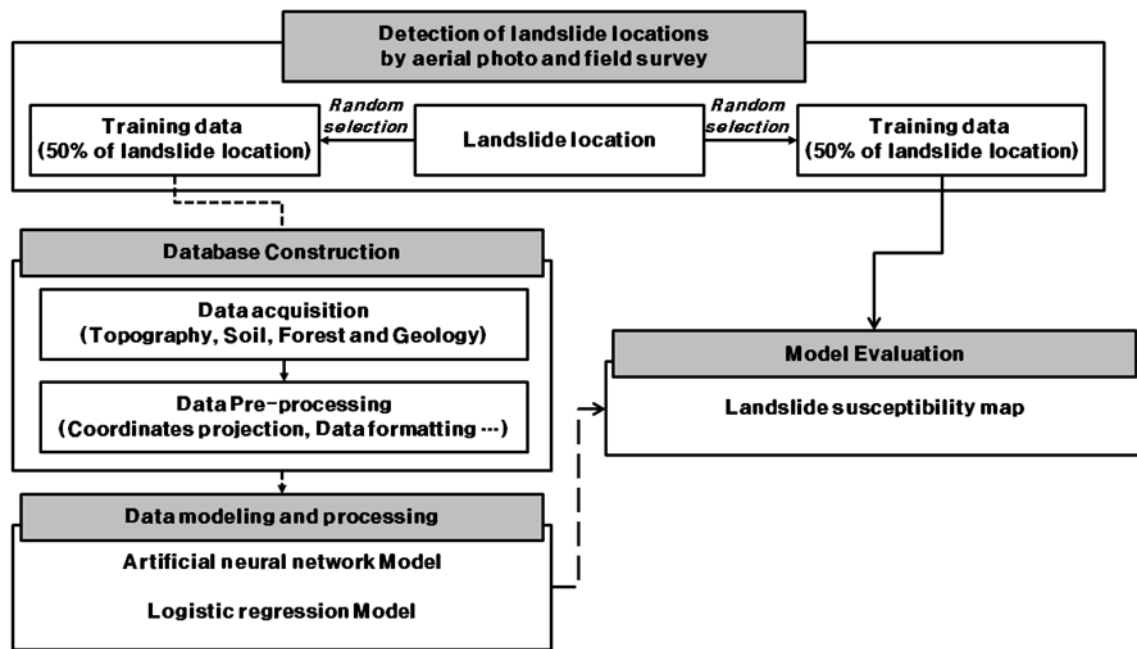


Figure 3: Overall methodology flow chart.

Table 1: Data layer of study area.

Classification	Factors	Data Type	Scale	Data Source
Geological Hazard	Landslide	Point	1:5,000	Result of Research
Topographic Map	Slope	Grid	1:5,000	National Geographic Information Institute
	Aspect			
	Curvature			
	TWI(Topographic Wetness Index)			
	SPI(Stream Power Index)			
Geological Map	Geology	Polygon	1:50,000	Korea Institute of Geoscience and Mineral Resources
Forest Map	Timber diameter	Polygon	1:25,000	Korea Forest Research Institute
	Timber type			
	Timber density			
	Timber age			
Soil Map	Topography	Polygon	1:25,000	National Academy of Agricultural Science
	Soil drainage			
	Soil material			
	Soil thickness			
	Soil texture			



graphical map of 1:5,000, and from the DEM. The slope indicates geographical angle, and the aspect indicates direction of geographical slope. The curvature is based on the 0 value, with slope concavity increasing as the value becomes negative, and becoming more convex with positive values. The value 0 indicates the slope is flat. As for soil factor, the soil topography, the soil texture, the soil drainage, the soil material, and the soil effective thickness were prepared from the detailed soil map (1:25,000). The topography is classified by the topography of the areas where soils are distributed, the soil texture indicates soil particle size, parent material refers to the rock floor from which the soils are formed, the drainage refers to water draining, and the effective soil thickness indicates the effective thickness of soil. Also, the timber type factor, the timber diameter, the timber age, and the timber density of forest are prepared from the forest map (1:25,000). Geological factor is from the geological map (1:250,000). The data with different formats, accuracies, scales and geometrics were almost not unusable to directly implement to this research. Due to this reason, these data were modified into applicable data input for analysis and artificial neural network along with logistic regression modeling (Figure 3).

Application of frequency ratio is known as to identifying correlations by grade between the landslide locations distributions and landslide relevant factors. Also it is to calculate landslide susceptibility indexes from the frequency ratio per grade of each factor to predict locations with landslide susceptibility of landslides. Table 2 is verifying the cause of using these factors.

### 3 Methodology

#### 3.1 Artificial neural network model

This study frequently used an artificial training algorithm as neural network method. By using a set of examples, associated values of input and output the artificial neural network model is trained. Until targeted minimal error is found between the desired and actual output values of the network, the back-propagation algorithm trains the network. Once the training is complete, the network is used as a feed-forward structure to produce the entire data classification [43].

The weight between layers was acquired by training the neural network, so the importance of each factor or contribution can be calculated. A GIS spatial database was used as landslide locations and input data were used as training regions. Of the various artificial neural net-

work methods, the back-propagation method was used. The program [47], using MATLAB, was partly converted for landslide analysis by adapting the input and output routines for the GIS data use.

The training sites were chosen from the landslide-relevant factors. The back-propagation algorithm was implemented to measure weights between the input layer, the hidden layer and output layer by modifying the number of hidden layers and the learning rate. The weights were enforced in the entire study. The calculated index values were transformed into an ARC/INFO GRID by applying the GIS. Also the landslide susceptibility map was built by using the GRID data.

