

Article

# Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea

Sung-Jae Park <sup>1</sup>, Chang-Wook Lee <sup>1,\*</sup>, Saro Lee <sup>2,3,\*</sup> and Moung-Jin Lee <sup>4</sup>

<sup>1</sup> Division of Science Education, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Korea; tres13@kangwon.ac.kr

<sup>2</sup> Geological Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), Gajeong-dong 30, Yuseong-gu, Daejeon 305-350, Korea

<sup>3</sup> Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 34113, Korea

<sup>4</sup> Center for Environmental Assessment Monitoring, Environmental Assessment Group, Korea Environment Institute, 370 Sicheong-daero, Sejong-si 399-007, Korea; leemj@kei.re.kr

\* Correspondence: cwlee@kangwon.ac.kr (C.-W.L.); leesaro@kigam.re.kr (S.L.); Tel.: +82-33-250-6731 (C.-W.L.); +82-42-868-3057 (S.L.)

Received: 21 August 2018; Accepted: 21 September 2018; Published: 25 September 2018



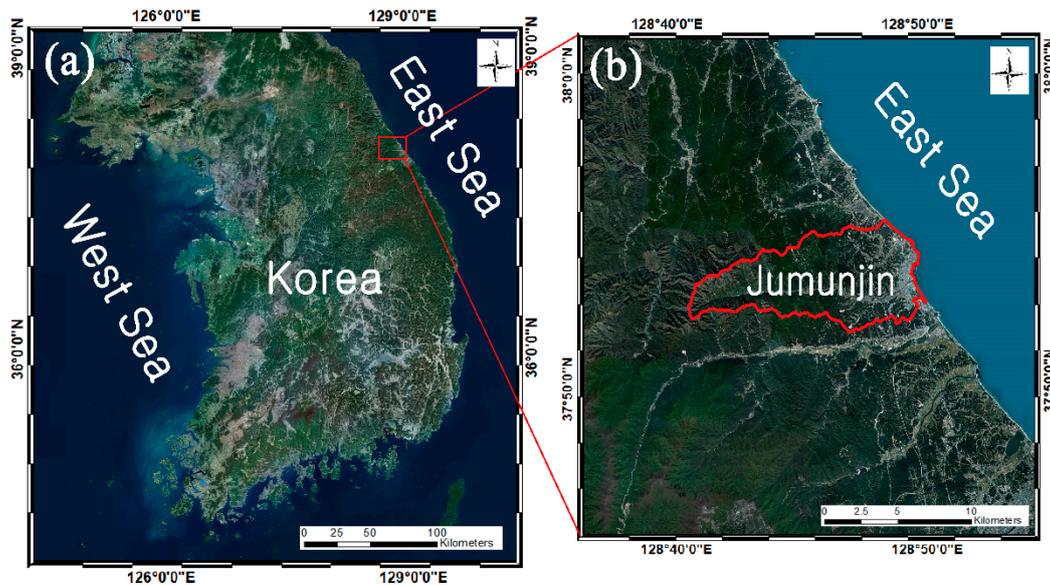
**Abstract:** We assessed landslide susceptibility using Chi-square Automatic Interaction Detection (CHAID), exhaustive CHAID, and Quick, Unbiased, and Efficient Statistical Tree (QUEST) decision tree models in Jumunjin-eup, Gangneung-si, Korea. A total of 548 landslides were identified based on interpretation of aerial photographs. Half of the 548 landslides were selected for modeling, and the remaining half were used for verification. We used 20 landslide control factors that were classified into five categories, namely topographic elements, hydrological elements, soil maps, forest maps, and geological maps, to determine landslide susceptibility. The relationships of landslide occurrence with landslide-inducing factors were analyzed using CHAID, exhaustive CHAID, and QUEST models. The three models were then verified using the area under the curve (AUC) method. The results showed that the CHAID model (AUC = 87.1%) was more accurate than the exhaustive CHAID (AUC = 86.9%) and QUEST models (AUC = 82.8%). The verification results showed that the CHAID model had the highest accuracy. There was high susceptibility to landslides in mountainous areas and low susceptibility in coastal areas. Analyzing the characteristics of the landslide control factors in advance will enable us to obtain more accurate results.

**Keywords:** landslide susceptibility; decision tree; CHAID; exhaustive CHAID; QUEST

## 1. Introduction

Landslides are natural disasters that can cause serious losses to both human life and property. In general, landslides are closely related to the slope of the terrain and high-slope terrain generally occurs in mountainous areas. With 70% of its total area covered by mountainous terrain, Korea is particularly susceptible to landslides. In addition to these topographical conditions, heavy rainfall in summer and high rainfall due to typhoons also increases the likelihood of landslides [1].

Jumunjin area is located in Gangneung-si, Gangwon-do, South Korea (Figure 1a,b). The main economic activities in the city of Gangneung-si are agriculture and tourism. The average annual rainfall is 1018.7 mm. The mountainous terrain is composed mostly of granite that was formed in the Mesozoic Jurassic, and the river basin, which contains the densely-populated residential areas, is composed of alluvial deposits from the Fourth Cenozoic era.



**Figure 1.** Location of study area from Daum map (a) Korea map and (b) Jumunjin area map marked by red boundary [2].

Typhoons Rusa in 2002 and Maemi in 2003 were accompanied by record rainfall of more than 80 mm/h. The two typhoons were responsible for more than 250 deaths and injuries. Typhoons Rusa and Maemi caused numerous landslides in the Sacheon-myeon (70.84 km<sup>2</sup>, population 4219, 1914 households) and Jumunjin area (60.55 km<sup>2</sup>, population 21,291, 8917 households). In particular, seven landslides occurred in Gangneung-si, and the city, as well as the present study area, was cut off by a landslide, which was caused by Typhoon Rusa in 2002. This storm also claimed the lives of three residents from Jumunjin area, which was cut off for more than ten days.

To choose the most appropriate approaches to minimize the damage caused both directly and indirectly by landslides, we must first identify the areas that are susceptible to landslides and where they are most likely to occur [3,4]. The most common approach used to identify landslide-susceptible areas is (geographic information system) GIS-based landslide susceptibility assessments, which include various classification-based methods, such as statistical, machine learning, and probabilistic approaches [5,6]. Probabilistic models are based on the same method as the frequency ratio (FR) [6–9], the weight of evidence (WOE), and the evidential belief function (EBF). Statistical models include the statistical index, analytical hierarchy process (AHP) [10], and logistic regression [11–13]. In recent years, fuzzy logic [14–16], fuzzy rule-based classifiers, neuro-fuzzy models [17–19], multivariate adaptive regression splines (MARS) [20], decision trees [21–24], neural networks [25–28], and support vector machines [29–31] have also been used. In this study, we applied probabilistic models to Gangneung-si in Gangwon-do using landslide data generated during typhoons in 2002 and 2003. Because the accuracy of the same model may vary for the mapping of landslide susceptibilities, it is important to select the most appropriate model for the region's characteristics. Decision tree models using CHAID, exhaustive CHAID, and QUEST were applied to each study area to select the more accurate algorithms.

