

Article

A Support Vector Machine for Landslide Susceptibility Mapping in Gangwon Province, Korea

Saro Lee ^{1,2}, Soo-Min Hong ^{3,*} and Hyung-Sup Jung ⁴

¹ Geological Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), Daejeon 305350, Korea; leesaro@kigam.re.kr

² Department of Geophysical Exploration, Korea University of Science and Technology, Daejeon 305350, Korea

³ Department of English Language and Literature, University of Seoul, Seoul 02504, Korea

⁴ Department of Geoinformatics, University of Seoul, Seoul 02504, Korea; hsjung@uos.ac.kr

* Correspondence: soomhong@uos.ac.kr; Tel.: +82-2-6490-2522

Academic Editor: Vincenzo Torretta

Received: 13 September 2016; Accepted: 23 December 2016; Published: 1 January 2017

Abstract: In this study, the support vector machine (SVM) was applied and validated by using the geographic information system (GIS) in order to map landslide susceptibility. In order to test the usefulness and effectiveness of the SVM, two study areas were carefully selected: the PyeongChang and Inje areas of Gangwon Province, Korea. This is because, not only did many landslides (2098 in PyeongChang and 2580 in Inje) occur in 2006 as a result of heavy rainfall, but the 2018 Winter Olympics will be held in these areas. A variety of spatial data, including landslides, geology, topography, forest, soil, and land cover, were identified and collected in the study areas. Following this, the spatial data were compiled in a GIS-based database through the use of aerial photographs. Using this database, 18 factors relating to topography, geology, soil, forest and land use, were extracted and applied to the SVM. Next, the detected landslide data were randomly divided into two sets; one for training and the other for validation of the model. Furthermore, a SVM, specifically a type of data-mining classification model, was applied by using radial basis function kernels. Finally, the estimated landslide susceptibility maps were validated. In order to validate the maps, sensitivity analyses were carried out through area-under-the-curve analysis. The achieved accuracies from the SVM were approximately 81.36% and 77.49% in the PyeongChang and Inje areas, respectively. Moreover, a sensitivity assessment of the factors was performed. It was found that all of the factors, except for soil topography, soil drainage, soil material, soil texture, timber diameter, timber age, and timber density for the PyeongChang area, and timber diameter, timber age, and timber density for the Inje area, had relatively positive effects on the landslide susceptibility maps. These results indicate that SVMs can be useful and effective for landslide susceptibility analysis.

Keywords: landslide; GIS; SVM; validation; sensitivity analysis

1. Introduction

Landslides are natural phenomena that can be very hazardous to humans, and thus, landslide susceptibility mapping is very important for the environmental, cultural, economic, and social sustainability of human beings. Throughout the world, many people have died and been injured by landslides, and thousands of hundreds of houses and buildings have been destroyed. As a consequence of this, many researchers have pursued work with the intention of predicting and preventing landslide hazards by using a wide variety of methods [1]. In particular, recent case studies have frequently applied soft computing technology to the assessment of landslide hazards. When creating soft computing models, artificial neural networks [2–6], neuro-fuzzy logic [2,7–9], decision trees [10–15], and support vector machines (SVMs) [10,15–19], have been applied in order to analyze landslide

susceptibility. Among the many soft computing models, SVMs were applied in the present study. They are a relatively new and promising pattern classification technique that were proposed by Vapnik and co-workers [20–23]. Nevertheless, the SVMs are not completely free from problems. The drawback of using SVMs is that they only cover the determination of the parameters for a given value of the regularization and kernel parameters, and choice of kernel [24]. Additionally, the machine learning algorithms, called kernel machines, can be quite sensitive to over-fitting the model selection criterion. Despite this limitation, SVMs are still one of the most popular machine learning algorithms and are considered to be the go-to method for the production of a high-performing algorithm with little tuning.

Approximately 70% of South Korea is mountainous, and landslides commonly occur, especially during the summer rainy season. The study areas for this research, PyeongChang and Inje, are located deep in the highlands of the Gangwon Province. The average annual temperature of these areas is 10.3 °C and 10.1 °C, respectively, and the average yearly rainfall value is 1082 mm and 1210.5 mm, respectively (<http://www.kma.go.kr/>). Very intense precipitation has occurred in the PyeongChang and Inje areas, which has caused many landslides (Figure 1). In 2006, between 12 July and 18 July, typhoon Ewiniar struck the PyeongChang and Inje regions, accompanied by heavy storms and excessive rainfall. Daily rainfall reached 202 mm and cumulative rainfall over the six-day period reached 650 mm. Subsequently, a very large number of landslides occurred in the study areas (Figures 2a and 3a). The typhoon claimed 40 lives and caused approximately 1 billion U.S. dollars worth of property damage. Landslides and the collapse of cut slopes were considered to be the leading causes of death [25]. The extensive damage from rainfall-induced landslides in the PyeongChang and Inje areas can be partly attributed to both the lack of landslide assessment and prediction, and the lack of response plans for minimizing the impacts of the landslides. Moreover, the PyeongChang 2018 Olympic Winter Games will be held in these areas and the construction work for the game venues are well underway. Therefore, it is imperative to enhance the nation’s capability to detect and predict geological hazards like landslides, and to prevent or reduce the risk to life, property, social and economic activities, and natural resources.

The main aim of the present study is to produce a landslide susceptibility map using an SVM. Because SVMs are high-performing machine learning algorithms, they were applied and validated in the PyeongChang and Inje areas of Korea, which served as case studies. Following this, the effect of each factor in the sensitivity analysis was evaluated by validation of landslide susceptibility maps, which excluded individual factors.

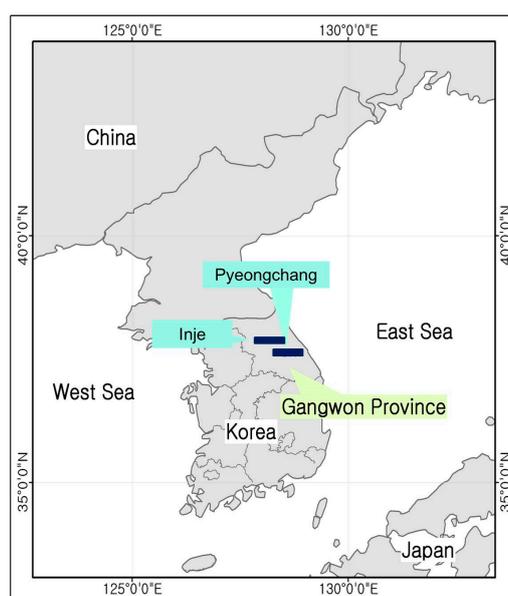


Figure 1. The study areas, the PyeongChang and Inje areas of Gangwon Province.