A three-layered feed-forward network was applied to MATLAB on the base of the framework presented by [47]. The back-propagation training algorithm is trained using a set of examples of associated input and output values. The hidden and output layer neurons process their inputs by multiplying each input by a corresponding weight, summing the product, and then processing the sum using a nonlinear transfer function to produce a result. The training in this study is a task whereby sites where landslides have and have not occurred are clearly perceived by the neural network, and then the neural network can calculate the results on the output, or the weights of landslide occurrence. According to [43], selecting representative sites is more important than counting the samples in training samples. Therefore, only 50% of the total cells of landslide sites are extracted to be landslide sites for training, and the values of the 15 landslide-related factors are regulated to be between 0.1 and 0.9, for the sigmoid function, used for improving inference in the neural network, which has a value between 0 and 1. Landslide susceptibility expected by the back-propagation algorithm is 0.9; weights were determined through repetitive back-propagation algorithm training in order to reduce the error between expected output and actual output to 0.1 (Table 3). To calculate weights, the neural network structure is established to be  $15(\text{input layer}) \times 32(\text{hidden layer}) \times 2(\text{output layer})$ , while the relative weights of the factors are calculated as the maximum repetitive number before reaching the targeted error of 2,000 and the learning rate of 0.01. As the calculated weights are granted to each factor, landslide susceptibility for the whole study area is prepared.

#### 3.2 Logistic regression model

Logistic regression allows for multivariate regression relations investigation between several independent variables and one.

**Table 2:** Frequency ratio between landslide and related factors.

	Class	No. of land- slide	% of land- slide	No. of pixels in do- main	% of pixels in do- main	Frequency ra- tio
Aspect	Flat	0	0	12418	0.85	0
	North	47	6.78	178128	12.14	0.56
	Northeast	38	5.48	144479	9.85	0.56
	East	114	16.45	129251	8.81	1.87
	Southeast	166	23.95	169033	11.52	2.08
	South	86	12.41	170758	11.64	1.07
	Southwest	86	12.41	196686	13.41	0.93
	West	83	11.98	231150	15.76	0.76
	Northwest	73	10.53	235075	16.02	0.66
Curvature	Concave	279	40.26	505504	34.46	1.17
	Flat	138	19.91	464390	31.66	0.63
	Convex	276	39.83	497084	33.88	1.18
Slope (degree)	0 5°	3	0.43	65606	4.47	0.1
	6 10°	4	0.58	84813	5.78	0.1
	11 15°	15	2.16	123479	8.42	0.26
	16 20°	23	3.32	166365	11.34	0.29
	21 25°	76	10.97	235666	16.06	0.68
	26 30°	113	16.31	255125	17.39	0.94
	31 35°	200	28.86	239647	16.34	1.77
	36 40°	129	18.61	171456	11.69	1.59
	41 90°	130	18.76	124821	8.51	2.2
TWI (Topographic Wetness Index)	1	141	20.35	147220	10.04	2.03
	2	112	16.16	146647	10	1.62
	3	122	17.6	147249	10.04	1.75
	4	92	13.28	146696	10	1.33
	5	83	11.98	147288	10.04	1.19
	6	59	8.51	146537	9.99	0.85
	7	35	5.05	146873	10.01	0.5
	8	27	3.9	146217	9.97	0.39
	9	10	1.44	146182	9.96	0.14
	10	12	1.73	146069	9.96	0.17
SPI (Stream Power Index)	1	61	8.8	174124	11.87	0.74
	2	42	6.06	143919	9.81	0.62
	3	68	9.81	143629	9.79	1
	4	77	11.11	144084	9.82	1.13
	5	75	10.82	143808	9.8	1.1
	6	93	13.42	143968	9.81	1.37
	7	79	11.4	143793	9.8	1.16
	8	88	12.7	143425	9.78	1.3
	9	55	7.94	143227	9.76	0.81
	10	55	7.94	143001	9.75	0.81
Geology	Banded gneiss	692	99.86	1375722	93.78	1.06
	Granite	1	0.14	91256	6.22	0.02
Timber diameter	Non forest area	231	33.33	605267	41.26	0.81
	Very small diameter	357	51.52	604447	41.2	1.25
	Small diameter	102	14.72	241453	16.46	0.89
	Medium diameter	3	0.43	15809	1.08	0.4

Table 2: Cont.