## 2. Data and Pre-Processing

Daum map is useful because it provides previous as well as current aerial photographs [30,31]. The aerial photographs without ground control point used to find landslide points in this study are two photographs taken in 2008 and 2014. It has 0.5 × 0.5 m resolution. 2008 data is used to analyze events after landslides because data in 2008 recorded many small-scale landslides that spread almost throughout the region. We identify small-scale landslides to prevent large-scale landslides when typhoon occurs. 2014 data used to watch before landslides occur because during this period large-scale landslides did not occur. There is a disadvantage in that the acquired photographs have no coordinate

system, but those have the advantage of obtaining higher resolution than satellite images. In order to solve the above, Georeferencing was performed to input GCPs of recent aerial photographs with coordinate system in 2008 and 2014 photographs [32–35]. This method can cause errors, because the users directly check to the reference point. To minimize this error, we used the same coordinates as the structure that is fixed in the study area and repeated until the RSME value was less than 0.5. The landslide point was selected in a naturally tree-grown area compared to the photographs that were obtained in 2008 and 2014 (Figure 2). From the digital aerial photograph, we identified areas that experienced landslides. All landslides are marked using point data. Using this process, a total of 548 locations of landslide occurrence were identified in the study area as landslide inventory (Figure 3). The total of 548 landslides were divided into two datasets; 50% of these were designated as training data and 50% as validation data. The training data were used to building landslide susceptibility map, while the validation data were used to accuracy assessment.

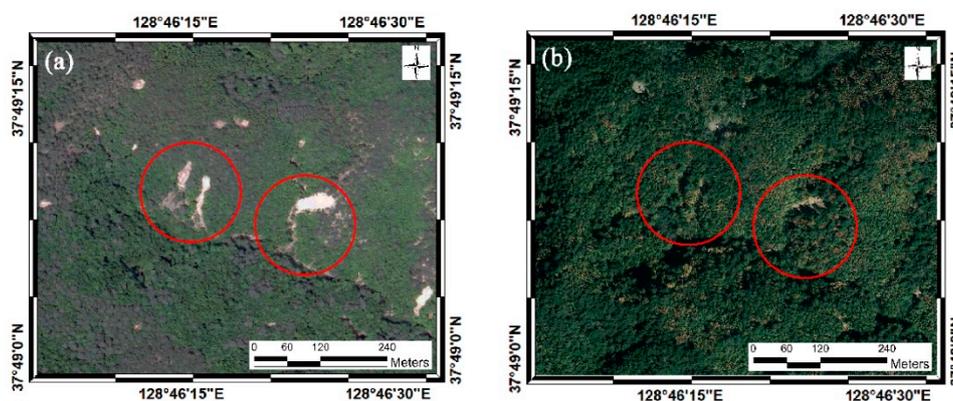


Figure 2. Landslide area of Jumunjin marked by red circle in 2008 (a) and 2014 (b) (Daum map) [2].

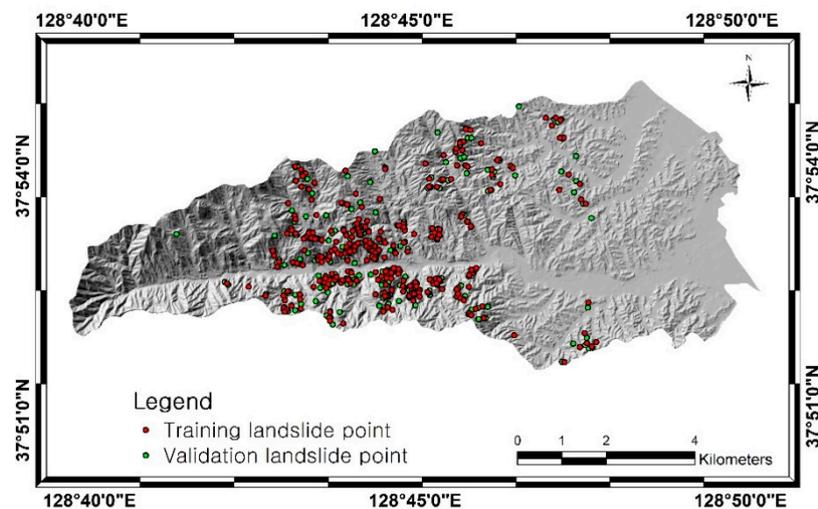


Figure 3. Landslide point of Jumunjin area marked by green circle on hill shade map.

To ensure validity, the parameters were selected by calculating the FR of all the calculated factors using the FR method. A similar procedure was used for Jumunjin area in Gangneung-si, Gangwon-do, which identified a total of 548 landslides that occurred in the area. The terrain data were interpolated using a DEM with a 1:5000 digital topographic map. The topographical factors extracted from the DEM of 10 × 10 m resolution for Jumunjin area were slope, slope direction, maximum curvature, lateral curvature, convexity, structure, surface area, central slope position (MSP), and topographic location index (TPI), which indicated the amount of flow aggregation and the topographic wetting index (TWI) [36–38]. The rock type can affect landslide occurrence because of the differences in the

rock strength and structural characteristics. Joints in the rock can contribute to landslides in areas that are susceptible to their occurrence. We used the distribution of rocks and a 1:25,000 monolayer on a 1:25,000 geological map [39]. Land cover, such as terrain and hydrological patterns, depends on the soil components. The soil components were used to produce soil covering at 1:5000 soil depths, because they can be a contributing factor of landslides caused by rainfall. In addition, the stability of a slope is affected by the distribution of plants on the slope [40]. Therefore, the clinical classification, grade, severity, and density were determined using factors such as forest type and the diameter, density, and age of trees in the forest (Table 1) [41,42]. The landslide conditioning factors were divided into a numerical and categorical factor. The numerical factor has continuous value for each class so we can select the class, for example slope, convexity, surface area, etc. The categorical factor were divided class by its category, for example aspect, forest factor, geology, etc.

**Table 1.** The input factors used in this study.

Category	Factors	Data Type	Scale	Source					
DEM	Topographic factors	Slope	Grid	1:5000	National Geographic Information Institute (NGII)				
		Aspect							
		Maximum curvature							
		Profile curvature							
		Convexity							
	Hydrologic factors	Texture							
		Surface area							
		Mid-slope position (MSP)							
		Terrain ruggedness index (TRI)							
		Topographic position index (TPI)							
Soil map	Land-cover Material	Polygon	1:5000	National Academy of Agricultural Science (NAAS)					
					Flow accumulation				
Forest map	Forest type Forest age Forest density Forest diameter	Polygon	1:5000	Korea Forest Research Institute (KFRI)					
					Topographic wetness index (TWI)				
					Geology	Lithology Distance from fault	Polygon	1:25,000	Korean Institute of Geoscience and Mineral Resources (KIGAM)

### 3. Method

The first step to make the landslide susceptibility map is analyzed the correlation between landslide point and landslide occurrence using aerial photographs. We surveyed the locations of landslides in the aerial photographs of Gangneung-si in Gangwon-do and selected Jumunjin area in Gangneung-si as study areas because they are strongly affected by landslides. A total of 548 landslides were identified using an analysis of aerial photographs; 50% of these were designated as training data and 50% as validation data. Therefore, the same number non-landslide pixels was randomly sampled from the free-landslide area, where landslide pixels were assigned a value of 1 and non-landslide pixels were assigned a value of 0. To construct the landslide susceptibility maps and evaluate their performance, the landslide inventory and landslide conditioning factor maps were converted into ASCII (American Standard Code for Information Interchange) format. ArcGIS was used to convert all of the input data into ASCII data. The transformed data were analyzed using SPSS; the number of data per factor was 1,046,756. Then, landslide training data and landslide conditioning factors were analyzed while using the decision tree method to calculate the landslide susceptibility index and building a landslide susceptibility map in the study area (Figure 4).

