

Article

Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and NaïveBayes Machine-Learning Algorithms

Hamid Reza Pourghasemi ¹, Amiya Gayen ², Sungjae Park ³, Chang-Wook Lee ^{3,*}
and Saro Lee ^{4,5,*}

¹ Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz 71441-65186, Iran; hr.pourghasemi@shirazu.ac.ir

² Department of Geography, University of Gour Banga, Malda 732103, India; mr.amiyagayen@gmail.com

³ Division of Science Education, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341 Korea; tres13@kangwon.ac.kr

⁴ Division of Geoscience Platform, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124 Gwahang-no, Yuseong-gu, Daejeon 305-350, Korea

⁵ Department of Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 305-350, Korea

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

Received: 9 August 2018; Accepted: 12 October 2018; Published: 15 October 2018



Abstract: The occurrence of landslide in the hilly region of South Korea is a matter of serious concern. This study tries to produce landslide susceptibility maps for Jumunjin Country in South Korea. Three machine learning algorithms, namely Logistic Regression (LR), LogitBoost (LB), and NaïveBayes (NB) are used, and their final model outcomes are compared to each other. Firstly, a landslide inventory map and the associated input data layers of the landslide conditioning factors were developed based on field verification, historical records, and high-resolution remote-sensing data in the geographic information system (GIS) environment. Seventeen landslide conditioning factors were prepared, including aspect, slope, altitude, maximum curvature, profile curvature, topographic wetness index (TWI), topographic positioning index (TPI), distance from fault, convexity, forest type, forest diameter, forest density, land use/land cover, lithology, soil, flow accumulation, and mid slope position. The result showed that the area under the curve (AUC) values of LR, LB, and NB models were 84.2%, 70.7%, and 85.2%, respectively. The results revealed that the LR and LB models produced reasonable accuracy than respect to NB model in landslide susceptibility assessment. The final susceptibility maps would be useful for preliminary land-use planning and hazard mitigation purpose.

Keywords: machine-learning algorithm; Logistic regression; LogitBoost; NaïveBayes; receiver operating characteristics

1. Introduction

A landslide is very complex natural phenomenon that affects human life, natural resources, and property worldwide [1]. Annually, landslides cause 1000 deaths and about USD 4 billion loss of property worldwide [2]. A landslide is described as a ground extent wide movement, which incorporates rock falls, shallow debris flow, and deep slope failures [3]. The most ideal approach to mitigate the magnitude and effect of landslides is related to appropriate disaster management and mitigation plan. Landslide modeling through machine-learning models is considered as an important tool of landslide mitigation and management plan [4]. There are number of factors which are co-related

with the occurrence of landslide such as geomorphology, geology, and hydrological factors [5,6] but all the factors do not have a uniform impact at every event. The impact of those factors are varies from different places in different ways [7]. For that, landslide susceptibility mapping at a regional scale becomes difficult and complex. However, in the last few decades a number of research institutes and governments have attempted to procedure landslide susceptibility maps for disaster prevention [8,9]. Several researchers express many ways to delineate landslide susceptibility zones. For the development of landslide susceptibility maps, a key requirement is the spatial relationship between a landslide and its conditioning factors [10]. In the present context, several researchers have applied numerous models to analysis the spatial relationship and then compared them by the geographic information system (GIS) environment [6,11].

The methods for landslide susceptibility assessment widely classified into three categories, such as quantitative [12], semi-quantitative [13], and qualitative [14]. A few researchers have utilized several quantitative techniques such as bivariate regression analysis [15,16], logistic regression [17–19], multivariate regression analysis [20,21], fuzzy logic [22], artificial neural network [11,23–25], decision tree [26,27], random forest [28], support vector machine [29], and adaptive neuro-fuzzy inference system [17,30–33].

In the present context, a machine-learning algorithm, as a powerful method of data driven tools, applied to learn the spatial relationship between landslide occurrence and landslide effecting factors, and overcome starting with an assumed structural model [34–36]. Statistical based models are required large amounts of data, whereas machine-learning algorithms based models can effectively overcome the limitation of data dependency [9,37].

The aim of this research is the identification and delineation of landslide susceptibility maps (LSM) based on machine-learning algorithms (i.e., LR, LB, NB models) for the Jumunjin Country in South Korea by taking into consideration relevant parameters, grouping them based on their proximity of characters. The outcome of models also compared and validated by the receiver operating characteristic (ROC) curve method for section of the best model in the study area and finally tried to identify the relative importance of landslide conditioning factors by LVQ (Learning Vector Quantization) model [38].

2. Study Area

The Gangneung-city of Gangwon-do is mainly active in agriculture and tourism, and the average annual rainfall is about 1018.7 mm (Figure 1). The mountainous terrain is composed mostly of granite formed in the Mesozoic Jurassic, and the river basin with dense residential area is composed of the alluvial deposits deposited in the 4th Cenozoic era.

Typhoon Rusa in 2002 and typhoon Maemi in 2003 recorded record rainfall of more than 80 mm per hour. These typhoons caused more than 250 deaths and injuries in Korea. Gangwon province, Gangneung city, Gangneung area (70.84 km², population 4219, 1914 households) and Jumunjin-eup (area 60.55 km², population 21,291, 8917 households) are areas where many landslides have occurred due to typhoons Rusa and Mica. In particular, seven landslides of Gangneung city including the study area were cut off by a landslide caused by typhoon Rusa in 2002, and three residents of Samgyunjin-eup were killed and many isolated for more than 10 days.