	Class	No. of land- slide	% of land- slide	No. of pixels in do- main	% of pixels in do- main	Frequency ra- tio
Timber density	Non forest area	231	33.33	605267	41.26	0.81
	Loose	357	51.52	604447	41.2	1.25
	Moderate	102	14.72	241453	16.46	0.89
	Dense	3	0.43	15809	1.08	0.4
Timber type	Non forest area	9	1.3	166668	11.36	0.11
	Mixed broad-leaf tree	27	3.9	86092	5.87	0.66
	Pine	305	44.01	528932	36.06	1.22
	Needle and broad	81	11.69	126759	8.64	1.35
	Artificial pine	26	3.75	17463	1.19	3.15
	Rigida pine	4	0.58	25395	1.73	0.33
	Korea nut pine	170	24.53	310110	21.14	1.16
	Artificial Larch	68	9.81	197038	13.43	0.73
	Larch	0	0	1569	0.11	0
	Artificial mixed broad-leaf	0	0	2849	0.19	0
	Poplat	3	0.43	4101	0.28	1.55
Timber age	Non forest area	13	1.88	193632	13.2	0.14
	1st age	218	31.46	411635	28.06	1.12
	2nd age	120	17.32	237161	16.17	1.07
	3rd age	92	13.28	175273	11.95	1.11
	4th age	246	35.5	406194	27.69	1.28
	5th age	4	0.58	43081	2.94	0.2
Soil drainage	No Data	0	0	4084	0.28	0
	Well drained	33	4.76	140333	9.57	0.5
	Somewhat poorly dray	403	58.15	520168	35.46	1.64
	Moderately well dray	257	37.09	802393	54.7	0.68
Soil thickness (cm)	No Data	0	0	4084	0.28	0
	20	254	36.65	743545	50.69	0.72
	50	26	3.75	76714	5.23	0.72
	100	406	58.59	573763	39.11	1.5
	150	7	1.01	68872	4.69	0.22
Soil material	No data	0	0	4084	0.28	0
	Valley alluvium	4	0.58	66756	4.55	0.13
	Gneiss residuum	17	2.45	62787	4.28	0.57
	Fluvial alluvium	0	0	2327	0.16	0
	Colluvium	511	73.74	784732	53.49	1.38
	Alluvial colluvium	161	23.23	546292	37.24	0.62
Soil texture	No data	0	0	3640	0.25	0
	Sandy loam	0	0	33337	2.27	0
	Rocky loam	404	58.3	540093	36.82	1.58
	Loam	18	2.6	68905	4.7	0.55
	Silt loam	3	0.43	29545	2.01	0.21
	Rocky sandy loam	0	0	4152	0.28	0
	Very rocky loam	7	1.01	54059	3.69	0.27
	Overflow area	261	37.66	733247	49.98	0.75

Table 2: Cont.

	Class	No. of land- slide	% of land- slide	No. of pixels in do- main	% of pixels in do- main	Frequency ra- tio
Topography	No data	0	0	4084	0.28	0
	Valley area	6	0.87	53673	3.66	0.24
	Valley and alluvial	0	0	36886	2.51	0
	Plains	0	0	1658	0.11	0
	Piedmont slope area	684	98.7	1345227	91.7	1.08
	Lower hilly area an	2	0.29	13781	0.94	0.31
	Fluvial plains	0	0	268	0.02	0
	Alluvial fan	1	0.14	11401	0.78	0.19

In the present situation, the dependent variable is binary, representing the presence or absence of landslides. Quantitatively, the relationship between the occurrence and its the dependency on several variables can be expressed as below:

$$p = 1/(1 + e^{-z}) \text{ or } p = ez/(1 + ez). \quad (1)$$

$P$  is an event occurring probability and  $e$  is the natural logarithm. In the present,  $p$  is an estimated landslide probability based on the intrinsic properties only, which is known as “susceptibility” in this context. The probability varies from 0 to 1 on an S-shaped curve and,  $z$  is the linear combination. It follows that logistic regression which involves fitting the data to an equation of the form

$$z = b_0x_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n. \quad (2)$$

While  $b_0$  is the model intercept,  $b_i$  ( $i = 0, 1, 2, \dots, n$ ) represents the slope coefficients of the logistic regression model, and  $x_i$  ( $i = 0, 1, 2, \dots, n$ ) are independent variables [48]. The linear model is then a logistic regression for the presence or landslides absence (present conditions) on the independent variables (pre-failure conditions).

Using the formulae, a landslide susceptibility map was built. The logistic regression analysis was presented by dividing the study area into grid squares of 5 m by 5 m. the 15 factors data were converted to an ASCII format for statistical package use.

The decision process for logistic multiple regression is, as with all multivariate application, setting the objectives is the first step in the analysis. The analysis proceeds with the derivation of the logistic function and the determination of whether a statistically significant function can be derived to separate the two groups. The logistic multiple regression results are then assessed for predictive accuracy by developing a classification matrix. Next, interpretation of the discriminant function determines which

of the independent variables contributes the most to discriminating between the groups. Finally, the logistic function should be validated with a holdout sample [16].

## 4 Result

### 4.1 Prediction of landslide susceptibility

The final weights between layers acquired during the artificial neural network training and the contribution or importance of each of the 15 factors to predict are shown in Table 3. Because the initial weights were assigned random values, the results were not the same. This research calculates ten times to allow the results to achieve similar values. 0.004 to 0.015 is the ranges of the Standard deviation (SD); so, there is not much effect on the results from random sampling. For easy interpretation, the average values were calculated, and the values were divided by the average of the weights of the factor with the minimum value. Among the weights, the slope presented the highest weight index, 1.330. Negative curvature values steep slope were more susceptible to landslides. Soil drainage (1.177) is the second major parameter contributing to landslide occurrence. Soil material (1.131) is the third important parameter contributing to landslide occurrence. The weights analysis shows that the less important parameters are soil thickness (1.022), soil texture (1.024) and timber type (1.053). The results show the most important factor is slope. The Gaussian nature of distribution of susceptibility zones statistically regarding the landslide areas approves the applicability of artificial neural network to landslide susceptibility mapping in the study (Figure 4). Based on the result, slope is the most important index for choosing priority order for managing continuous landslide, along with



restoration and prevention, on the research. Furthermore, index should be managed with its index weight.

By using this method, the significance (Sig.) and the logistic multiple regression coefficients (B) of related variables were calculated (Table 4). By applying the maximum-likelihood model, the coefficients were estimated. Because the relationship between independent variables and the probability was nonlinear in the logistic multiple regression model, parameter estimation [30] needs an iterative algorithm. Most values at a significance (Sig.) level) were less than 0.05, which means the factor is influential in landslide occurrence. 0.278 was the output value corresponding to the Hosmer and Lemeshow goodness-of-fit test [49].