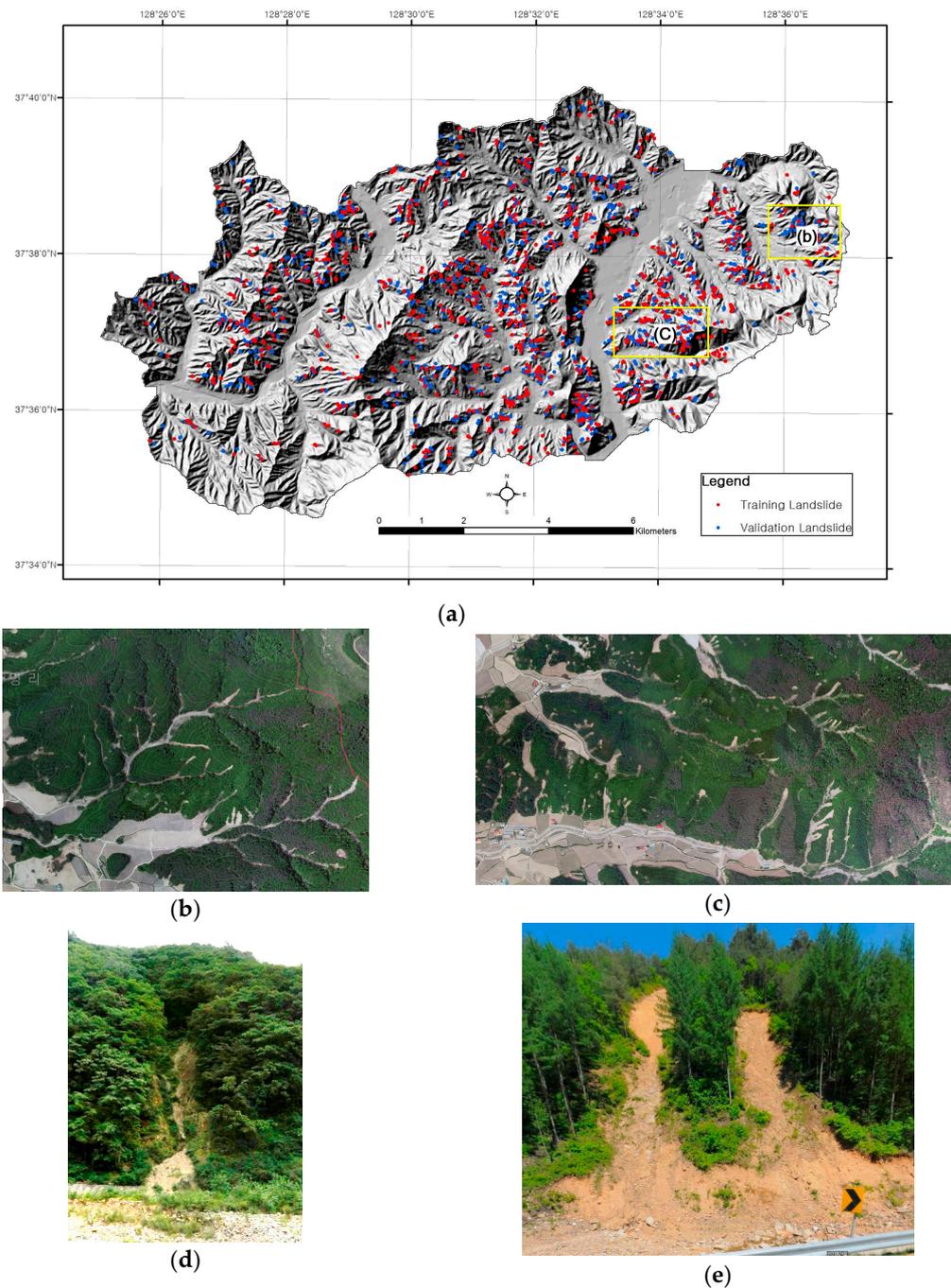


Figure 2. (a) Hillshade map and landslide occurrence in the PyeongChang area; (b,c) respectively show digital aerial photographs of the yellow rectangles shown in (a), and represent areas in which many landslides occurred; (d,e) show typical landslide photographs in the yellow rectangles, respectively.

2. Data

For the accurate detection of landslide locations, digital aerial photographs with ground resolutions of 50 cm were collected from the Daum website (<http://map.daum.net>) (Figure 2b,c and Figure 3b,c). Web-based digital aerial photographs taken from across Korea are readily available on the Daum web portal. We used photographs taken on the 27 May 2008 using the UltraCamX sensor (Microsoft, Graz, Austria). They were taken by Samah Aerial Survey Co. (<http://www.samah.com>) after landslides had occurred during the rainy season of 2006. High-resolution photographs were

rectified via ground control points (GCPs) from digital topographic features. As a result of this, landslide locations could be accurately detected by visual interpretation of the aerial photographs taken after landslide occurrences and checks by field investigations (Figure 2d,e and Figure 3d,e). Consequently, it was reported that both rainfall-triggered shallow landslides and channelized debris flows, occurred widely in the study areas. Most landslides had approximate lengths of 20–3000 m, widths of 5–50 m, and depths of less than 3 m. The landslides were mapped as initiation points. The total number of landslides was 2099 in the PyeongChang area and 2580 in the Inje area. The location of each landslide was denoted using a pixel of 10 m × 10 m. The study areas were delineated based on the basin boundary.