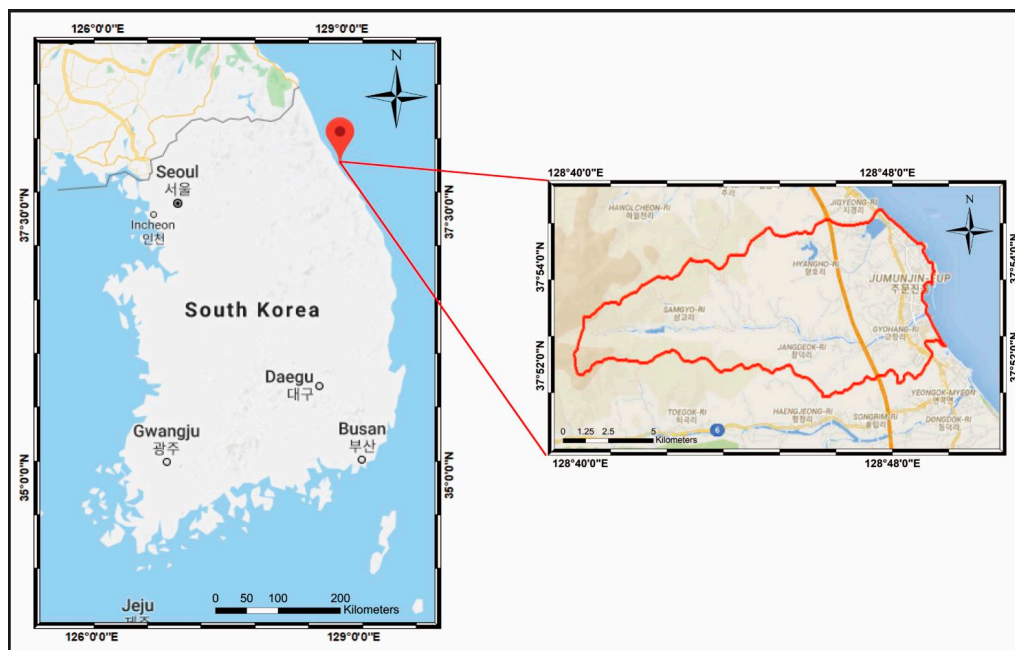


Figure 1. The location of the Jumunjin area, from Google maps.

3. Data Used

Landslide inventory data are considered the most important tools for foreseeing future landslide incidents [31]. An inventory map is required to study the spatial relationship between landslide distribution and the effective factors of landslide. Regarding this investigation, 548 landslide locations were identified in the Jumunjin Country in South Korea (Figure 2). Different methods were applied for the landslide location identification such as inventory reports, high-resolution satellite images, and extensive field surveys. In this study, 17 landslide effective factors are used based on the literature review and data availability [6,22,26,39–43]. These factors include aspect, slope, altitude, maximum curvature, profile curvature, topographic wetness index (TWI), topographic positioning index (TPI), distance from fault, convexity, forest type, forest diameter, forest density, land use and land cover (LULC), lithology, soil, flow accumulation, and mid slope position. The data used in this present study are shown in Table 1.

Table 1. Details of data used in landslide susceptibility assessment.

Data	Sources	Scale/Resolution
Digital elevation model	National Geographic Information Institute (NGII)	1:5000
Satellite image	Daum map	0.5 × 0.5 m
Soil map	National Academy of Agricultural Science (NAAS)	1:5000
Lithology map	Korean Institute of Geoscience and Mineral Resources (KIGAM)	1:25,000
Fault line	Korean Institute of Geoscience and Mineral Resources (KIGAM)	1:25,000

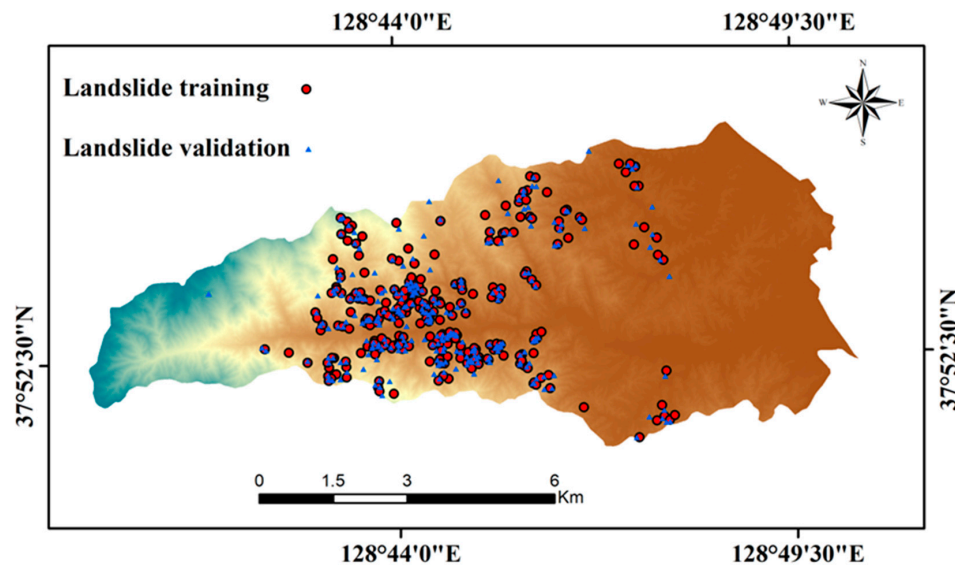


Figure 2. Landslide inventory map with altitude of the study area.