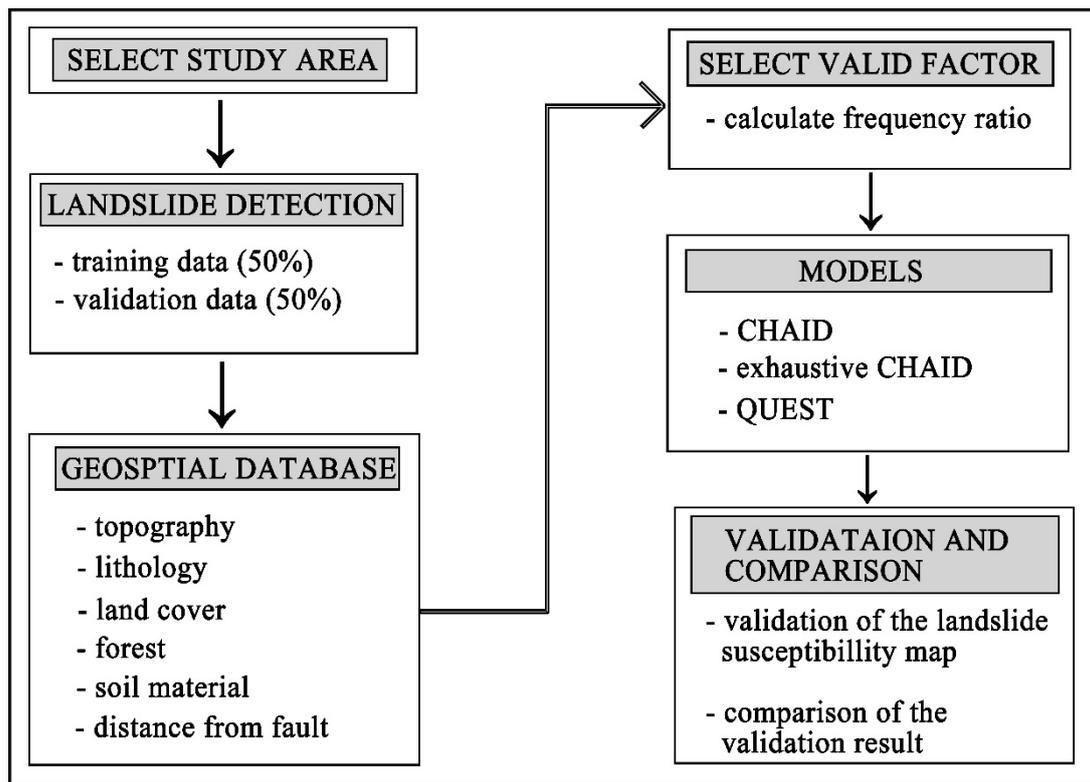


Figure 4. Workflow in this study.

Decision trees are an analytical technique used to perform a decision analysis. They are used to search for and model the relationships, patterns, and rules that exist in large datasets. The structure of a decision tree consists of nodes, starting with the root node and continuing to the child nodes until each branch reaches the end node based on separation criterion, a stopping rule, pruning, and so on. The node at the end of the tree, where no further branching occurs, is called the end node, and the distance from the root node to the terminal node is referred to as the depth [16]. The process of decision tree analysis proceeds in order from decision tree formation through pruning feasibility evaluation and interpretation to prediction [43,44]. Decision tree analysis using dependent and independent variables. In this research, the dependent variable data used is landslide inventory while the independent data used is landslide conditioning factors. In this study, we used three algorithms from the decision tree method to analyse landslide susceptibility map that is Chi-square automatic interaction detection (CHAID), exhaustive CHAID, Quick, Unbiased, and Efficient Statistical Tree (QUEST).

Chi-square automatic interaction detection (CHAID) is an algorithm that performs Dodge separation using the chi-square test (categorical target variable) or F-test (continuous target variable). The CHAID algorithm uses the Pearson's chi-square statistic or likelihood ratio chi-square statistic as a separation criterion when the target variable is categorical. Here, the likelihood ratio squared statistic can be used if the target variable is an ordered or pre-grouped continuous-type variable [43–46]. The exhaustive CHAID was modified by applying the basic algorithm of CHAID. The exhaustive CHAID algorithm continues to merge categories, regardless of their importance until each of the two categories remains for each predictor. If you need to analyze large amounts of data or variables, it can take a long time. Finally, it takes more time to calculate branches that are considered to be more important than when calculating common branches [47–49]. Both CHAID and exhaustive CHAID algorithms consist of three steps: merging, splitting, and stopping. Splitting and stopping steps in the Exhaustive CHAID algorithm are the same as those in CHAID. But, merging step uses an exhaustive search procedure to merge any similar pair until only a single pair remains.

Quick, Unbiased, and Efficient Statistical Tree (QUEST) is an algorithm that uses a statistical test method and a binary division decision tree algorithm for classification and data mining. The QUEST tree growth process consists of selecting the partition predictor, selecting and stopping the partition point for the selected predictor. This algorithm only considers univariate partitioning. It performs an edge separation by identifying a separation variable and a separation point in a selected separation parameter [50]. As the separation criterion, the continuous probability variable of the analysis of variance (ANOVA) F-statistic is calculated, and the categorical variable is selected as that with the least significant probability by calculating the significance probability of the chi-square test statistic.

The decision tree used nominal (categorical) data as dependent variable, then Equations (1) and (2) of Pearson chi-squared will be used [51].

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \tag{1}$$

where,

$$n_{ij} = \sum_{n \in D} f_{nI}(x_n = i \cap y_n = j), m_{ij} = \frac{n_i \cdot n_j}{n_{..}} \tag{2}$$

$n_{ij}$ , is the observed cell frequency and  $m_{ij}$ , is the estimated expected cell frequency for  $(x_n = i, y_n = j)$  following the independence model. The corresponding  $p$  value given by  $p = Pr(x_{de} > x^2)$  [52].

#### 4. Results

We analyzed the relationships between the landslide occurrence sites and landslide occurrence factors in the study area. We randomly selected 279 sites (i.e., half of the 548 landslide occurrence sites) for the verification process. The numeric data was reclassified into five classes by quantile for ease of visual interpretation. The landslide susceptibility of the three algorithms was compared with the training data and the susceptibility was verified by comparing with validation data. The results from CHAID, exhaustive CHAID, and QUEST analyses were calculated by selecting the effective eye factor by calculating the FR, i.e., the rate of occurrence of landslides according to the grade of the element. Then, we recategorized the value of each element based on the results from the analysis and created a new factor map (Figure 5 and Figure S1). The resolution of the factor map is  $10 \times 10$  m, which is the same resolution as DEM, because it is extracted from DEM.

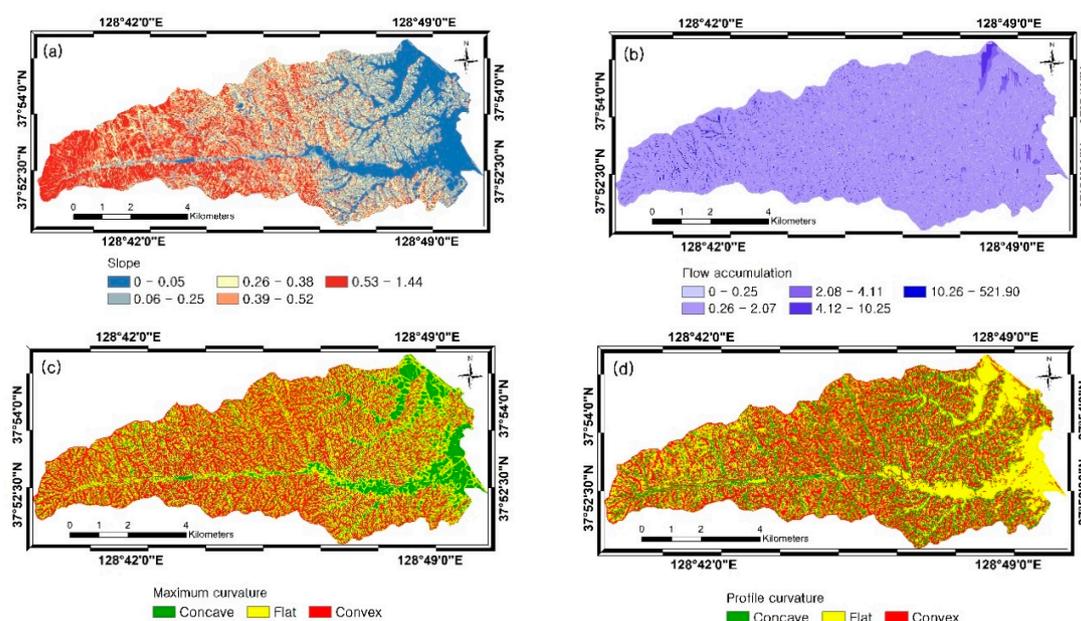


Figure 5. Cont.