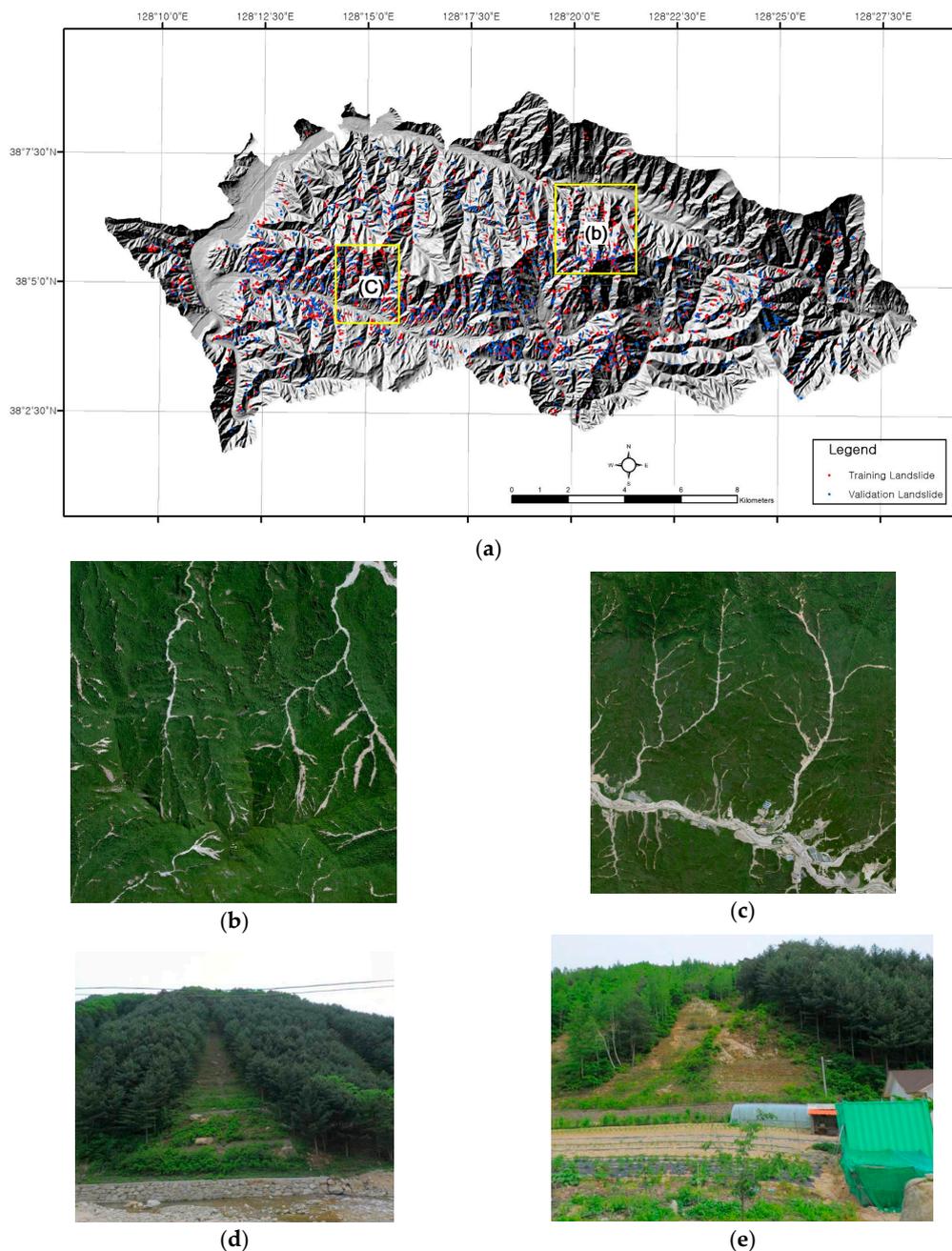


Figure 3. (a) Hillshade map and landslide occurrence in the Inje area; (b,c) respectively show digital aerial photographs of the yellow rectangles shown in (a), and represent areas in which many landslides occurred; (d,e) show typical landslide photographs in the yellow rectangles, respectively.

It is generally believed that landslides result from the interaction of complex factors. The selection of factors, and preparation of corresponding thematic data layers, are crucial for models used in landslide susceptibility mapping [26]. The instability factors responsible for landslides include lithology, geological structure, slope steepness, seismicity, morphology, climate, land use, stream evolution, groundwater conditions, vegetation cover, and human activity. Among these factors, specially topography, geology, soil, forest and land use factors, were taken into account in this study (listed in Table 1), and were collected from available maps and field investigations. A digital elevation model (DEM) with 10 m \times 10 m resolution was prepared by digitization of contours at 5 m intervals from the topographical maps. The slope gradient, slope aspect, plan curvature, slope length, topographic wetness index (TWI) [27], and stream power index (SPI) [28], were calculated by using the DEM. The geology was replaced with the slope length in the Inje area because only one type of geology exists in this region. The pattern of structural lineaments was detected by an interpretation of a hill shade map, produced by a structural geologist with extensive experience working as an interpreter. The selected factors are assumed to have a dominant influence on the occurrence of landslides. Previous studies have analyzed these factors using the same parameters and frequency ratio model in South Korea, including a similar area [13,29–33]. The probability-likelihood ratio method was applied to Boun, Korea [29], and the probability logistic regression method was used for the statistical analysis of land slide susceptibility at Yongin, Korea [30]. Landslide detection was applied with the frequency ratio, weight of evidence, logistic regression, and artificial neural network models in Jinbu, Korea [31]. The decision tree was used for the landslide susceptibility mapping in Pyeongchang, Korea [13], and the integration of frequency ratio and neuro-fuzzy models was used to forecast and validate the landslide susceptibility in the Seorak mountain area [32]. Landslides in the Inje area were mapped using the frequency ratio with the condition of rainfall probability [33].

Table 1. Data layer considered as predisposing factors in the study areas.

Category	Factors	Data Type	Scale
Topographic map [34]	Slope gradient Slope aspect Curvature TWI (Topographic Wetness Index) SPI (Stream Power Index) Slope Length (Inje only)	GRID	1:5000
Geological map [35]	Geology Distance from fault	Polygon GRID	1:50,000
Soil map [36]	Topography Soil drainage Soil material Soil depth Soil texture	Polygon	1:25,000
Forest map [37]	Timber diameter Timber type Timber density Timber age	Polygon	1:25,000
Land use map [38]	Land use (PyeongChang only)	Polygon	1:5000

In order to consider the factors, a geological map was provided with polygon coverage at a scale of 1:50,000, and soil and forest maps (Tables 2 and 3) were presented with polygon coverage at a scale of 1:25,000. These maps were created by the following three institutes: the Korea Institute of Geoscience and Mineral Resources (KIGAM), the Korea Forest Research Institute (KFRI), and the National Academy of Agricultural Science (NAAS). The land-use map with polygon coverage at a scale of 1:5000, was provided by the Korea Ministry of Environment (KME). The land-use types were classified by using a 10 m \times 10 m spatial resolution panchromatic SPOT-5 image, acquired in November 2007. The diameter, density, type, and age of timber, were derived from the forest maps (see Table 3), while soil topography, soil thickness, soil texture, soil material, and soil drainage, were all acquired from the soil maps (see Table 2).

Landslide occurrences were constructed in a vector spatial database using a GIS software package. Seventeen factor maps were generated from the maps, and were then converted into a 10 m × 10 m raster format. Consequently, the dimensions of the study area grids were 1770 columns by 1028 rows for a total of 1,819,560 cells (about 182 km²) in PyeongChang, and 2884 columns by 1299 rows for 3,746,316 cells (about 375 km²) in Inje.

Table 2. Soil map factor classification *.

Soil Topography	Soil Texture	Soil Drainage	Soil Material	Soil Effective Thickness
Water	Water	Water	Water	Water
Fluvial plains	Sandy loam	Somewhat poorly drained	Fluvial alluvium	0–20 cm
Valley and alluvial fan	Fine sandy loam	Moderately well drained	Alluvial-Colluvium	20–50 cm
Lower hilly area	Gravelly sandy loam	Well drained	Okcheon system residuum formation	50–100 cm
Hilly area	Gravelly silt loam	Excessively drained	Colluvium	100–150 cm
Piedmont slope area	Loam	Poorly drained	Diluvium	
Diluvium	Silt loam		Valley alluvium	
Valley area	Gravelly loam		Granite residuum	
Valley and piedmont slope area	Loamy fine sand		Alluvium	
Mountain and hilly area	Overflow area		Phyllite residuum formation	
Mountainous area	Rocky silt loam			
	Rocky sandy loam			

* The terrain unit is 0.25 ha (the distance between survey is 100–200 m). But the unit can be changed according to the condition.

Table 3. Forest map factor classification *.