3.1. Aspect

It is emphasized that the aspect is a vital factor in instigating landslide susceptibility mapping [1,39]. Physically, aspect-related factors include orientation of discontinuities controlling landslide, sunlight, wind and precipitation [44–46]. In this work, a digital elevation model (DEM) of 5 m spatial resolution was used to compute the aspect map (Figure 3a).

3.2. Slope Gradient

The slope gradient has considerable influence on slope stability Bednarik [47]. Many researchers have shown that slope gradient is the main factor for landslide susceptibility maps (LSM) [48,49]. Slope map was also developed based on the DEM; this map reveals that the maximum basin area falls under steep to very steep slopes (Figure 3b).

3.3. Altitude

Altitude is another importance conditioning factors because it is affected by several geomorphic and geological processes [48,50]. Generally, landslides have more tendencies to occur at high altitude [45]. In this study area, the altitude ranges from 0.30 m to 956 m from mean sea level (Figure 3c).

3.4. Curvature (Maximum and Profile)

The curvature values represent slope shape and terrain morphology and can affect LSM in several ways as proposed by Haigh and Rawat (2012) [51]. The maximum curvature represents the flow acceleration and erosion and deposition rate, and profile curvature affects the water flow velocity variation down the slope [9,52]. Therefore, total and profile curvature maps were generated by the spatial analysis tool (Figure 3d,e).

3.5. Topographic Wetness Index (TWI)

The topographical wetness index is describe by the flow accumulation theoretical measure, where it affects soil moisture and ground water flow that affect the slope material [53,54]. It is expressed as:

$$TWI = \ln (\alpha / \tan \beta) \quad (1)$$

where, α is the cumulative upslope area drainage through a point and $\tan \beta$ is the slope angle at the point. In this study, the TWI was considered as an important contributing factor (Figure 3f).

3.6. Topographic Positioning Index (TPI)

A positive value of the TPI indicates higher locations comparing to averages and negative values indicates lower locations than the surrounding specified by neighborhood. The TPI maps of the present study area (Figure 3g) show an index value ranging from -30.86 to 50.83 indicating the coexistence of low land (depression, channels and valley); highlands (ridges, and residual hills); and constant slope areas (river plain, and agricultural fields), leading to rough topography over the whole catchment.

3.7. Distance from Fault

The landslide takes place mainly along the faults. Faults are the zones of weakness that not only decrease the rock strength; but also fracturing and unstable slope conditions [55,56]. A Euclidean distance method was adopted to produce the distance to fault factor (Figure 3h).

3.8. Convexity

Convexity also considered as an important parameter for LSM, because slope movement is a natural phenomenon that happens in a gravitational field; a greater the degree of convexity indicates greater susceptibility to slope fails. Thus, convexity is an important factor affecting slope stability [57]. In this area, the convexity ranges from 0 to 80.64 (Figure 3i).

3.9. Forest Factors (Forest Type, Forest Diameter, and Forest Density)

Forest factors such as the types of forest, forest diameter, and forest density are the most importance factors for landslide susceptibility assessment. In the forest type, PK means *Pinus koraiensis* stand, D means *Pinus densiflora* stand, L means Arable land, PL means *Larix kaempferi* stand, PD means artificial *Pinus koraiensis* stand, M means Mixed stand and H means Hardwood stand. The diameter of the tree is less than 6 cm class 0, 6~16 cm is class1, class 2 is less than 18~28 cm, more than 30 cm was classified as class 3 (Figure 3j–l). The landslide occurrence is maximum in areas covered by artificial pine and non-forested area, and has low probability under paper palm trees and artificial rigid pine because the root systems of broad-leaf trees have more soil-holding capacity than needle-leaf trees. The landslide probability value is lower for medium diameter and is higher for non-forest and small diameter trees due to medium diameter trees having more roots and greater capacity to maintain the water and soil-pore pressure during heavy rain periods [23]. In the case of trees age, landslide probability is higher in younger timber non-forested area and lower probability in older trees, again because an older tree has more roots [23].

3.10. Land Use/Land Cover (LULC)

Land use and land cover are affected by human activities and also alterations in the environment. It is also plays a significant role for the slope stability [58]. In this research, LULC map is generated based on field observation and satellite images and classified into seven classes: 100, 200, 300, 400, 500, 600, and 700 (Figure 3m).

3.11. Lithology

The lithology can be regarded as one of the most important factors in the geo-hazard assessment [49,59], because strength, the rocks, soil permeability, and weathering are influence by lithological characteristics [60,61]. The characteristics of the component materials of a slope such as permeability, porosity, and strength, play a crucial role in the stability of the slope, and are mostly different for various lithological units, so, this factor greatly influences the probability of landslide occurrence [58]. In this work, the lithological map (Figure 3n) was prepared by the Geological survey of South Korea with 1: 50,000 (Table 2).

Table 2. Lithological units in the study area.