By using the maximum-likelihood method, the coefficients (B) were estimated, and the coefficients that made the observed results most 'likely' were selected. Because the relationship between the independent variables and probability was nonlinear in the logistic regression model, for parameter estimation [41], an iterative algorithm was used. If a coefficient was positive, its transformed log value was greater than 1, meaning that the modeled event is more likely to occur. If a coefficient was negative, its transformed log value was less than 1, and the event is less likely to occur. A coefficient of zero (0) had a transformed log value of 1.0, meaning that this coefficient does not affect the likelihood of the event. In the case of numerical data, positive associations were observed with slope, and negative associations were observed with TWI and SPI. For example, in the categorical data, the cases of 'No Data' and 'Valley alluvium' both yielded negative effects, whereas Gneiss residuum had a positive effect. After the interpretation, three equations were developed to predict the probability of landslide occurrence.

$$\begin{aligned}
 z = & (0.023 \times \text{Slope}) + (-0.005 \times \text{SPI}) + (-0.030 \times \text{TWI}) \\
 & + \text{Aspect} + \text{Curvature} + \text{Geology} + \text{Soil drainage} \\
 & + \text{Soil material} + \text{Soil texture} + \text{Soil thickness} \\
 & + \text{Timber diameter} + \text{Timber type} + \text{Timber density} \\
 & + \text{Timber age} + \text{Topography} - 57.710. \quad (3)
 \end{aligned}$$

A landslide susceptibility map was built based on formula above. The logistic regression analysis was used by dividing the study area into a 5 m × 5 m size grid. The factors were fitted to this and changed to an ASCII file to the statistical package use. By using the logistic regression coefficient (Table 4) and Equations (3), the landslide probability was calculated for the nine cases. As there was no coefficient was available for a certain class, the average value (i.e. unity) was used. The computed probability values were mapped to allow interpretation as illustrated in

Figure 5. The values were classified into equal areas and then grouped into five classes for visual.

As illustrated in Figure 4 and Figure (5), the computed susceptibility values were mapped to allow interpretation. To ease visual interpretation, the susceptibility values were classified into five classes (No Data: 0%, Very high: 100~88, High: 87~76, Medium: 75~58 and Low: 57~1) based on area. As increase value of the susceptibility, the landslide susceptibility increases; as well as a lower value indicates a lower susceptibility. Figure 4 is landslide susceptibility map considering artificial neural network and Figure 5 is landslide susceptibility map with logistic regression.

## 4.2 Validation

The produced prediction indexes are necessary for verification because of assessment values. In this study, the susceptibility map accuracy by each analysis method is verified using a Landslide Susceptibility Indices (LSI), which is expressed as a ratio value of the prediction index value from the prediction on an area of landslide per equal area. In this study validation, the calculated LSI values for all cells in the study were sorted in descending order. Then, the landslides (%) were divided into classes of the accumulated area ratio (%) according to the LSI value. In order to quantitatively compare the results, the areas under the curve (AUC) were calculated again with the total area of 1 indicating perfect prediction accuracy. In order to qualitatively assess the prediction accuracy, the AUC can be used.

In order to compare the quantitative result, the areas under the curve were calculated again as the total area is 100% which represents the accuracy of perfect prediction. Therefore the area under a curve can be used to qualitatively assess the accuracy of prediction. The results of the landslide susceptibility maps were validated using the landslide locations. For landslide location, remote sensing data has been validated through field survey. The total number of landslide locations is 694. Approximately 50% (347) of its total landslide locations were used for artificial neural network, logistic regression and susceptibility analysis. For validation, the remaining 50% (347) were used.

The success rate validation results, from comparing the susceptibility calculation results and landslide occurrence location using an artificial neural network and logistic regression models, are shown in appear in Figure 6. Although, neural network model validation result (80.10%) is a little better than the ones from the logistic regression model validation result (77.05%). Neural network model

**Table 3:** Artificial neural network weight between landslide and related factors.

	1	2	3	4	5	6	7	8	9	10	Average	S.D.	N.W.*
Aspect	0.069	0.068	0.068	0.071	0.065	0.089	0.066	0.07	0.081	0.065	0.071	0.008	1
Curvature	0.077	0.07	0.08	0.075	0.079	0.073	0.074	0.069	0.071	0.075	0.074	0.004	1.044
Slope	0.098	0.094	0.089	0.087	0.092	0.096	0.098	0.094	0.095	0.104	0.095	0.005	1.33
TWI (Topographic Wetness Index)	0.078	0.069	0.075	0.081	0.074	0.086	0.062	0.067	0.073	0.084	0.075	0.008	1.049
SPI (Stream Power Index)	0.076	0.077	0.077	0.07	0.079	0.069	0.064	0.086	0.089	0.071	0.076	0.008	1.067
Geology	0.083	0.066	0.07	0.068	0.074	0.075	0.072	0.09	0.082	0.086	0.077	0.008	1.076
Timber diameter	0.082	0.085	0.074	0.056	0.068	0.076	0.09	0.074	0.076	0.069	0.075	0.01	1.053
Timber density	0.075	0.103	0.085	0.084	0.09	0.059	0.07	0.082	0.056	0.091	0.08	0.015	1.117
Timber type	0.057	0.089	0.08	0.086	0.085	0.063	0.064	0.073	0.08	0.06	0.074	0.012	1.035
Timber age	0.081	0.072	0.074	0.069	0.078	0.078	0.057	0.082	0.062	0.088	0.074	0.009	1.041
Soil drainage	0.078	0.074	0.093	0.08	0.087	0.094	0.083	0.086	0.085	0.078	0.084	0.007	1.177
Soil thickness	0.074	0.071	0.07	0.072	0.073	0.074	0.081	0.075	0.07	0.068	0.073	0.004	1.022
Soil material	0.089	0.074	0.088	0.091	0.08	0.081	0.087	0.072	0.069	0.074	0.081	0.008	1.131
Soil texture	0.074	0.078	0.067	0.077	0.073	0.077	0.079	0.057	0.081	0.066	0.073	0.007	1.024
Topography	0.084	0.065	0.067	0.086	0.073	0.087	0.086	0.071	0.08	0.088	0.079	0.009	1.105
* Neural network Weight													