Timber Type	Timber Diameter	Timber Age	Timber Density
Non-forest	Non-forest	Non-forest	Non-forest
Rigida pine Pine	Very small diameter (timber diameter is below 6 cm)	1st age More than (50% 1–10 years old timber)	Loose (Less than 50% forest area)
Needle and broad Artificially afforested broad leaf tree		2nd age More than (50% 11–20 years old timber)	
Korea nut pine Larch	Small diameter (timber diameter is 6–16 cm)	3rd age More than (50% 21–30 years old timber)	Moderate Less than (51%–70% forest area)
Broad leaf tree Field		4th age More than (50% 31–40 years old timber)	
Cultivated land Chestnut tree Poplar Ranch	Medium diameter (wood diameter is 16–28 cm)	5th age More than (50% 41–50 years old timber)	Dense More than (71% forest area)

* The terrain unit is 0.1 ha for artificial forest area and 0.5 ha for natural forest area.

The topographic factors reflect the geomorphological characteristics of the study areas. Slope gradient, slope aspect, and plan curvature can all influence landslide initiation [30,39,40]. The hillslope profile indicates the thickness and composition of soil horizons, which vary with not only position on a hillslope, but also with water drainage. In the hillslope domain, between the drainage divides and the stream network, several topographic attributes are distributed, such as slope, curvature, and TWI [41–44]. The topography has a vital role in the spatial variation of hydrological conditions, such as soil moisture, groundwater flow, and slope stability. Topographic indices such as SPI and TWI are used to describe spatial soil moisture patterns [27,28]. Lithology plays an important role in landslide occurrences because different

lithologic units have varied inherent characteristics, including strength, composition, and structure, producing varied resistance against landslides [45–47]. Fault lines are the expression of structural brittle deformation of rocks due to tectonics, and thus landslide susceptibility is higher along these features. The occurrence of landslides varies with land-use pattern, which is an indication of the stability of hillslopes [48]. Forest cover and soil properties also affect various geomorphologic and hydrologic processes, including surface erosion, hillslope change, and the rate of landslide occurrences [49,50].

3. Methods

The detailed workflow for landslide susceptibility map creation is shown in Figure 4. In this study, landslide occurrence locations were identified by using digital aerial photographs. Following this, 50% of the landslide occurrences were randomly selected as training data, and the others were determined to be validation data. Geology, topography, soil, forest, and land-use datasets, were entered into a GIS-based spatial database, and the 18 landslide-related factors were extracted from the database. In this way, SVMs were applied for landslide susceptibility mapping.

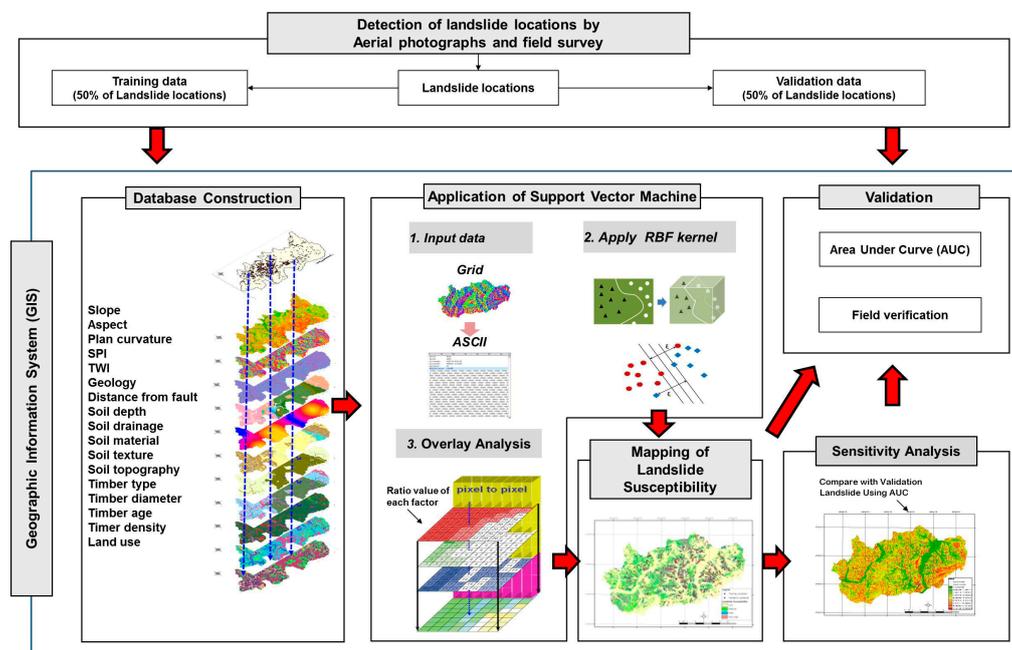


Figure 4. Flowchart of the study procedures.

An SVM is a supervised learning method based on statistical learning theory and the principle of structural risk minimization [22]. Using the training data, the SVM implicitly maps the original input space onto a high-dimensional feature space [51]. Subsequently, in the feature space, the optimal hyperplane is determined by maximizing the margins of class boundaries [52]. The SVM intends to minimize the upper bound of the generalization error by maximizing the margin between the separating hyperplane and the data [53]. The training points that are closest to the optimal hyperplane are called support vectors. The aim of SVM classification is to find an optimal separating hyperplane that can distinguish between the two classes (i.e., landslides and no landslides), and the set of training data [10]. Two main concepts underlie SVM modeling for discriminant-type statistical problems. The first of these concepts is an optimum linear separating hyperplane that separates data patterns. The second, is the use of kernel functions for converting the original nonlinear data patterns, into a format that is linearly separable in a high-dimensional feature space [54].

Detailed descriptions of two-class SVM modeling were provided by [2,17,54] and were summarized in the following explanation. Consider a set of linear separable training vectors x_i

($i = 1, 2, \dots, n$). The training vectors consist of two classes, denoted as $y_i = \pm 1$. The goal of the SVM is to search an n -dimensional hyperplane, differentiating the two classes by their maximum gap (Figure 5a). This is expressed as:

$$\frac{1}{2} \|w\|^2 \quad (1)$$

subject to the following constraints:

$$y_i((w \cdot x_i) + b) \geq 1 \quad (2)$$

where $\|w\|$ is the norm of the hyperplane, b is a scalar base, and (\cdot) denotes the scalar product operation. Using the Lagrangian multiplier, the cost function can be defined as:

$$L = \frac{1}{2} \|W\|^2 - \sum_{i=1}^n \lambda_i (y_i((w \cdot x_i) + b) - 1) \quad (3)$$

where λ_i is the Lagrangian multiplier. The solution can be achieved by dual minimization of Equation (3), with respect to w and b through standard procedures. For the non-separable case (Figure 5b), the constraints can be modified by introducing slack variables ζ_i [22]:

$$y_i((w \cdot x_i) + b) \geq 1 - \zeta_i \quad (4)$$

and Equation (1) becomes:

$$L = \frac{1}{2} \|W\|^2 - \frac{1}{vn} \sum_{i=1}^n \zeta_i \quad (5)$$

where $v(0,1)$ is introduced in order to account for misclassification [55,56]. In addition, Vapnik [22] introduced a kernel function $K(x_i, x_j)$ to account for the nonlinear decision boundary.