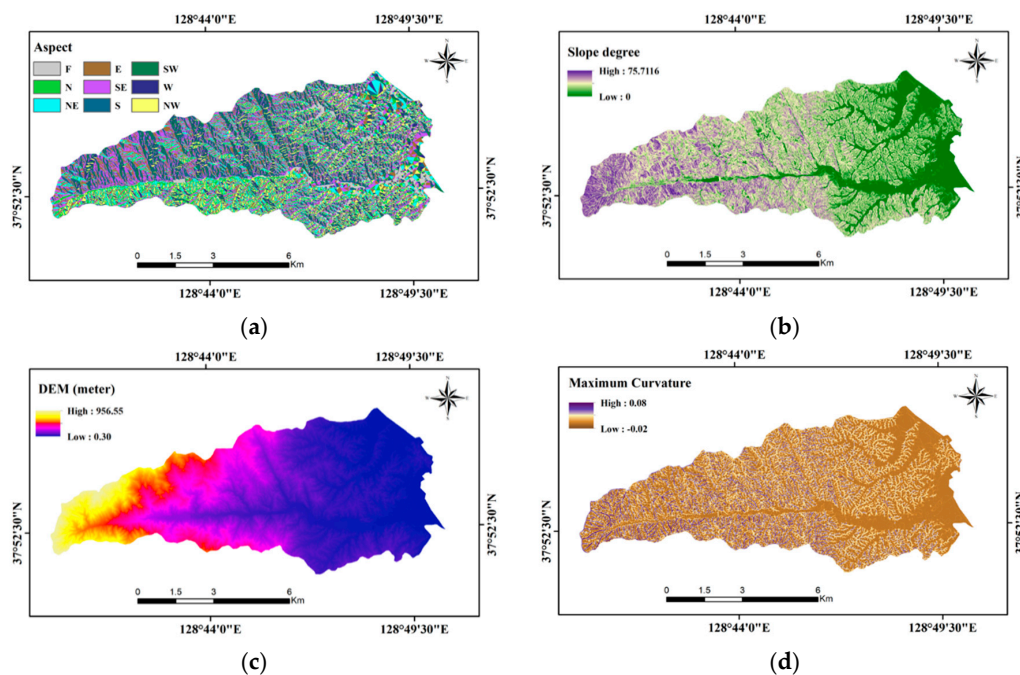
Code	Formation	Lithology	Geological Age
Qa	Alluvium		Quaternary
PCEbgn	Banded gneiss	Quartzite and hornblende	Precambrian
Null	-	-	-
Jpbgr	Porphyritic biotite granite	Porphyritic biotite granite	Jurassic
Jjgr	Jumunjin granite	Jumunjin granite	Jurassic

3.12. Soil

The soil is one of the most important effective factors for landslide assessment (Figure 3o) with the increase of soil depth and the tendency of soil to absorb moisture, so this reduces surface runoff rate. Soil moisture directly affects the slope material, by which pore water pressure diminishes soil stability. On the other hand, shallow soil depth is considered to be more unstable and leading to landslides [62]. In this area, the maximum area falls under the SmF2 soil group having a shallow depth on steep slopes represents vulnerable condition for the occurrence of landslides [62].

3.13. Flow Accumulation

The flow accumulation is the quantity of water, which will move to each pixel from its neighbors and eventually accumulated to it [63]. This is also an important factor for landslide susceptibility assessment [64] (Figure 3p). This map was generated by using the ArcGIS 10.2.2 hydrology tool.