**Table 4:** Logistic regression coefficient between landslide and related factors.

Factor	Class	Logistic regression coefficient (B)	Significance level (Sig.)
Slope	-	0.023	0.128
TWI (Topographic Wetness Index)	-	-0.03	0.087
SPI (Stream Power Index)	-	-0.005	0.865
Aspect	Flat	-11.453	0
	North	0.205	
	Northeast	0.238	
	East	1.072	
	Southeast	0.585	
	South	-0.029	
	Southwest	-0.098	
	West	-0.173	
Curvature	Concave	0.072	0.489
	Flat	-0.125	
	Convex	0	
Geology	Banded gneiss	2.747	0.007
	Granite	0	
	Water	0	
Timber diameter	Non forest area	-11.138	0.016
	Very small diameter	1.141	
	Small diameter	-15.583	
	Medium diameter	0	
	Large	0	
Timber type	Non forest area	11.933	0.001
	Mixed broad-leaf tree	-0.897	
	Needle and broad	-0.136	
	Rigida pine	15.273	
	Pine	0.059	
	Artificial Larch	0.426	
	Larch	0	
	Korea nut pine	-0.004	
	Artificial pine	0.337	
Timber density	Non forest area	0	0.147
	Loose	0.228	
	Moderate	0.532	
	Dense	0	
Timber age	Non forest area	0	0.032
	1st age	0	
	2nd age	13.586	
	3rd age	15.836	
	4th age	0.906	
	5th age	0	
	6th age	0	
Soil drainage	No Data	-4.981	0.999
	Well drained	-0.927	
	Somewhat poorly dra	-0.649	
	Moderately well dra	0	

Table 4: Cont.

Factor	Class	Logistic regression coefficient (B)	Significance level (Sig.)
Soil material	No Data	-11.632	0
	Valley alluvium	-12.643	
	Gneiss residuum	2.902	
	Fluvial alluvium	0	
	Colluvium	0	
	Alluvial colluvium	0	
Soil thickness	No Data	0	0.99
	20m	-12.108	
	50m	-11.024	
	100m	-10.854	
Soil texture	No Data	-12.614	0.821
	Very rocky loam	0.045	
	Sandy loam	-19.546	
	Rocky loam	-0.562	
	Rocky sandy loam	-11.449	
	Overflow area	-8.361	
Topography	No Data	0	0.77
	Valley area	-0.627	
	Valley and alluvial	-10.245	
	Plains	-0.265	
	Piedmont slope area	9.587	
	Lower hilly area	11.07	
	Alluvial fan	0	

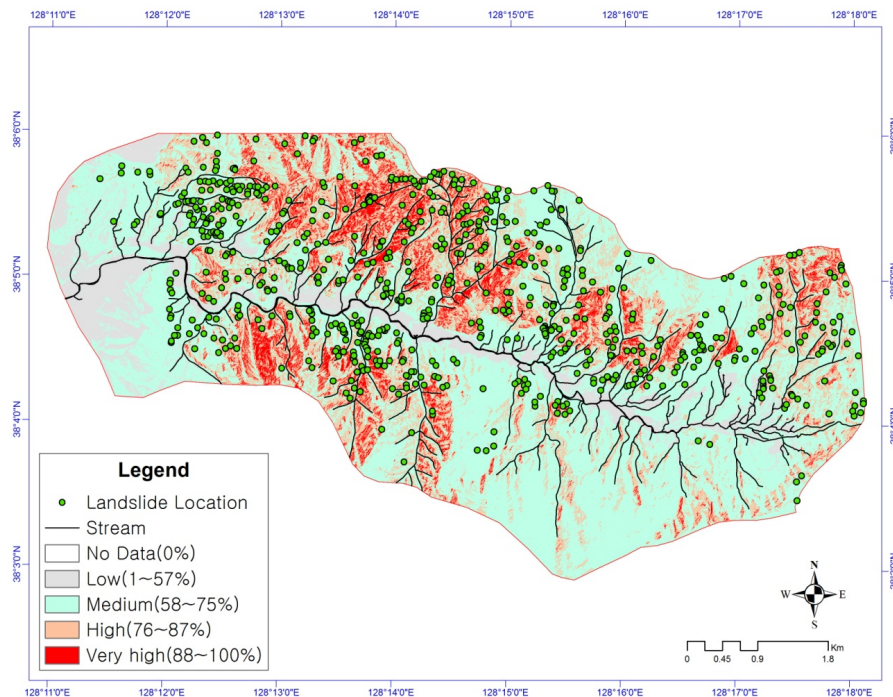
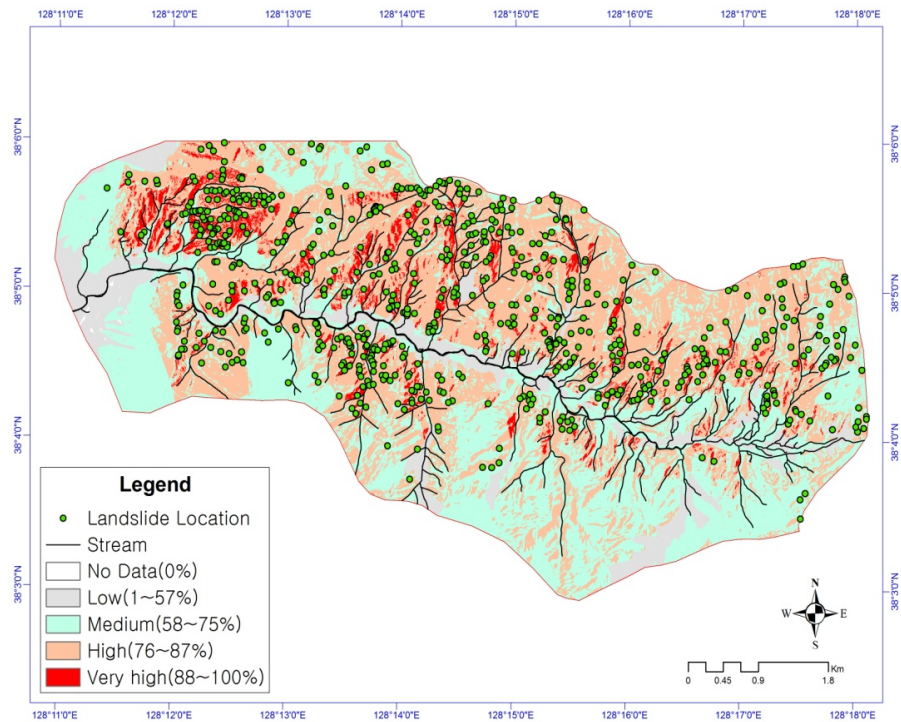