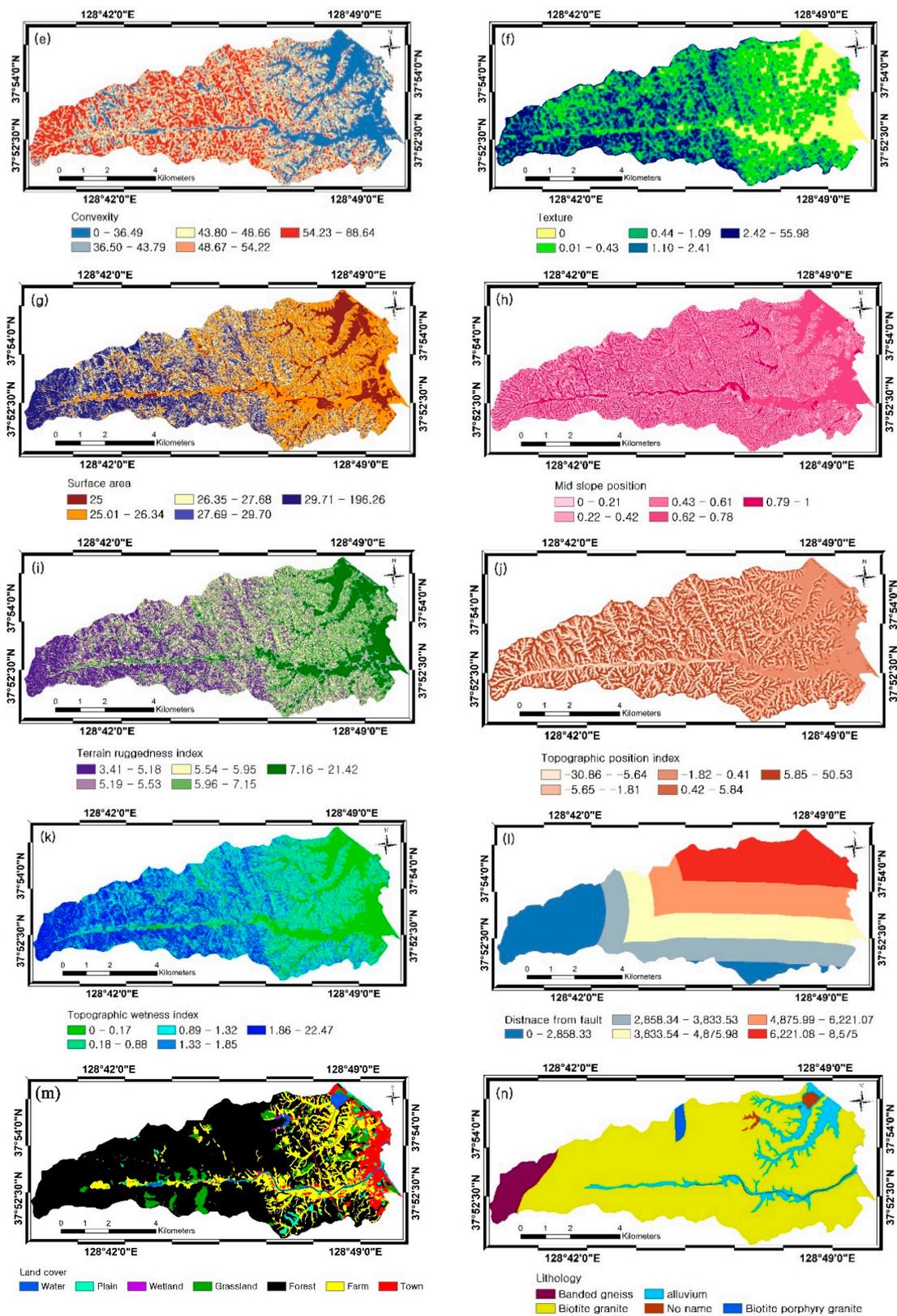
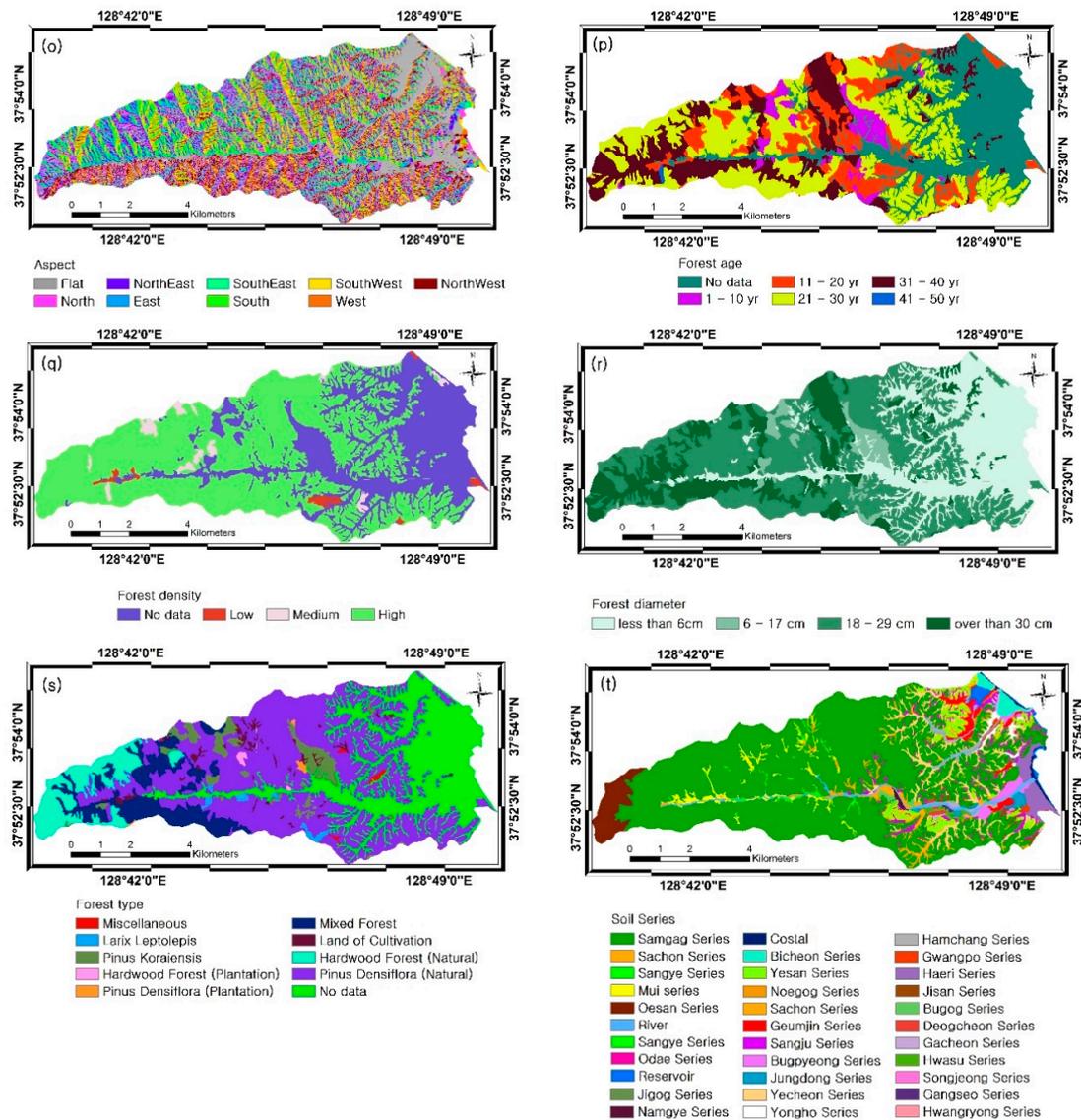


Figure 5. Cont.