**Figure 3.** Cont.

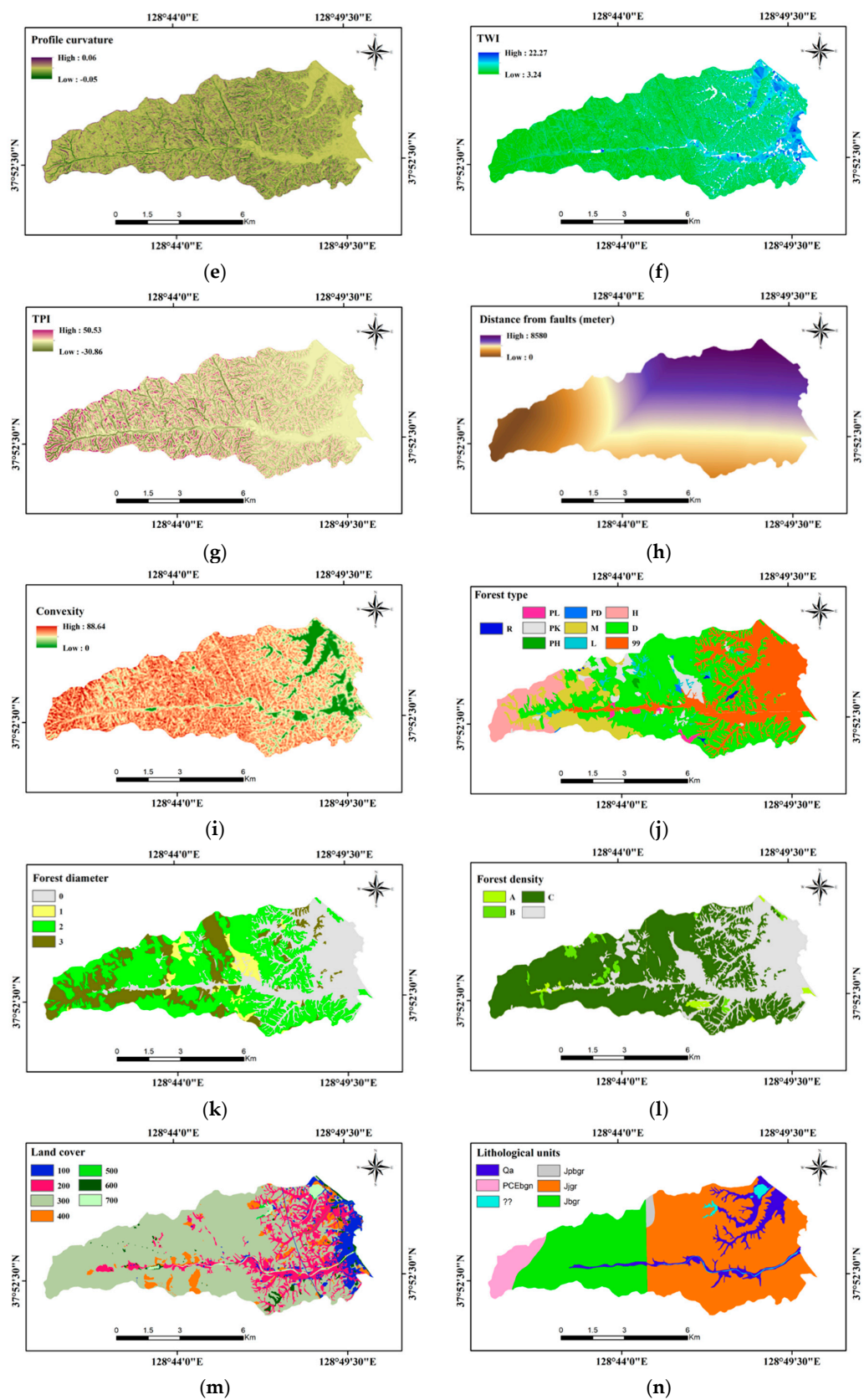


Figure 3. Cont.

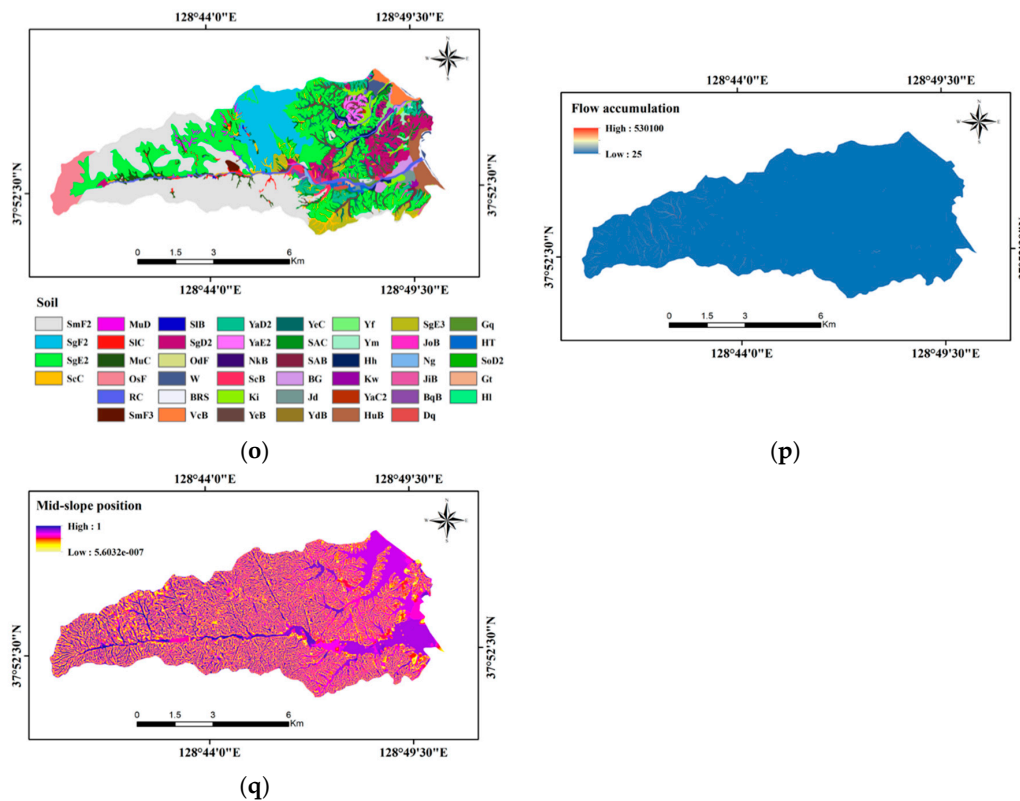


Figure 3. Landslide conditioning factor maps used in this study: (a) aspect; (b) slope; (c) altitude; (d) maximum curvature; (e) profile curvature; (f) topographic wetness index (TWI); (g) topographic positioning index (TPI); (h) distance from fault; (i) convexity; (j) forest types; (k) forest diameter; (l) forest density; (m) land use and land cover (LULC); (n) lithology; (o) soil types; (p) flow accumulation; and (q) mid slope position.

3.14. Mid-Slope Position

The mid-slope position indicates the relative vertical distance from the middle of the slope of the ridge or valley. The larger value means greater distance [65]. This is chosen as an important factor when modeling [66].

4. Multicollinearity of Landslide Effective Factors

Multicollinearity is a statistical analysis in which numbers of independent factors in a multiple regression model are strongly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy [67,68].

Normally, there are two exponents for considering multicollinearity of variables; those are Variance Inflation Factors (VIF) and Tolerance, defined as follows (Equations (2) and (3)):

$$Tolerance = 1 - R_J^2 \quad (2)$$

$$VIF = \left[\frac{1}{Tolerance} \right] \quad (3)$$

where, R_J^2 is the coefficient of determination of a regression of explanatory J on all the other explanatory. A tolerance of less than 0.10 and variance inflation factors (VIF) >5 indicate multicollinearity problems [69,70].

5. Modeling for Landslide Susceptibility Zonation

In the current research, the models of Logistic Regression (LR), LogitBoost (LB), and NaïveBayes (NB) were used, and their results compared. For computing those models first, landslide inventory map were generated. Then, three methods (i.e., LR, LB, and NB) were used to model the spatial correlation between landslide conditioning factors and landslide occurrence in each scenario using R statistical software. Finally, the accuracy and variable contribution analysis of all models was evaluated by the ROC curve and LVQ methods, respectively.