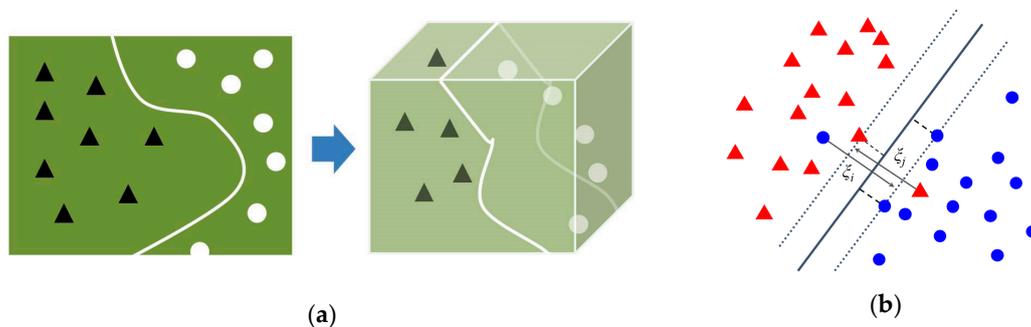


Figure 5. Explanation of the support vector machine (SVM) principle. (a) n -Dimensional hyperplane differentiating the two classes by the maximum gap; (b) non-separable case and the slack variables ξ_i [54].

In the present SVM study, the Environment for Visualizing Images (ENVI) (ENVI 4.4 2006, Harris Corporation, Jersey City, NJ, USA) was used. The ENVI 4.4 SVM classifier provides four types of kernels: linear, polynomial, radial basis function (RBF), and sigmoid. The mathematical representations of each kernel (linear, polynomial, radial basis function, and sigmoid, respectively) are listed below (ENVI 4.4 2006):

$$\begin{aligned} \text{Linear} : K(x_i, y_i) &= x_i^T \cdot x_j, \\ \text{Polynomial} : K(x_i, y_i) &= (\gamma \cdot x_i^T \cdot x_j + r)^d, \quad \gamma > 0, \\ \text{Radial basis function} : K(x_i, y_i) &= e^{-\gamma(x_i - x_j)^2}, \quad \gamma > 0, \\ \text{Sigmoid} : K(x_i, y_i) &= \tanh(\gamma \cdot x_i^T \cdot x_j + r) \end{aligned} \quad (6)$$

where γ , r , and d are parameters of the kernel functions and are entered manually. In the present study, the RBF kernel was used (often called the Gaussian kernel) because it is one of the most

powerful kernels [10,17]. The RBF kernel is the default kernel, and it works well in most of the cases (ENVI 4.4 2006). Moreover, in many studies and cases (especially in nonlinear problems), RBF provides better prediction results for landslide susceptibility mapping than other kernels [2,10,57,58]. In order to perform the landslide susceptibility mapping using an SVM, the following steps had to be taken:

- (1) Preparation of the 18 landslide-related factors as GIS data;
- (2) Opening of the landslide-related factors to use ENVI software through a TIFF image file;
- (3) Defining the region of interest (ROI) through using the landslide location data;
- (4) Running the SVM classification algorithm by using the RBF kernel for each factor;
- (5) Summarizing the result of SVM classification for each factor;
- (6) Validation of the summarized result by using the area-under-the-curve (AUC) method.

A sensitivity analysis showed how a solution might be changed when the input factors are also altered. If the selected factor results in a relatively large change in the outcome, the outcome is believed to be affected by that specific factor. The factors that have the greatest impact on the calculated landslide susceptibility map can therefore be identified through using sensitivity analysis. In the present study, sensitivity analysis was conducted by excluding each factor in turn during the summation stage:

$$\text{LSI SEN}_i = \text{SVM}_{\text{all}} - \text{SVM}_i \quad (i = 1, 2, \dots, n) \quad (7)$$

where LSI SEN_i is the landslide susceptibility index (LSI) of a factor omitted from the sensitivity analysis, SVM_{all} is the sum of the result of classification when using the SVM of all factors, and SVM_i is the result of classification when using the SVM of a particular factor. Here, n is the total number of input factors. The LSI was used to map the landslide susceptibility. Finally, the landslide susceptibility was mapped with the result of the classification, and the sensitivity analysis was validated through the use of existing landslide locations, that were not used to train the model and the AUC method.

4. Results

The SVM was applied and the results were used to produce landslide susceptibility maps of the study areas. In order to create a landslide susceptibility map, four classes were established, based on areas for simple and visual interpretation. These classes were: very high (10%), high (10%), medium (20%), and low (60%) (Figure 6a for PyeongChang area and Figure 6b for Inje area). The classification was useful for both estimating the possibility of a landslide in each class, and for visually delineating susceptible zones in residential and facility areas.

A landslide susceptibility map should be able to make an effective prediction of possible landslide areas. It can also be validated through incorporating data acquired from new landslide occurrence locations, if landslides do occur. In this study, the validation of the landslide susceptibility analysis was performed by using 50% of total landslide occurrences that were not used as the training data. The validation of the landslide susceptibility map was completed by following these four steps: (1) sorting the calculated LSI values in all cells into descending order; (2) breaking down the ordered cell values into 100 classes with cumulative 1% intervals; (3) adapting the above procedure for the landslide occurrence cells by comparing the 100 classes; and (4) making a graph to compare the two sets of classifications.

As a result of this, in the case of the PyeongChang area, the 90%–100% (10%) class of the study areas in which the LSI had a higher rank, explained 40% of the entire landslides. In addition, the 80%–100% (20%) class of the study areas in which the LSI had a higher rank, explained 63% of the landslides. In order to compare the results quantitatively, AUCs were recalculated as the total area [59,60]. Therefore, the AUC could be used to assess the prediction accuracy qualitatively. From the validation of the landslide susceptibility maps, the RBF kernel produced AUC values, indicating the accuracy of the landslide susceptibility maps, and these were 81.36% for the PyeongChang area, and 77.49% for the Inje area (Figure 7). There were some differences in accuracy between the study areas,

because the previous studies [2–19,29–33] showed that the spatial distribution is subject to change, according to the area and event. However, the accuracy was usually high enough, displaying figures of above 80%.