**Figure 5.** Spatial database of factors in Jumunjin area, slope (a), flow accumulation (b), maximum curvature (c), profile curvature (d), convexity (e), texture (f), surface area (g), mid-slope position (h), terrain ruggedness index (i), topographic position index (j), topographic wetness index (k), distance from fault (l), land cover (m), lithology (n), aspect (o), forest age (p), forest density (q), forest diameter (r), forest type (s), and soil material (t).

The results for each element are shown in the Table 2 below. In the Table 2, % landslide (+) means the percentage of landslide points in the class and the % domain (+) means the percentage of the total area occupied by each class.

**Table 2.** The Calculated frequency ratio in Jumunjin area.

Factor	Class	% Landslide (+)	% Domain (+)	FR Value
aspect	Flat	13.44	10.82	1.24
	North	15.42	10.37	1.49
	NorthEast	15.42	12.03	1.28
	East	7.51	11.00	0.68
	SouthEast	5.14	11.39	0.45
	South	9.09	11.23	0.81
	SouthWest	8.30	11.04	0.75
	West	12.25	10.80	1.13
	NorthWest	13.44	11.33	1.19

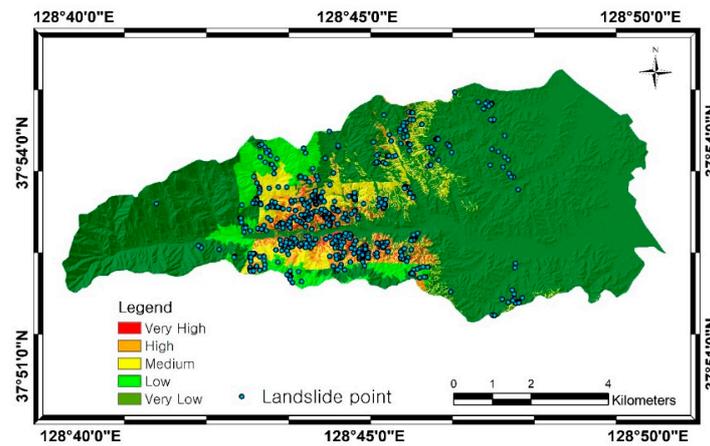
Table 2. Cont.

Factor	Class	% Landslide (+)	% Domain (+)	FR Value
convexity	0–36.49	0.73	19.68	0.04
	36.50–43.79	10.95	19.25	0.57
	43.80–48.66	20.44	19.72	1.04
	48.67–54.22	30.29	20.87	1.45
	54.23–88.64	37.59	20.48	1.84
	0.25–2.07	33.58	22.86	1.47
	2.08–4.11	22.99	21.54	1.07
	4.12–10.25	13.87	18.02	0.77
mid slope position	10.26–521.90	10.22	17.44	0.59
	0–0.21	29.56	19.74	1.50
	0.43–0.61	20.80	19.36	1.07
	0.62–0.78	9.85	20.75	0.47
slope	0.79–1	14.60	20.33	0.72
	0–0.05	1.82	19.90	0.09
	0.06–0.25	8.03	19.91	0.40
	0.39–0.52	28.10	19.80	1.42
surface area	0.53–1.44	47.45	19.91	2.38
	25	0.73	12.83	0.06
	25.01–26.34	16.79	37.35	0.45
	26.35–27.68	17.52	19.50	0.90
texture	29.71–196.26	33.58	13.82	2.43
	0	0.36	13.90	0.03
	0.01–0.43	9.12	34.28	0.27
	0.44–1.09	19.71	18.62	1.06
tpi	1.10–2.41	32.48	18.15	1.79
	–30.86–5.64	12.41	19.12	0.65
	–5.65–1.81	14.96	19.14	0.78
	–1.82–0.41	10.22	20.08	0.51
	0.42–5.84	30.29	20.97	1.44
	5.85–50.53	32.12	20.69	1.55
	5.19–5.53	8.03	21.03	0.38
	5.54–5.95	14.23	20.31	0.70
	5.96–7.15	29.93	20.05	1.49
twi	7.16–21.42	46.72	19.48	2.40
	0–0.17	44.89	19.04	2.36
	0.89–1.32	18.98	20.94	0.91
	1.33–1.85	11.68	20.07	0.58
Lithology	1.86–22.47	0.36	18.80	0.02
	Biotite granite	100	83.91	1.19
	Soil			
	Samgag Series	95.24	67.04	2.33
	Sangye Series	0.36	0.59	0.62
Forest type	River	0.36	2.08	0.18
	Yesan Series	0.36	2.27	0.16
	Yecheon Series	1.09	4.84	0.22
	Pinus Koraiensis	6.57	4.57	1.44
	No data	0.73	28.50	0.03
Forest age	No data	0.73	30.28	0.02
	21–30 yr	44.89	32.40	1.39
	31–40 yr	20.80	18.07	1.15
Forest diameter	less than 6 cm	1.08	30.28	0.04
	18–29 cm	67.80	46.91	1.45
	over than 30 cm	21.42	18.21	1.18
Forest density	No data	8.39	34.87	0.24
	Medium	2.19	2.06	1.07
Land cover	Farm	0.36	16.73	0.02
	Grassland	16.06	5.22	3.08
	6221.08–8575	10.58	20.31	0.52
	flat	16.06	31.28	0.51
	convex	51.82	38.08	1.36
	SgE3	1.82	2.70	0.68

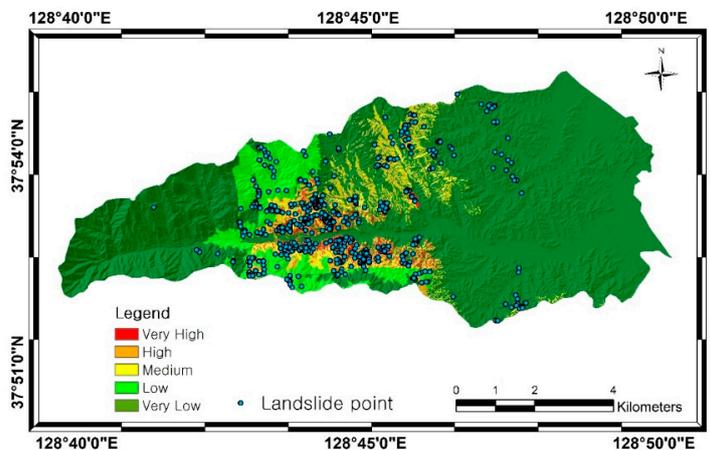
Table 2. Cont.