Figure 4: Landslide susceptibility map based on artificial neural network.



**Figure 5:** Landslide susceptibility map based on logistic regression.

success result is 84.9% and logistic regression model success result is 80.01%. In addition, a difference of validation and success result of neural network model is about 2.28%, and the difference between the results of validation and success of the logistic regression model is about 4.8%.

## 5 Discussion and conclusion

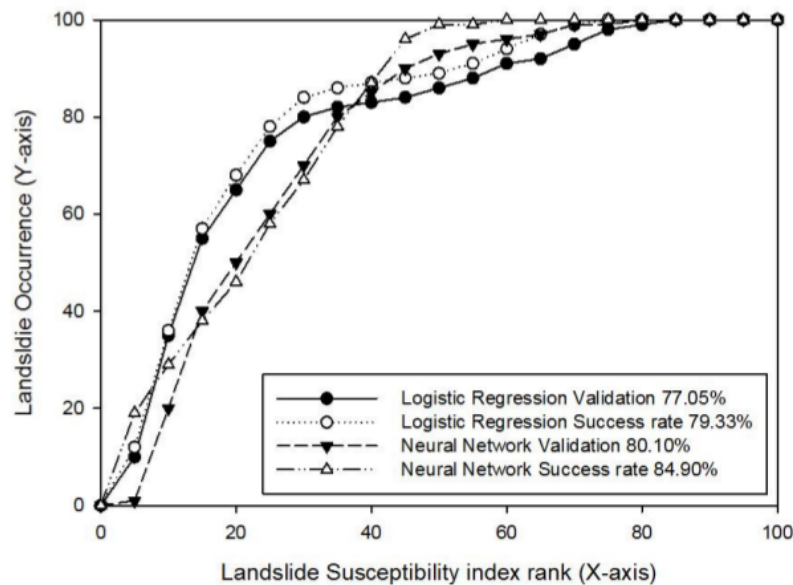
In the area of Inje, Korea, Landslide location was identified using aerial photographs and a landslide-related database was constructed for the landslide susceptibility analysis. Using 15 factors, artificial neural network and logistic regression models were applied and validated for the study using GIS. The relative importance and weight of factors were calculated through the artificial neural network use. The 'slope' showed the highest weight value (1.330), followed by the 'soil drainage' with a value of 1.177. The 'aspect' presented the lowest value at 1.000, and the 'soil thickness' was 1.023. The results show that the 'slope' was the most important factor as well as it was 1.4 times more important than 'aspect' in landslide susceptibility mapping. Additionally, logistic regression model got the output value (0.278) using Hosmer and Lemeshow Goodness-of-Fit test [42]. The output value more than 0.05 means that

the logistic regression model is valid. Thus the methodology of this study is using logistic regression model.

Next, landslide susceptibility maps were constructed using artificial neural network and logistic regression models. These maps revealed high levels of prediction accuracy: 80.10% and 77.05% for the artificial neural network and logistic regression models, respectively. Therefore, the artificial neural network model yielded more accurate results than the logistic regression model. Based on this validation result, the resulting susceptibility map is considered to be a satisfactory agreement between the computed results and the landslide inventory. Usually, the validation results revealed satisfactory agreement between the existing data and susceptibility map for landslide location.

The maps resulting from the use of the artificial neural network and logistic regression models had similar spatial distribution patterns. The middle north and southwest areas of the site were predicted to have respectively high and very high susceptibility. These areas have steep (sandy) loam soil, slopes, a thick soil layer, and are hilly or mountainous. The areas of high and very high susceptibility should be a priority concern during landslide-prevention planning. The site's middle north region was found to have low and very low potential in all the susceptibility maps. Almost all of these regions are categorized by low-lying land, river areas, silt loam soil, and a thin soil layer.