5.1. Logistic Regression (LR)

Logistic regression is one of the best-fitting models to explore the relationship between a dependent variable and a set of independent explanatory variables, which could be categorical, continuous, or binary variables [71,72]. The significant of the LR algorithm is that variables do not require normal distribution [17]. Independent variables in the logistic regression model must be denoted as 0 and 1; those represent the landslide absence and presence. Finally, the model output represents probability value within a range of 0 and 1. The logistic regression is based on the logistic algorithm P_i , denoted as [42]:

$$P = \frac{\exp(z)}{1 + \exp(z)} \quad (4)$$

where, P represents the probability value co-related to a certain observation, and z expressed as:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (5)$$

where, β_0 represents the intercept of the function, β also represents the independent variable's contribution X_1 , and n represents the number of influence factors [71].

5.2. LogitBoost (LB)

The logitBoost algorithm is generally used for best fitting the linear logistic regression functions at the tree node [73,74]. It is used to solve the over-fitting problems by the built of logitBoost algorithm [75]. The LogitBoost regression executes additive logistic regression with least-squares fits for individual class i.e., landslide occurrence and landslide non-occurrence as the following equation [76]:

$$Lc(x) = \sum_{i=1}^D \beta_i x_i + \beta_0 \quad (6)$$

where, D denotes the number of landslide-dependent factors and β_i is the coefficient of the i -th component within the input vector x . The posterior probabilities are developed by applying the linear logistic regression method, with the help of Equation (7) [73]:

$$p(C/x) = \exp(Lc(x)) / \sum_{C'=1}^C \exp(Lc'(x)) \quad (7)$$

where, C is the number of classes and the least-square fits $Lc(x)$ are resolved such $\sum_{C=1}^C L_C(x) = 0$ that establish the least number of instances per node of logistic model trees.

5.3. NaïveBayes (NB)

A NaïveBayes classifier is a machine-learning algorithm classifier system based on Bayes' law, which works under an independent assumption of variables [77]. The significant of the NB classifier is that it is very user-friendly for modeling without a requirement for any complicated iterative parameter estimation schemes [78]. Therefore, this model has been successfully implemented for

assessing landslide susceptibility mapping [9,26,38,79]. In this research, the NB model was calculated by applying a R 3.0.2 software environment using the “rminer” package [80].

Within an observation component of K attributes x_i , $i = 1, 2, \dots, k$ (x_i is landslide dependent factor), y_j , j is landslide, no landslide within the output class. NaïveBayes estimates the probability value $P(x_i/y_j)$ for each and every possible output class. The largest posterior probability prediction is calculated based on following equation [36,76]:

$$Y_{NB} = \operatorname{argmax} P(y_j) \prod_{i=1} P(x_i/y_j) \quad (8)$$

where, $y_j \in$ are landslide and non-landslide.

The conditional probability is calculated by applying flowing equation [31]:

$$P\left(\frac{x_i}{y_j}\right) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x_i - \mu)^2}{2\delta^2}} \quad (9)$$

where, μ and δ are mean and standard deviation of x_i respectively.

5.4. Analysis of Spatial Relationship between Landslide Location and Effective Factors Based on Frequency Ratio (FR)

The frequency ratio (FR) is the ratio of the area where landslide has occurred in respect to the total basin area, which is expressed as the ratio of the landslide probability in respect to the non-occurrences for a given attribute [2]. The FR values of each class of explanatory variables are representing the degree of landslide occurrence. Based on Equation (10), the FR value will be calculated for different landslide conditioning factors:

$$FR_i = \left(\frac{Npix(Sxi)}{\sum_{i=1}^m Sxi} \right) / \left(\frac{Npix(Xj)}{\sum_{j=1}^n Npix(Xj)} \right) \quad (10)$$

where, $Npix(Sxi)$ represents the number of pixels of landslide within class i for factor x , $Npix(Xj)$ represent the number of pixels within factor Xj , m is the number of classes in the parameter Xi and ‘ n ’ is the number of conditioning factors in the study area.

5.5. Analysis of Independent Variable's Importance

The models’ ability to develop landslide susceptibility predictions by the influences of the various landslide factors and therefore recognise their impact is a significant work. To find out the relationship between landslide events and conditioning factors importance analysis, learning vector quantization (LVQ) is implemented [81]. The LVQ algorithm is efficiently applied in different fields for landslide susceptibility mapping [82], groundwater potential mapping [83], mineral potential mapping [84], and environmental science [84]. Recently, Razavi Termeh et al. (2018) [81] and Pourghasemi and Rahmati (2018) [9] have utilized the LVQ algorithm to quantify variables importance for landslide susceptibility modeling. The Euclidean distance is assessed as the major rule of antagonism. The distance (Di) of neuron i is computed as the difference amongst the training vectors X and the reference vector Zi in following equation [85]:

$$Di = Zi - X = \sqrt{\sum Zij - Xj} \quad (11)$$

where, Xi and Zi are j th components of X and Zj , respectively. The importance of conditioning factors has been analyzed improving the learning equation of Zi . If the neuron stands in the false class [81,86]:

$$Zi(t+1) = Zi(t) - \lambda(t)hi(t)(X(t) - Zi(t)) \quad (12)$$

or the neuron stands in the correct class, then,

$$Z_i(t+1) = Z_i(t) + \lambda(t)h_i(t)(X(t) - Z_i(t)) \quad (13)$$

where, 1 represented if the i th neuron is a winner or 0 (otherwise),

$$\lambda(t) = \begin{cases} 0, \\ 1, \end{cases} \quad (14)$$

The neuron signifies the rate of exhilaration. $\lambda(t)$ represents the learning intensity at time t . This LVQ algorithm functioned in R statistical software. In the present study, the comparative performance of the conditioning factors (independent factor) to landslide events (i.e., dependent factor) are computed applying the LVQ method. The details of the LVQ method can be noted in the articles of Kohonen (1995) [87], and Kohonen et al. (1996) [86].