Factor	Class	% Landslide (+)	% Domain (+)	FR Value
Forest type	PK	6.57	4.57	1.44
	D	74.45	46.72	1.59
	PL	5.47	0.74	7.36
	99	0.73	28.50	0.03
	PD	1.09	0.21	5.22
	M	11.68	9.81	1.19
Forest age	0	0.73	30.28	0.02
	1	7.66	4.59	1.67
	2	25.91	14.51	1.79
	3	44.89	32.40	1.39
	4	20.80	18.07	1.15
Forest diameter	0	1.08	30.28	0.04
	1	9.71	4.59	2.12
	2	67.80	46.91	1.45
	3	21.42	18.21	1.18
Forest density	0	8.39	34.87	0.24
	C	89.42	61.96	1.44
	B	2.19	2.06	1.07
Land cover	200	0.36	16.73	0.02
	300	83.58	67.73	1.23
	400	16.06	5.22	3.08
Distance from Fault	1	3.28	19.57	0.17
	2	30.29	19.80	1.53
	3	41.61	20.15	2.07
	4	14.23	20.18	0.71
	5	10.58	20.31	0.52
maximum curvature	concave	18.61	30.25	0.62
	flat	32.48	37.03	0.88
	convex	48.91	32.72	1.49
profile curvature	concave	32.12	30.64	1.05
	flat	16.06	31.28	0.51
	convex	51.82	38.08	1.36

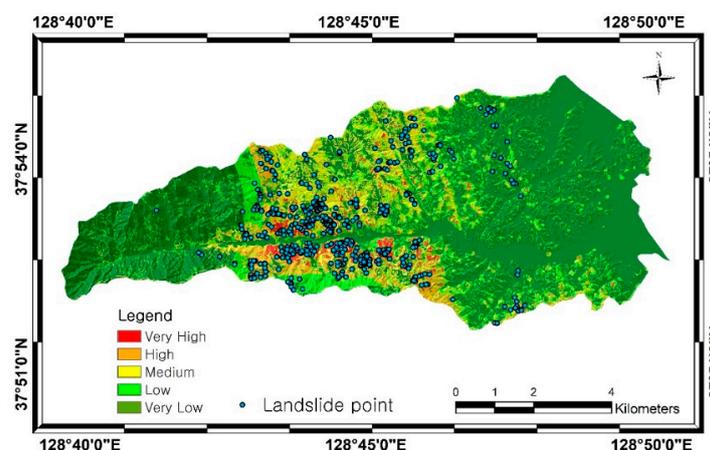
The range of CHAID values was 430–4248. Using these calculated values, we generated a map of landslide susceptibility indices classified into the five stages using the natural break method. The blue area indicates low landslide susceptibility, and the red area shows where the landslide is susceptibility is high. Spatial analysis was performed by referring to the map of landslide susceptibility that is generated using the three analytical methods. Most of the five rivers in Jumunjin area are located in mountainous areas, which had the highest landslide susceptibility, while the coastal areas showed the lowest landslide susceptibility. Landslide susceptibility should effectively predict landslide area and can be used to validate existing landslide location data. Therefore, a validity test was performed using the analysis results for landslide susceptibility. The landslide susceptibilities that were generated by the three algorithms were compared with the initial classified training data, and the generated susceptibilities were verified using the validation data. For this, 547 landslide occurrence points were randomly classified into 50% training data and 50% validation data. The training data were used for processing and the validation data were used for validation. A quantitative comparison among the three algorithms was performed using the AUC method to confirm the processing results. Figures 6–8 show the degree of landslide susceptibility using the CHAID, exhaustive CHAID, and QUEST rating values.



**Figure 6.** Landslide susceptibility map of Chi-square automatic interaction detection (CHAID) algorithm. red area means Landslide susceptibility is very high, orange area means high, yellow area means medium, light green area means low, dark green area means very low.



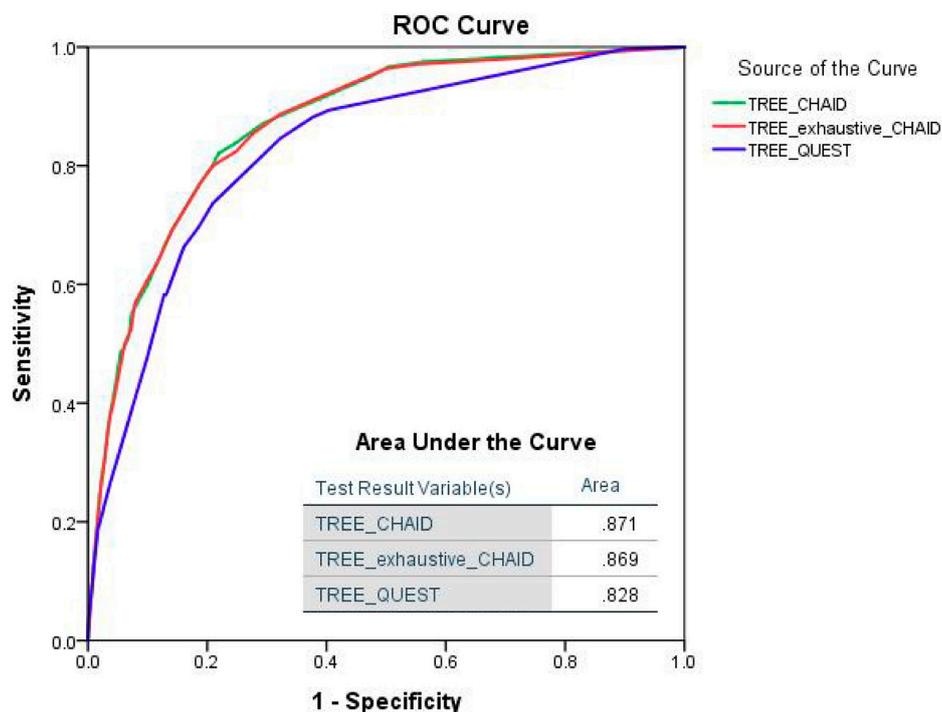
**Figure 7.** Landslide susceptibility map of exhaustive CHAID algorithm. red area means Landslide susceptibility is very high, orange area means high, yellow area means medium, light green area means low, dark green area means very low.



**Figure 8.** Landslide susceptibility map of QUEST algorithm. red area means Landslide susceptibility is very high, orange area means high, yellow area means medium, light green area means low, dark green area means very low.

The algorithm with the highest AUC was considered to be the best algorithm for the study. The AUC value can be obtained by plotting the Receiver operating characteristic (ROC) curve and

calculating the area of the curve. In the ROC curve, the  $x$  and  $y$  axes represent True Positive Rate (TPR) and False Positive Rate (FPR), respectively, where TPR is the rate at which the true value is correctly predicted, and FPR is the rate at which false is predicted as true. TPR and FPR are inversely related to each other. In this study, TPR is the relationship between landslide susceptibility and landslide points, and FPR is the relationship between landslides and landslide points. This curve was the result of comparing the training data with landslide susceptibility. Figure 9 shows that the CHAID algorithm had the highest AUC value (0.871), followed by the exhaustive CHAID (0.869) and QUEST algorithms (0.828). The accuracies of the CHAID, exhaustive CHAID, and QUEST algorithms were 87.1%, 86.9%, and 82.8%, respectively; the CHAID algorithm produced the best estimate of susceptibility to landslides. The results show that the study area has a very small difference of 0.2% between CHAID and exhaustive CHAID. This is because both of the algorithms use chi-square test and f-test as the basic algorithm.



**Figure 9.** ROC curve result of Jumunjin area CHAID (Green), exhaustive CHAID (Red), QUEST (Blue).  $x$ -axis means true positive rate,  $y$ -axis means false positive rate.