**Figure 6:** Illustration of cumulative frequency diagram showing landslide susceptibility index rank (x-axis) occurring in cumulative percent of landslide occurrence (y-axis) for susceptibility.

An artificial neural network model has more advantages than logistic regression model. An artificial neural network model is simple and the process of input, calculation, and output is easily understood. Moreover, the artificial neural network model can be calculated by factor's weights. The weighting given to the various factors that are important in our landslide-susceptibility analysis offers their relative significance ranking. Data is needed in the statistical package use for the logistic regression model, and later reconverted to incorporate it into the GIS database. Furthermore, large amounts of data cannot be processed by the statistical package easily and quickly. However, the degree of landslide susceptibility rating can be analyzed quantitatively. Using artificial neural network model, susceptibility can be qualitatively analyzed, and there are advantages, including continuous and discrete data processing, extraction of a good result for a complex problem and a multi-faceted approach to a solution.

This study holds significance by applying more than two statistical methodologies based on the GIS and conducting comparative analysis. However, there are also several limitations to this research. First, the use of multiple scales can undermine the accurate interpretation of data. Second, the categories of TWI and SPI generated from DEM can innately include errors in DEM. Lastly, despite the significance of the water flow and confluence in the occurrence of landslides, this study simply applied TWI and SPI. SM parameters (e.g. TVDI) [46, 47] for land drainage can improve the susceptibility assessment for landslides.

Landslides are known as one of the most hazardous natural disasters. So, government and research institutions have attempted to assess the landslide risk and hazard and to show its spatial distribution. The landslide susceptibility maps helps planners and engineers to select areas for further detail survey and locations for development. The results provide basic data to assist slope management and land use planning in the Inje area. The used methods are valid for assessment purposes and generalized planning, although the methods might be less useful at the site-specific scale where geographic diversity and local geological may prevail. For the models to be more generally applied, more landslide and more case studies conducted are needed.

**Acknowledgement:** This research were supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Minister of Science, ICT and Future Planning of Korea and Korea Environment Institute funded by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A1002704).

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: <http://www.textcheck.com/certificate/zwXCwg>

## References

- [1] Lee S., Ryu, J.H., Won J.S., Park H.J., Determination and application of the weights for landslide susceptibility mapping using an artificial neural network, *Eng. Geol.*, 2004, 71, 289-302.
- [2] Yalcin A., GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations, *Catena*, 2008, 72, 1-12.
- [3] Shou K., Chen Y., Liu H., Hazard analysis of Li-shan landslide in Taiwan, *Geomorphology*, 2009, 103, 143-153.
- [4] Tangestani M.H., A comparative study of Dempster-Shafer and fuzzy models for landslide susceptibility mapping using a GIS: An experience from Zagros Mountains, SW Iran, *J. Asian Earth Sci.*, 2009, 35, 66-73.
- [5] Yilmaz I., A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks, *Bull. Eng. Geol. Environ.*, 2009, 68, 297-306.
- [6] Pradhan B., Manifestation of an advanced fuzzy logic model coupled with Geo-information techniques to landslide susceptibility mapping and their comparison with logistic regression modelling, *Environ. Ecol. Stat.*, 2010, 18, 471-493.
- [7] Pradhan B., Lee S., Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia, *Landslides*, 2010, 7, 13-30.
- [8] Li Y., Chen G., Tang C., Zhou G., Zheng L., Rainfall and earthquake-induced landslide susceptibility assessment using GIS and Artificial Neural Network, *Nat. Hazards Earth Sys*, 2012, 12, 2719-2729.
- [9] Xu C., Xu X., Dai F., Saraf A.K., Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China, *Comput. Geosci.*, 2012, 46, 317-329.
- [10] Ramakrishnan D., Singh T.N., Verma A.K., Gulati A., Tiwari K.C., Soft computing and GIS for landslide susceptibility assessment in Tawaghat area, Kumaon Himalaya, India, *Nat. Hazards*, 2013, 65, 315-330.
- [11] Bui D.T., Pradhan B., Lofman O., Revhaug I., Dick O.B., Regional prediction of landslide hazard using probability analysis of intense rainfall in the HoaBinh province, Vietnam, *Nat. Hazards*, 2013, 66, 707-730.
- [12] Lee S., Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data, *Int. J. Remote Sens.*, 2005, 26, 1477-1491.
- [13] Lee S., Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea, *Earth Surf. Processes*, 2007, 32, 2133-2148.
- [14] Bai S., Lü G., Wang J., Zhou P., Ding L., GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang, China, *Environ. Earth Sci.*, 2010, 62, 139-149.
- [15] Nandi A., Shakoor A., A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses, *Eng. Geol.*, 2010, 110, 11-20.
- [16] Oh H.J., Lee S., Cross-validation of logistic regression model for landslide susceptibility mapping at Geneoung areas, Korea, *Disaster Advances*, 2010, 3, 44-55.
- [17] Pradhan, B., Lee S., Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models, *Environ. Earth Sci.*, 2010, 60, 1037-1054.
- [18] Yalcin A., Reis S., Aydinoglu A.C., Yomralioglu T., A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey, *Catena*, 2011, 85, 274-287.
- [19] Akgun A., Kincal C., Pradhan B., Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey), *Environ. Monit. Assess.*, 2012, 184, 5453-5470.
- [20] Bai S., Wang J., Zhang Z., Cheng C., Combined landslide susceptibility mapping after Wenchuan earthquake at the Zhouqu segment in the Bailongjiang Basin, China, *Catena*, 2012, 99, 18-25.
- [21] Dahal R.K., Hasegawa S., Bhandary N.P., Poudel P.P., Nonomura, A., Yatabe, R., A replication of landslide hazard mapping at catchment scale, *Geomatics, Nat. Hazards Risk J.*, 2012, 3, 161-192.
- [22] Lepore C., Kamal S.A., Shanahan P., Bras R.L., Rainfall-induced landslide susceptibility zonation of Puerto Rico, *Environ. Earth Sci.*, 2012, 66, 1667-1681.
- [23] Miller S., Degg M., Landslide susceptibility mapping in North-East Wales, *Geomatics, Nat. Hazards Risk J.*, 2012, 3, 133-159.
- [24] Devkota K.C., Regmi A.D., Pourghasemi H.R., Yoshida K., Pradhan B., Ryu I.C., Dhital M.R., Althuwaynee, O.F., Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya, *Nat. Hazards*, 2013, 65, 135-165.
- [25] Ayalew, L., Yamagishi, H., The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, *Geomorphology*, 2005, 65, 15-31.
- [26] Yilmaz I., Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine, *Environ. Earth Sci.*, 2010, 61, 821-836.
- [27] Akgun A., A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at Izmir, Turkey, *Landslide*, 2012, 9, 93-106.
- [28] Lee S., Min K., Statistical analysis of landslide susceptibility at Yongin, Korea, *Environ. Geol.*, 2001, 40, 1095-1113.
- [29] Lee S., Pradhan B., Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia, *J. Earth System Scim.*, 2006, 115, 661-672.
- [30] Oh H.J., Lee S., Chotikasathien, W., Kim, C., Kwon, J., Predictive landslide susceptibility mapping using spatial information in the Pechabun area of Thailand, *Environ. Geol.*, 2009, 57, 641-651.
- [31] Ozdemir A., Landslide susceptibility mapping of vicinity of Yaka Landslide (Gelendost, Turkey) using conditional probability approach in GIS, *Environ. Geol.*, 2009, 57, 1675-1686.
- [32] Vahidnia M.H., Alesheikh A.A., Alimohammadi A., Hosseinali F., Landslide Hazard Zonation Using Quantitative Methods in GIS, *Int. J. Civil Eng.*, 2009, 7, 176-189.
- [33] Regmi N.R., Giardino J.R., Vitek J.D., Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA, *Geomorphology*, 2010, 115, 172-187.