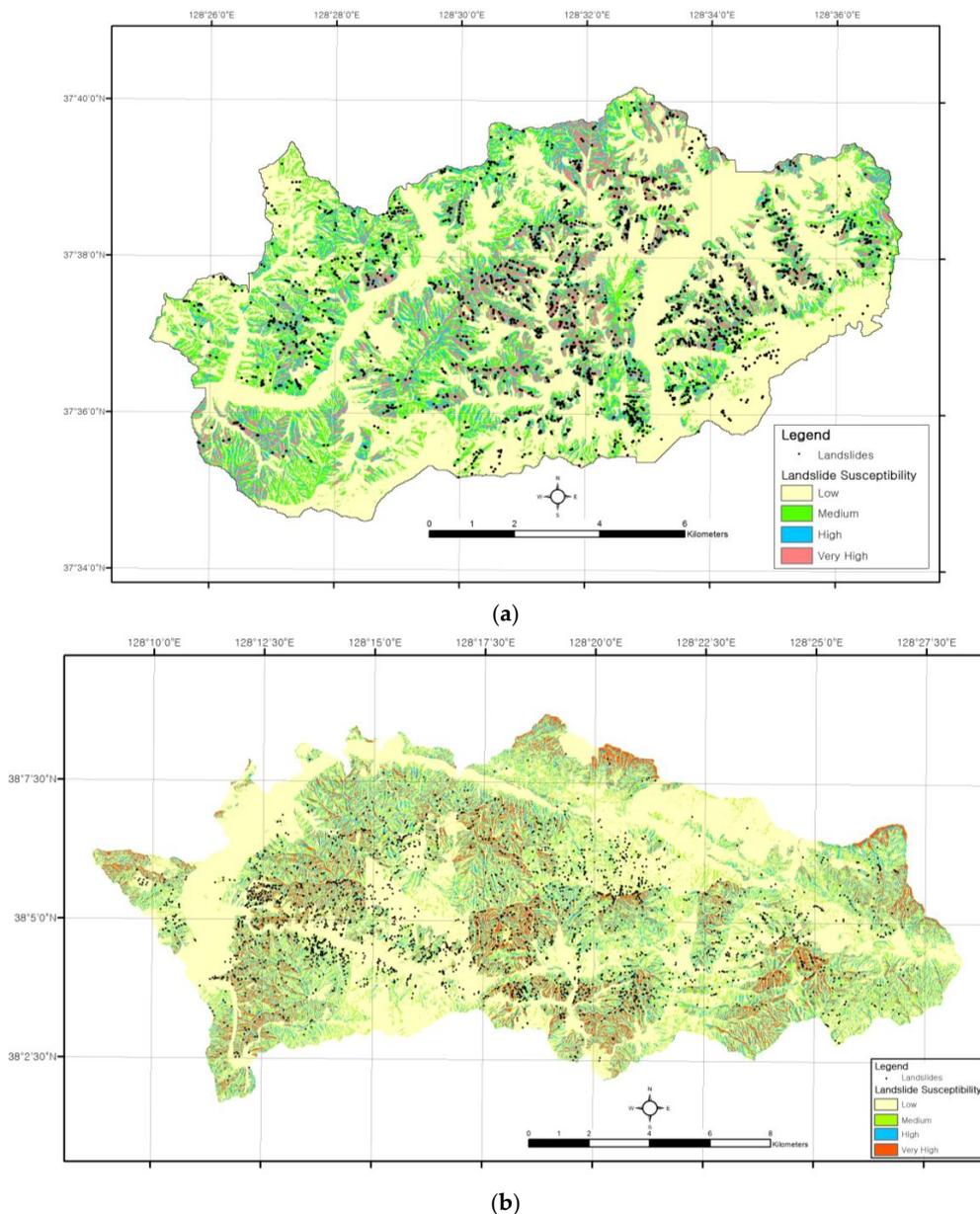


Figure 6. Landslide susceptibility maps created support vector model. The index was classified into four classes based on an area for simple and visual interpretation: very high, high, medium, and low index ranges in 10%, 10%, 20%, and 60% of the study area, respectively. (a) The PyeongChang area and (b) the Inje area.

The sensitivity analyses were conducted by excluding each factor (Equation (7)) in turn during the summation stage of the SVM, before the effect of each factor was evaluated. In this way, a sensitivity analysis was performed in order to make sure that the model system used was susceptible to various factor selection. The model outputs were compared with the expected output changes. In order to conduct the sensitivity analysis, we re-used the rate curve, as well as the AUC method. In accordance with the landslide susceptibility validation of the PyeongChang area by the sensitivity analysis (Table 4), all of the factors, including aspect, SPI, TWI, slope, land use, geology, plan curvature, distance from

fault, timber type, and soil depth, exerted a minor positive influence on the landslide susceptibility maps. On the contrary, the remaining factors, including soil topography, soil material, soil drainage, soil texture, timber age, timber diameter, and timber density, exercised a minor negative influence on the landslide susceptibility. Moreover, for the Inje area, SPI, slope length, aspect, slope, TWI, plan curvature, geology, distance from fault, soil material, soil depth, soil topography, soil texture, soil drainage, and timber type, in sequence, also had small positive influences on the landslide susceptibility maps. In stark contrast, timber diameter, timber density, and timber age, had small negative influences on the landslide susceptibility maps. This is because the lower the value of the AUC gets, the greater the effect of the factor on the landslide susceptibility maps will be. Conversely, a larger AUC means that the factor has a more negative effect on the landslide susceptibility maps.

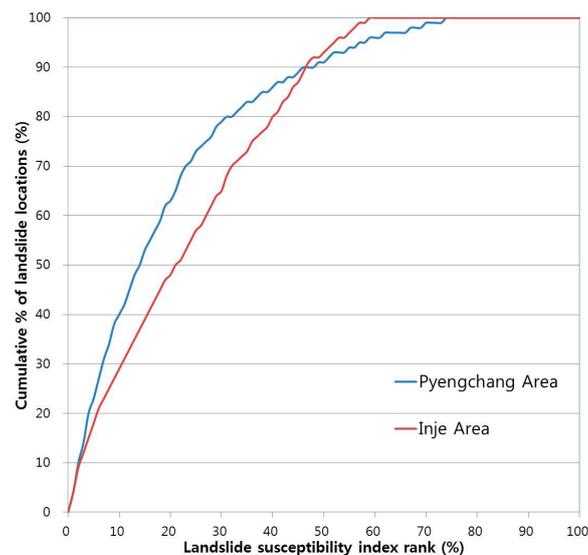


Figure 7. Cumulative frequency diagram showing the landslide susceptibility index rank (x -axis) occurring in the cumulative percent of landslide locations (y -axis). From the validation of the landslide susceptibility maps, SVM approaches the produced area-under-the-curve (AUC) values of accuracy of 81.36% (PyeongChang area with all factors) and 77.49% (Inje area with all factors).

Table 4. Result of sensitivity analysis.

PyeongChang			Inje		
Factor	AUC	Effect	Factor	AUC	Effect
Aspect	78.66	Positive	SPI	75.93	Positive
Land use	80.32	Positive	Slope Length	76.22	Positive
SPI	80.70	Positive	Aspect	76.28	Positive
Slope	80.72	Positive	Slope	76.68	Positive
TWI	80.81	Positive	TWI	76.76	Positive
Geology	80.89	Positive	Plan curvature	77.18	Positive
Plan curvature	81.06	Positive	Geology	77.22	Positive
Distance from fault	81.07	Positive	Distance from fault	77.32	Positive
Timber type	81.18	Positive	Soil material	77.37	Positive
Soil depth	81.30	Positive	Soil depth	77.37	Positive
All factor used	81.36		Soil topography	77.42	Positive
Soil topography	81.36	Negative	Soil texture	77.42	Positive
Soil drainage	81.37	Negative	Soil drainage	77.43	Positive
Soil material	81.39	Negative	Timber type	77.47	Positive
Soil texture	81.55	Negative	All factor used	77.49	
Timber diameter	81.81	Negative	Timber diameter	77.79	Negative
Timber age	81.93	Negative	Timber density	77.88	Negative
Timber density	82.22	Negative	Timber age	78.35	Negative

AUC: area-under-the-curve.