## 5. Discussion

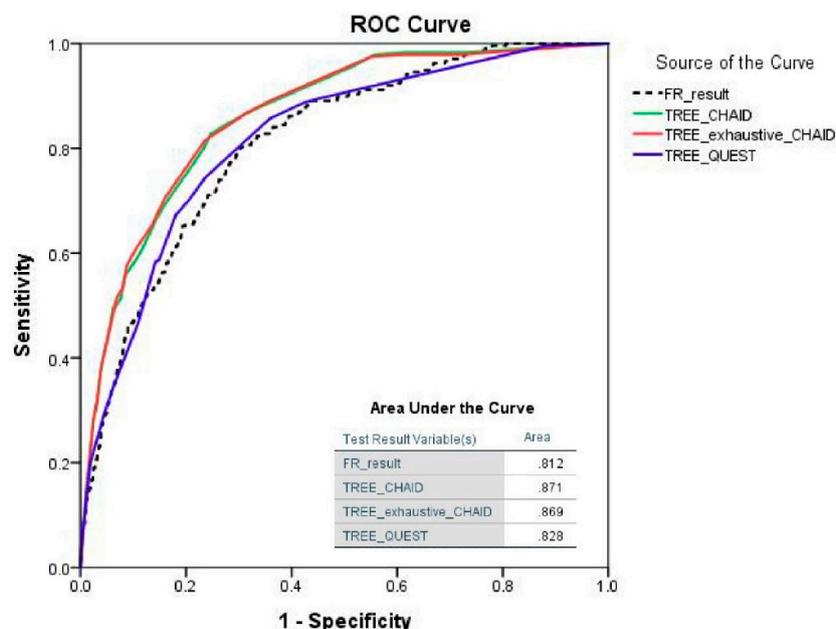
In this study, we used digital aerial photographs that were taken at high resolution to identify landslides. It is very difficult to distinguish small landslides in the study area using satellite imagery. The use of aerial photo analysis avoids time-consuming and costly field surveys. Landslides were identified using pattern classification, and the factors that were related to landslides were identified and analyzed together with the landslide location. The analysis used 20 factors (soil, hydrological features, and geological and forest map data). The algorithm for creating landslide susceptibility used CHAID, exhaustive CHAID, and QUEST algorithms in decision tree models.

In Korea, particularly in the rainy season, many landslides result from heavy rainfall over a short period. Therefore, it is necessary to analyze the susceptibility to landslides by selecting the appropriate factors that are related to landslides and classifying them by examining the relationships between these factors and landslide location. Twenty factors, including topography, hydrology, soil map, and clinical map, were selected to analyze landslide probability using CHAID, exhaustive CHAID, and QUEST algorithms in decision tree models. These algorithms are used in various fields as sophisticated

modeling techniques and were applied to determine the effects of environmental factors on landslides and landslide susceptibilities. The landslide susceptibility maps were divided into five grades (very low, low, medium, high, and very high) for the ease of visual interpretation. The maps were verified against training and validation data. Specifically, 548 landslides were randomly divided into two sets of data; 50% of the landslides were used for training and the remaining 50% were used for validation using the ROC curve. The CHAID algorithm had the highest AUC (0.871), followed by the exhaustive CHAID algorithm (0.869), and the QUEST algorithm (0.828). The CHAID algorithm had the highest decision time (DT) accuracy (87.1%), followed by the exhaustive CHAID algorithm (86.9%), and the QUEST algorithm (82.8%). Hence, the accuracy was greater than 80% for all algorithms, and all of the algorithms were valid. The results from this study showed that slope, topographical solidity index, surface area, and convexity were positively correlated with landslide susceptibility in both of the study areas. These factors are considered related to landslide susceptibility because they increase the instability of the slope as the size of the area increases. In contrast, the TWI and the flow accumulation are negatively correlated with landslide susceptibility. The TWI and the flow aggregation correspond to hydrologic factors, and landslide susceptibility increases due to decreased cohesion from moisture as the slope becomes drier. Finally, our study revealed that those factors that increase the instability of the slope and those with less effect on the hydrological factors will increase landslide susceptibility.

## 6. Conclusions

Landslide susceptibility maps are of much interest in the landslide research community to improve performance. Map quality is controlled in an adapted way and new machine learning techniques have proven to be effective in terms of predictive performance. Therefore, in this study, we investigated the application of three decision tree method CHAID, exhaustive CHAID, and QUEST to the assessment of landslide susceptibility. According to the literature, such investigations are rare. Especially, it is based on a case study at Jumunjin area. The results of this study confirm that the performance of the landslide mapping is improved by using the machine learning ensemble. For comparison, we also considered traditional model frequency ratios with an AUC of 0.812 (Figure 10).



**Figure 10.** ROC curve result of Jumunjin area, FR (Dotted), CHAID (Green), exhaustive CHAID (Red), QUEST (Blue). *x*-axis means TPR, *y*-axis means FPR.

Decision tree model predictions were improved by 5.9% for CHAID, 5.7% for exhaustive CHAID and 1.6% for QUEST. These results are reasonable because the techniques used in the classifier ensemble

framework can reduce bias and dispersion and over-avoid problems with the primary classifier to improve predictability. Among the three decision tree models, the CHAID algorithm provided the greatest improvement. The main advantage of the three decision tree models is that it automates the process of examining multiple databases to gather valuable information. As a result of the analysis, the landslide occurrence point, which is a dependent variable, was classified as 1 and the point without landslide was classified as 0. So, the model assessed the landslide risk and quantitatively analyzed the correlation between landslide and various factors. Accuracy has increased by up to 5.9% over previous models. On the other hand, Decision Tree model has disadvantage. The disadvantages are that the preprocessing can take a long time because it has many procedures. Landslide susceptibility maps can help you to make decisions about building choices and planning. This map can be the basis for landslide risk management studies. Decision tree modeling combined with RS and GIS spatial data provided reasonably accurate landslide forecasting. In this study, we only examined Gangneung-si in Gangwon-do. In future studies, to provide data that are more widely applicable, the occurrence of landslides in other parts of Korea should be examined. Further research will enable effective monitoring and the prevention of landslides in susceptible areas if the appropriate factors are selected for assessing landslide occurrence.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-4292/10/10/1545/s1>, Figure S1: Spatial database of factors in Jumunjin area, slope (a), flow accumulation (b), maximum curvature (c), profile curvature (d), convexity (e), texture (f), surface area (g), mid-slope position (h), terrain ruggedness index (i), topographic position index (j), topographic wetness index (k), distance from fault (l), land cover (m), lithology (n), aspect (o), forest age (p), forest density (q), forest diameter (r), forest type (s), soil material (t).

**Author Contributions:** S.-J.P. processed decision tree models for estimation of landslide susceptibility and wrote the paper. C.-W.L. organized the paper. S.L. suggested the idea and interpreted the results. M.-J.L. designed the experiments. All of the authors contributed to the writing of each part.

**Funding:** This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Ministry of Science and ICT, and The National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2017R1A2B4003258) and (NRF-2018R1D1A1B07041203).

**Acknowledgments:** The first author acknowledges the support of the Kangwon National University Scholarship for his master degree studies.