6. Results and Discussion

6.1. Multicollinearity Analysis

In this research, multicollinearity among the conditioning factors of landslide locations was identified with the help of VIF and tolerance (Table 3). The highest VIF values represent as lowest tolerance level. The highest and lowest VIF values are, respectively, related to slope gradient (4.182) and aspect (1.039). The results revealed that there is no collinearity among the 17 landslide conditioning factors in this study.

Table 3. Multicollinearity analysis for the landslide conditioning factors.

Factors	Collinearity Statistics	
	Tolerance	VIF
Aspect	0.962	1.039
Convexity	0.409	2.445
Altitude	0.280	3.572
Distance from fault	0.482	2.076
Flow accumulation	0.736	1.358
Forest density	0.420	2.382
Forest diameter	0.293	3.417
Forest type	0.539	1.857
Land use/land cover (LU/LC)	0.817	1.224
Lithology	0.479	2.089
Maximum curvature	0.369	2.706
Mid slope position	0.621	1.612
Profile curvature	0.479	2.089
Slope	0.239	4.182
Soil types	0.567	1.764
TPI	0.323	3.096
TWI	0.259	3.862

6.2. Spatial Relationship between Landslide Locations and Effective Factors

The spatial relationship between effective factors and landslide locations were calculated based on FR bivariate statistical technique (Table 4), the continuous value of the factor extracted from the DEM was reclassified in 5 classes using the quantile for ease of analysis.

Table 4. Spatial relationship between each effective factor and gully erosion locations using frequency ratio (FR) model.

Factor	Class or Type	Landslide	%Landslide	Domain	%Domain	FR
Aspect	F	34	12.41	261,192	10.87	1.14
	N	39	14.23	252,216	10.50	1.36
	NE	39	14.23	284,988	11.86	1.20
	E	19	6.93	264,632	11.02	0.63
	SE	13	4.74	272,460	11.34	0.42
	S	23	8.39	269,289	11.21	0.75
	SW	21	7.66	265,477	11.05	0.69
	W	48	17.52	260,769	10.86	1.61
Slope angle	1	5	1.82	478,026	19.90	0.09
	2	22	8.03	478,284	19.91	0.40
	3	40	14.60	492,094	20.48	0.71
	4	77	28.10	475,589	19.80	1.42
	5	130	47.45	478,282	19.91	2.38
Surface area	1	2	0.73	308,094	12.83	0.06
	2	46	16.79	897,345	37.35	0.45
	3	48	17.52	468,400	19.50	0.90
	4	86	31.39	396,462	16.50	1.90
	5	92	33.58	331,974	13.82	2.43
Maximum curvature	concave	51	18.61	726,645	30.25	0.62
	flat	89	32.48	889,637	37.03	0.88
	convex	134	48.91	785,993	32.72	1.49
Profile curvature	concave	88	32.12	736,127	30.64	1.05
	flat	44	16.06	751,387	31.28	0.51
	convex	142	51.82	914,761	38.08	1.36
TWI	1	123	44.89	457,389	19.04	2.36
	2	66	24.09	508,138	21.15	1.14
	3	52	18.98	503,008	20.94	0.91
	4	32	11.68	482,196	20.07	0.58
	5	1	0.36	451,544	18.80	0.02
TPI	1	34	12.41	459,382	19.12	0.65
	2	41	14.96	459,818	19.14	0.78
	3	28	10.22	482,381	20.08	0.51
	4	83	30.29	503,713	20.97	1.44
	5	88	32.12	496,981	20.69	1.55
Distance of Fault (m)	1	9	3.28	470,732	19.60	0.17
	2	83	30.29	476,084	19.82	1.53
	3	114	41.61	481,862	20.06	2.07
	4	39	14.23	485,262	20.20	0.70
	5	29	10.58	488,335	20.33	0.52
Convexity	1	2	0.73	472,798	19.68	0.04
	2	30	10.95	462,403	19.25	0.57
	3	56	20.44	473,777	19.72	1.04
	4	83	30.29	501,251	20.87	1.45
Forest type	PK	18	6.57	109,781	4.57	1.44
	D	204	74.45	1,121,961	46.70	1.59
	R	0	0.00	7279	0.30	0.00
	L	0	0.00	35,535	1.48	0.00
	PL	15	5.47	17,874	0.74	7.36
	99	2	0.73	684,994	28.51	0.03
	PH	0	0.00	5390	0.22	0.00
	PD	3	1.09	5038	0.21	5.22
	M	32	11.68	235,676	9.81	1.19
	H	0	0.00	178,747	7.44	0.00
Forest density	0	23	8.39	838,044	34.89	0.24
	C	245	89.42	1,488,114	61.95	1.44
	B	6	2.19	49,407	2.06	1.06
	A	0	0.00	26,710	1.11	0.00
Forest diameter	0	7	1.08	727,808	30.30	0.04
	1	63	9.71	110,236	4.59	2.12
	2	440	67.80	1,126,519	46.89	1.45
	3	139	21.42	437,712	18.22	1.18

Table 4. Cont.