5. Discussion and Conclusions

In the present study, the SVM was applied for predicting landslide occurrences in the areas susceptible to these phenomena in the PyeongChang and Inje areas, where landslides are expected to continue to strike in the future. For application of the SVM, the RBF kernel was used. In order to validate the maps, landslide locations that were not used in the training, were used to assess predictive capacity of the landslide susceptibility maps. The maps were developed by using an RBF kernel that had high accuracy rates (81.36% for the PyeongChang area and 77.49% for the Inje area), as validated by the AUC method. Therefore, the SVM can be used efficiently for landslide susceptibility analysis and may be used widely for the prediction of various spatial events.

In order to assess the sensitivity of the factors, each factor was excluded from the analysis in turn, and its sensitivity was validated through the use of the landslide location data. In the Inje area, topography, geology, and soil-related factors, had relatively positive effects (especially SPI, slope length, aspect, slope, and TWI) on the landslide susceptibility maps. On the contrary, in the PyeongChang area, topography-related, and geology-related factors, had relatively positive effects (especially aspect, land use, SPI, slope, and TWI) on the landslide susceptibility maps. From the sensitivity analysis, we know that the topography-derived factors, such as slope, aspect, SPI, TWI, and slope length, are important for landslide susceptibility mapping. The landslide susceptibility mapping uncertainty comes from the input factor error, the positioning error of aerial photograph, and the landslide location visual interpretation error. The SVM is a useful and flexible supervised classifier, that is suitable for a wide range of classification problems, even if the problems are in high dimensions and are not linearly separable. In addition, SVMs are flexible when considering the choice of the threshold form separating susceptible areas from non-susceptible areas, by introducing the kernel. The kernel implicitly contains a non-linear transformation; therefore, any particular assumptions of the functional form of the transformation do not have to be made [61]. When using the SVM, decision rules provide a general method of function estimation that is performed by solving a convex (quadratic) optimization problem. In summary, SVMs have a significant advantage when compared to other machine learning algorithms: the uniqueness of the solution [57]. Nevertheless, the SVM is still a binary classifier, so multi-class classification is not eligible in this case. Moreover, the SVM does not offer many choices for controlling values, neither does it directly provide probability and statistics estimates in the results and procedure, as it is a non-parametric technique.

The present study identified factors that may be involved in landslides, and the results and methods that can be used for landslide susceptibility mapping in other regions beyond the study areas. Landslide susceptibility maps are able to help implement a guide for planning mass evacuation of residents in the case of a landslide, and also to prevent or reduce the disruptive impacts of a natural disaster on surrounding communities. Landslide susceptibility mapping is expected to be readily applied to other areas, yielding a comprehensive and useful analysis model. Especially when considering more practical use of landslide management, landslide hazard and risk analysis must be completed [62]. The analysis will be based on the relationship between the landslide occurrence and precipitation. Therefore, in the not-too-distant future, a landslide warning system may be established by forecasting landslides induced by rainfall. Nevertheless, caution should be exercised when using the models for specific site development, and the scale of the analysis should be considered with great care. In sum, the models used in the present study are not valid for specific planning and assessment purposes, but for general planning and assessment purposes.

Acknowledgments: Saro Lee was 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. Hyung-Sup Jung and Soo-Min Hong were supported by the "Development of Scene Analysis & Surface Algorithms" project, funded by ETRI, which is a subproject of "Development of Geostationary Meteorological Satellite Ground Segment (NMSC-2016-01)" program funded by NMSC (National Meteorological Satellite Center) of KMA (Korea Meteorological Administration).

Author Contributions: Saro Lee collected data and processed input data. Soo-Min Hong managed the paperwork and interpreted the results. Hyung-Sup Jung suggested the idea and prepared the data. All of the authors contributed to the writing of each part.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bertazzi, P.A. Disasters, Natural and Technological. Available online: <http://www.ilocis.org/documents/chpt39e.htm> (accessed on 26 December 2016).
2. 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. [[CrossRef](#)]
3. Tien, B.D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility assessment in the HoaBinh province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. *Geomorphology* **2012**, *171–172*, 12–29. [[CrossRef](#)]
4. Alkhasawneh, M.S.; Ngah, U.K.; Tay, L.T.; Isa, N.A.M. Determination of importance for comprehensive topographic factors on landslide hazard mapping using artificial neural network. *Environ. Earth Sci.* **2014**, *72*, 787–799. [[CrossRef](#)]
5. Bi, R.; Schleier, M.; Rohn, J.; Ehret, D.; Xiang, W. Landslide susceptibility analysis based on ArcGIS and Artificial Neural Network for a large catchment in Three Gorges region, China. *Environ. Earth Sci.* **2014**, *72*, 1925–1938. [[CrossRef](#)]
6. Gelisli, K.; Kaya, T.; Babacan, A.E. Assessing the factor of safety using an artificial neural network: Case studies on landslides in Giresun, Turkey. *Environ. Earth Sci.* **2015**, *73*, 8639–8646. [[CrossRef](#)]
7. Song, K.Y.; Oh, H.J.; Choi, J.; Lee, S. Prediction of landslides using ASTER imagery and data mining models. *Adv. Space Res.* **2012**, *49*, 978–993. [[CrossRef](#)]
8. Dehnavi, A.; Aghdam, I.N.; Pradhan, B.; Morshed, V.M.H. A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena* **2015**, *135*, 122–148. [[CrossRef](#)]
9. Lee, M.J.; Song, W.K.; Won, J.S.; Park, I.; Lee, S. Spatial and temporal change in landslide hazard by future climate change scenarios using probabilistic based frequency ratio model. *Geocarto Int.* **2014**, *29*, 639–662.
10. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
11. Althuwaynee, O.F.; Pradhan, B.; Park, H.J.; Lee, J.H. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides* **2014**, *11*, 1063–1078. [[CrossRef](#)]
12. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2014**, *11*, 425–439. [[CrossRef](#)]
13. Park, I.; Lee, S. Spatial prediction of landslide susceptibility using a decision tree approach: A case study of the PyeongChang area, Korea. *Int. J. Remote Sens.* **2014**, *35*, 6089–6112. [[CrossRef](#)]
14. Wang, L.J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosci. J.* **2016**, *20*, 117–136. [[CrossRef](#)]
15. Wu, X.; Ren, F.; Niu, R. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ. Earth Sci.* **2014**, *71*, 4725–4738. [[CrossRef](#)]
16. 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. [[CrossRef](#)]
17. Xu, C.; Dai, F.; Xu, X.; Lee, Y.H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. [[CrossRef](#)]