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

## References

- Gariano, S.L.; Rianna, G.; Petrucci, O.; Guzzetti, F. Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale. *Sci. Total Environ.* **2017**, *596*, 417–426. [[CrossRef](#)] [[PubMed](#)]
- Daum Map. Available online: <http://map.daum.net/> (accessed on 20 December 2017).
- Tsangaratos, P.; Ilija, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* **2016**, *13*, 305–320. [[CrossRef](#)]
- Reichenbach, P.; Rossi, M.; Malamud, B.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
- Van Westen, C.J.; Van Asch, T.W.J.; Soeters, R. Landslide hazard and risk zonation—Why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [[CrossRef](#)]
- Lee, S.; Choi, J.; Woo, I. The effect of spatial resolution on the accuracy of landslide susceptibility mapping: A case study in Boun, Korea. *Geosci. J.* **2004**, *8*, 51–60. [[CrossRef](#)]
- Hong, H.; Naghibi, S.A.; Pourghasemi, H.R.; Pradhan, B. GIS-based landslide spatial modeling in Ganzhou city, China. *Arab. J. Geosci.* **2016**, *9*, 1–26. [[CrossRef](#)]
- Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
- Lee, M.J.; Park, I.; Won, J.S.; Lee, S. Landslide hazard mapping considering rainfall probability in Inje, Korea. *Geomat. Nat. Hazards Risk* **2016**, *7*, 424–446. [[CrossRef](#)]
- Althuwaynee, O.F.; Pradhan, B.; Lee, S. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. *Int. J. Remote Sens.* **2016**, *37*, 1190–1209. [[CrossRef](#)]

11. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2007**, *4*, 41–47. [[CrossRef](#)]
12. Steger, S.; Brenning, A.; Bell, R.; Petschko, H.; Glade, T. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* **2016**, *262*, 8–23. [[CrossRef](#)]
13. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Honh, H.; Tien Bui, D.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]
14. Feizizadeh, B.; Shadman, R.M.; Jankowski, P.; Blaschke, T. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. *Comput. Geosci.* **2014**, *73*, 208–221. [[CrossRef](#)] [[PubMed](#)]
15. 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.* **2011**, *18*, 471–493. [[CrossRef](#)]
16. Park, I.; Lee, J.; Lee, S. Ensemble of ground subsidence hazard maps using fuzzy logic. *Cent. Eur. J. Geosci.* **2014**, *6*, 207–218. [[CrossRef](#)]
17. Dehnavi, A.; Aghdam, I.N.; Pradhan, B.; Varzandeh, M.H.M. A new hybrid model using step-wise weight assessment ratio analysis (SWAM) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena* **2015**, *135*, 122–148. [[CrossRef](#)]
18. Nasiri Aghdam, I.; Varzandeh, M.H.M.; Pradhan, B. Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environ. Earth Sci.* **2016**, *75*, 553–572. [[CrossRef](#)]
19. 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](#)]
20. Conoscenti, C.; Ciaccio, M.; Caraballo-Arias, N.A.; Gomez-Gutierrez, A.; Rotigliano, E.; Agnesi, V. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Bence River basin (western Sicily, Italy). *Geomorphology* **2015**, *242*, 49–64. [[CrossRef](#)]
21. 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](#)]
22. Pham, B.T.; Tien Bui, D.; Dholakia, M.B.; Prakash, I.; Pham, H.V. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotech. Geol. Eng.* **2016**, *34*, 1807–1824. [[CrossRef](#)]
23. Pradhan, B. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *J. Indian Soc. Remote Sens.* **2010**, *38*, 301–320. [[CrossRef](#)]
24. Lee, S.; Park, I. Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines. *J. Environ. Manag.* **2013**, *127*, 166–176. [[CrossRef](#)] [[PubMed](#)]
25. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* **2013**, *68*, 1443–1464. [[CrossRef](#)]
26. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model Softw.* **2010**, *25*, 747–759. [[CrossRef](#)]
27. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [[CrossRef](#)]
28. Tsangaratos, P.; Benardos, A. Estimating landslide susceptibility through a artificial neural network classifier. *Nat. Hazards* **2014**, *74*, 1489–1516. [[CrossRef](#)]
29. Lee, S.; Hong, S.; Jung, H. A support vector machine for landslide susceptibility mapping in Gangwon province, Korea. *Sustainability* **2017**, *9*, 48. [[CrossRef](#)]
30. Tien Bui, D.; Tuan, T.A.; Hoang, N.D.; Thanh, N.Q.; Nguyen, D.B.; Liem, N.V.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458. [[CrossRef](#)]

31. Oh, H.J. Landslide detection and landslide susceptibility mapping using aerial photos and artificial neural networks. *Korean J. Remote Sens.* **2010**, *26*, 47–57.
32. Lee, S.; Kim, G.; Lee, C. Preliminary study for tidal flat detection in Yeongjong-do according to tide level using Landsat images. *Korean J. Remote Sens.* **2016**, *32*, 639–645. [[CrossRef](#)]
33. Lee, S.K.; Lee, C.W. Predicting the hazard area of the volcanic ash caused by Mt. Ontake Eruption. *Korean J. Remote Sens.* **2014**, *30*, 777–786. [[CrossRef](#)]
34. Jang, J.; Eom, J.; Cheong, D.; Lee, C. Monitoring of the Estuary Sand Bar Related with Tidal Inlet in Namdaecheon Stream using Landsat Imagery. *Korean J. Remote Sens.* **2017**, *33*, 481–493. [[CrossRef](#)]
35. Eom, J.; Lee, C. Analysis on the area of deltaic Barrier Island and suspended sediments concentration in Nakdong River using satellite images. *Korean J. Remote Sens.* **2017**, *33*, 201–211. [[CrossRef](#)]
36. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122. [[CrossRef](#)]
37. Iwahashi, J.; Pike, R.J. Automated classifications of topography from dams by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* **2007**, *86*, 409–440. [[CrossRef](#)]
38. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [[CrossRef](#)]
39. Oh, C.Y.; Kim, K.T.; Choi, C.U. Analysis of landslide characteristics of Inje area using SPOT5 images and GIS analysis. *Korean J. Remote Sens.* **2009**, *25*, 445–454.
40. Schwarz, M.; Preti, F.; Giadrossich, F.; Lehmann, P.; Or, D. Quantifying the role of vegetation in slope stability: A case study in Tuscany (Italy). *Ecol. Eng.* **2010**, *36*, 285–291. [[CrossRef](#)]
41. Schmidt, K.M.; Roering, J.J.; Stock, J.D.; Dietrich, W.E.; Montgomery, D.R.; Schaub, T. The variability of root cohesion as an influence on shallow landslide susceptibility in the Oregon Coast Range. *Can. Geotech. J.* **2001**, *38*, 995–1024. [[CrossRef](#)]
42. Chi, K.H.; Shin, J.S.; Park, N.W. Quantitative Analysis of GIS-based Landslide Prediction Models Using Prediction Rate Curve. *Korean J. Remote Sens.* **2001**, *17*, 199–210.
43. Loh, W.Y.; Shih, Y.S. Split selection methods for classification trees. *Stat. Sin.* **1997**, *7*, 815–840.
44. Song, Y.S.; Chae, B.G. Development to Prediction Technique of Slope Hazards in Gneiss Area using Decision Tree Model. *J. Eng. Geol.* **2008**, *18*, 45–54.
45. 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](#)]
46. Yeon, Y.K.; Han, J.G.; Ryu, K.H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [[CrossRef](#)]
47. Peng, L.; Niu, R.Q.; Huang, B.; Wu, X.L.; Zhao, Y.N.; Ye, R.Q. Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology* **2014**, *204*, 287–301. [[CrossRef](#)]
48. 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. [[CrossRef](#)]
49. Xu, M.; Watanachaturaporn, P.; Varshney, P.K.; Arora, M.K. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* **2005**, *97*, 322–336. [[CrossRef](#)]
50. Biggs, D.; Barry, D.V.; Suen, E. A method of choosing multi way partitions for classification and decision trees. *J. Appl. Stat.* **2006**, *18*, 49–62. [[CrossRef](#)]
51. Alkhasawneh, M.S.; Ngah, U.K.; Tay, L.T.; Isa, M.; Ashidi, N.; Al-Batah, M.S. Modeling and testing landslide hazard using decision tree. *J. Appl. Math.* **2014**. [[CrossRef](#)]
52. Baker, S.; Cousins, R.D. Clarification of the use of chi-square and likelihood functions in fits to histograms. *Nucl. Instrum. Methods Phys. Res.* **1984**, *221*, 437–442. [[CrossRef](#)]