- [34] Yilmaz I., The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks, *Environ. Earth Sci.*, 2010, 60, 505-519.
- [35] Oh H.J., Park N.W., Lee S.S., Lee S., Extraction of landslide-related factors from ASTER imagery and its application to landslide susceptibility mapping, *Int. J. Remote Sens.*, 2012, 33, 3211-3231.
- [36] Pradhan B., Chaudhari A., Adinarayana J., Buchroithner M.F., Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: A case study at Penang Island, Malaysia, *Environ. Monit. Assess.*, 2012, 184, 715-727.
- [37] Choi J., Oh H.J., Lee H.J., Lee C., Lee S., Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS, *Eng. Geol.*, 2012, 124, 12-23.
- [38] Youssef A. M., Pradhan B., Jebur M. N., El-Harbi H. M., Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia, *Environ. Earth Sci.*, 2014, doi: 10.1007/s12665-014-3661-3
- [39] Marco B., Accuracy of radar rainfall estimates for stream flow simulation, *J. Hydrol.*, 2002, 267, 26-39.
- [40] Herr H.D., Krzysztofowicz R., Generic probability distribution of rainfall in space: the bivariate model, *J. Hydrol.*, 2005, 306, 234-263.
- [41] Cabus P., River flow prediction through rainfall-runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium, *Agr. Water Manag.*, 2008, 95, 859-868.
- [42] Jaiswal P., Van Westen C.J., Jetten V., Quantitative estimation of landslide risk from rapid debris slides on natural slopes in the Nilgiri hills, India, *Nat. Hazards. Earth System. Sci.*, 2011, 11, 1723-1743.
- [43] Paola J.D., Schowengerdt R.A., A review and analysis of back propagation neural networks for classification of remotely sensed multi-spectral imagery, *Int. J. Remote Sens.*, 1995, 16, 3033-3058.
- [44] Jebur M. N., Pradhan B., Tehrany M. S., Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale, *Remote Sens. Environ.*, 2014, 152, 150-165.
- [45] Jebur M., Pradhan B., Tehrany M., Manifestation of LiDAR-Derived Parameters in the Spatial Prediction of Landslides Using Novel Ensemble Evidential Belief Functions and Support Vector Machine Models in GIS, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, doi: 10.1109/JSTARS.2014.2341276
- [46] Zawadzki J., Kędzior M. A., Statistical analysis of soil moisture content changes in Central Europe using GLDAS database over three past decades, *Centr. Eur. J. Geosci.*, 2014, 6, 344-353.
- [47] Zawadzki J., Przeździecki K., Metoda wyznaczania wskaźnika suszy TVDI i jego analiza statystyczna na przykładzie Kampinoskiego Parku Narodowego, *Acta Astrophys.*, 2013, 20, 495-507.
- [48] Hines J.W., Fuzzy and neural approaches in engineering, John Wiley and Sons, Inc., New York, 1997.
- [49] Dai F.C., Lee C.F., Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, *Geomorphology*, 2002, 42, 213-228.
- [50] Hosmer Jr D.W., Lemeshow S., *Applied Logistic Regression*, Wiley & Sons, Inc. 2013, 45-52.