18. Ren, F.; Wu, X.; Zhang, K.; Niu, R. Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. *Environ. Earth Sci.* **2015**, *73*, 4791–4804. [[CrossRef](#)]
19. Hong, H.; Pradhan, B.; Jebur, M.N.; Bui, D.T.; Xu, C.; Akgun, A. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **2016**, *75*, 1–14. [[CrossRef](#)]
20. Boser, B.; Guyon, I.; Vapnik, V. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992.
21. Cortes, C.; Vapnik, V. Support vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
22. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
23. Vapnik, V. *Statistical Learning Theory*; Wiley: Hoboken, NJ, USA, 1998.
24. Cawley, C.; Talbot, N.L.C. Over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
25. Ministry of Public Safety and Security, 2006. Hazard Annals of South Korea in 2006. Available online: http://www.mpss.go.kr/board/board.do?boardId=bbs_0000000000000042&mode=view&cntId=16&category=&pageIdx= (accessed on 26 12 2016).
26. Sarkar, S.; Kanungo, D.P. An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 617–625. [[CrossRef](#)]
27. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
28. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological and biological applications. *Hydrol. Proc.* **1991**, *5*, 3–30. [[CrossRef](#)]
29. Lee, S.; Choi, J.; Min, K.-D. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *Int. J. Remote Sens.* **2004**, *25*, 2037–2052. [[CrossRef](#)]
30. Lee, S.; Min, K.-D. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ. Geol.* **2001**, *40*, 1095–1111.
31. Lee, S.; Song, K.-Y.; Oh, H.-J.; Choi, J. Detection of landslide using web-based aerial photographs and landslide susceptibility mapping using geospatial analysis. *Int. J. Remote Sens.* **2012**, *33*, 4937–4966.
32. Lee, M.-J.; Park, I.; Lee, S. Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: A case study of Seorak mountain area in Korea. *Environ. Earth Sci.* **2015**, *74*, 413–429. [[CrossRef](#)]
33. Lee, M.-J.; Park, I.; Won, J.-S.; Lee, S. Landslide hazard mapping considering rainfall probability in Inje, Korea. *Geomat. Nat. Haz. Risk* **2016**, *7*, 424–446. [[CrossRef](#)]
34. Land Information Platform. Available online: <http://map.ngii.go.kr> (accessed on 17 December 2016).
35. Geological Information System. Available online: <http://mgeo.kigam.re.kr> (accessed on 17 December 2016).
36. Soil Environment Information System. Available online: <http://soil.rda.go.kr/soil/soilmap/characteristic.jsp> (accessed on 17 December 2016).
37. Forest Geospatial Information Service. Available online: <http://116.67.44.22/forest> (accessed on 17 December 2016).
38. Environmental Geospatial Information Service. Available online: <https://egis.me.go.kr> (accessed on 17 December 2016).
39. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2012**, *42*, 213–228. [[CrossRef](#)]
40. 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. [[CrossRef](#)]
41. Pirotti, F.; Tarolli, P. Suitability of lidar point density and derived landform curvature maps for channel network extraction. *Hydrol. Proc.* **2010**, *24*, 1187–1197. [[CrossRef](#)]
42. Gangodagamage, C.; Belmont, P.; Foufoula, G.E. Revisiting scaling laws in river basins: New considerations across hillslope and fluvial regimes. *Water Resour. Res.* **2011**. [[CrossRef](#)]
43. Tarolli, P.; Borga, M.; Chang, K.; Chiang, S. Modeling shallow landsliding susceptibility by incorporating heavy rainfall statistical properties. *Geomorphology* **2011**, *133*, 199–211. [[CrossRef](#)]
44. Lee, S.; Hwang, J.; Park, I. Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu, Korea. *Catena* **2013**, *100*, 15–30. [[CrossRef](#)]

45. Kincal, C.; Akgun, A.; Koca, M.Y. Landslide susceptibility assessment in the İzmir (West Anatolia, Turkey) city center and its near vicinity by the logistic regression method. *Environ. Earth Sci.* **2009**, *59*, 745–756. [[CrossRef](#)]
46. Chauhan, S.; Sharma, M.; Arora, M.K. Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model. *Landslides* **2010**, *7*, 411–423. [[CrossRef](#)]
47. Lee, S.; Won, J.S.; Jeon, S.W.; Park, I.; Lee, M.J. Spatial Landslide Hazard Prediction Using Rainfall Probability and a Logistic Regression Model. *Math. Geosci.* **2015**, *47*, 565–589. [[CrossRef](#)]
48. Anbalagan, R. Landslide susceptibility evaluation and zonation mapping in mountainous terrain. *Eng. Geol.* **1992**, *32*, 269–277. [[CrossRef](#)]
49. Edeso, J.M.; Merino, A.; Gonzalez, M.J.; Marauri, P. Soil erosion under different harvesting managements in steep forest lands from northern Spain. *Land Degrad. Dev.* **1999**, *10*, 79–88. [[CrossRef](#)]
50. Dhakal, A.S.; Sidle, R.C. Long-term modeling of landslides for different forest management practices. *Earth Surf. Proc. Land* **2003**, *28*, 853–868. [[CrossRef](#)]
51. Kanevski, M.; Pozdnoukhov, A.; Timonin, V. *Machine Learning Algorithms for Geospatial Data. Theory, Applications and Software*; PPUR EPFL-Press: Lausanne, Switzerland, 2009.
52. Abe, S. *Support Vector Machines for Pattern Classification*; Springer: New York, NY, USA, 2010.
53. Amari, S.; Wu, S. Improving support vector machine classifiers by modifying kernel functions. *Neural Netw.* **1999**, *12*, 783–789. [[CrossRef](#)]
54. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]
55. Scholkopf, B.; Smola, A.; Williamson, R.C.; Bartlett, P.L. New support vector algorithms. *Neural Comput.* **2000**, *12*, 1207–1245. [[CrossRef](#)] [[PubMed](#)]
56. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining Inference and Prediction*; Springer: New York, NY, USA, 2001.
57. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Xu, C.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [[CrossRef](#)]
58. Taner, S.B. An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The Candir catchment area (western Antalya, Turkey). *Int. J. App. Earth. Obs. Geoinform.* **2004**, *26*, 399–412.
59. Lee, S.; Dan, N.T. Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: Focus on the relationship between tectonic fractures and landslides. *Environ. Geol.* **2005**, *48*, 778–787. [[CrossRef](#)]
60. Lee, S.; Sambath, T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ. Geol.* **2006**, *50*, 847–855. [[CrossRef](#)]
61. Auria, L.; Moro, R.A. Support Vector Machines (SVM) as a Technique for Solvency Analysis. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1424949 (accessed on 17 December 2016).
62. Einstein, H.H. Landslide Risk Assessment Procedure. In Proceedings of the Fifth International Symposium on Landslide, Lausanne, Switzerland, 10–15 July 1988.