Factor	Class or Type	Landslide	%Landslide	Domain	%Domain	FR
Land cover	100	0	0.00	155,472	6.47	0.00
	200	1	0.36	402,248	16.74	0.02
	300	229	83.58	1,626,301	67.70	1.23
	400	44	16.06	125,540	5.23	3.07
	500	0	0.00	5324	0.22	0.00
	600	0	0.00	42,429	1.77	0.00
	700	0	0.00	44,961	1.87	0.00
Geology	Biotite porphyry	0	0.00	21,258	0.88	0.00
	Jumunjin granite	107	39.05	1,319,764	54.94	0.71
	Alluvium	0	0.00	205,364	8.55	0.00
	Banded gneiss	0	0.00	131,399	5.47	0.00
	Biotite granite	167	60.95	695,831	28.97	2.10
	Noname	0	0.00	28,659	1.19	0.00
Soil	SmF2	163	59.49	615,013	25.60	2.32
	SgF2	31	11.31	221,518	9.22	1.23
	SgE2	58	21.17	575,574	23.96	0.88
	ScC	0	0.00	49,114	2.04	0.00
	MuD	1	0.36	14,197	0.59	0.62
	SiC	1	0.36	14,200	0.59	0.62
	MuC	6	2.19	34,360	1.43	1.53
	OsF	0	0.00	75,573	3.15	0.00
	RC	1	0.36	49,918	2.08	0.18
	SmF3	3	1.09	6315	0.26	4.17
	SiB	0	0.00	1023	0.04	0.00
	SgD2	1	0.36	129,163	5.38	0.07
	OdF	0	0.00	91	0.00	0.00
	W	0	0.00	16,686	0.69	0.00
	BRS	0	0.00	9403	0.39	0.00
	VcB	0	0.00	26,155	1.09	0.00
	YaD2	1	0.36	54,643	2.27	0.16
	YaE2	0	0.00	22,274	0.93	0.00
	NkB	0	0.00	9570	0.40	0.00
	ScB	0	0.00	16,263	0.68	0.00
	Ki	0	0.00	35,943	1.50	0.00
	YeB	2	0.73	78,790	3.28	0.22
	YeC	1	0.36	37,489	1.56	0.23
	SAC	0	0.00	54,763	2.28	0.00
	SAB	0	0.00	23,182	0.97	0.00
	BG	0	0.00	24,473	1.02	0.00
	Jd	0	0.00	13,064	0.54	0.00
	YdB	0	0.00	2227	0.09	0.00
	Yf	0	0.00	8782	0.37	0.00
	Ym	0	0.00	6584	0.27	0.00
	Hh	0	0.00	14,957	0.62	0.00
	Kw	0	0.00	6810	0.28	0.00
	YaC2	0	0.00	2310	0.10	0.00
	HuB	0	0.00	42,744	1.78	0.00
	SgE3	5	1.82	64,836	2.70	0.68
	JoB	0	0.00	1068	0.04	0.00
	Ng	0	0.00	3245	0.14	0.00
	JiB	0	0.00	8284	0.34	0.00
	BqB	0	0.00	8370	0.35	0.00
	Dq	0	0.00	4438	0.18	0.00
	Gq	0	0.00	5517	0.23	0.00
	HT	0	0.00	9881	0.41	0.00
	SoD2	0	0.00	1299	0.05	0.00
	Gt	0	0.00	1458	0.06	0.00
	HI	0	0.00	708	0.03	0.00

Table 4. Cont.

Factor	Class or Type	Landslide	%Landslide	Domain	%Domain	FR
Flow accumulation	1	53	19.34	483,922	20.14	0.96
	2	92	33.58	549,109	22.86	1.47
	3	63	22.99	517,391	21.54	1.07
	4	38	13.87	432,821	18.02	0.77
	5	28	10.22	419,032	17.44	0.59
Mid slope position	1	81	29.56	474,165	19.74	1.50
	2	69	25.18	476,110	19.82	1.27
	3	57	20.80	465,025	19.36	1.07
	4	27	9.85	498,482	20.75	0.47
	5	40	14.60	488,493	20.33	0.72

Aspect, slope, altitude, maximum curvature, profile curvature, topographic wetness index (TWI), topographic positioning index (TPI), distance from fault, convexity, forest type, forest diameter, forest density, land use and land cover, lithology, soil, flow accumulation, and mid slope position.

The highest FR value (1.61) was found in western facing aspect class, followed by northern facing aspect with a FR value of 1.36. Slope angle is highly positively correlated with the spatial occurrence of landslides as steep slope (5th class) areas represents most of the identified landslides (FR = 2.38). The highest FR value (2.43) was found in the 5th elevation class, followed by the 4th class with a FR value of 1.90. The FR value was maximum of convex curvature for both cases i.e., maximum curvature as well as profile of curvature. The FR value was 1.49 and 1.36 for maximum and profile of curvature respectively, due to it having great influence on the susceptibility to landslide occurrence. The highest influence of convex curvature on landslide occurrence is found on areas where additional influences of convergences and divergence of running water during downhill flow are found [26]. Naturally, the TWI has an intimate spatial correlation with landslide but it varies according to the varying surface topography regions. In the present scenario, only the 1st class of TWI has a greater FR value (FR = 2.40) followed by 2nd (FR = 1.14) and 3rd classes (FR = 0.91), respectively. The TPI represents the low land, high land, and constant slope areas. The FR value is maximum (1.55) for the 5th TPI classes that can represent high landslide occurrence rates. The distance from the fault is an important influencing factor for landslide occurrence because being near to the fault line represents the more weak zones due to it being characterized by heavily fractured rocks [88]. There is a trend of negative relationship between distance from fault and landslide occurrence. The highest FR value (2.07) was found in the 3rd class of distance from the fault line factor. Forest factors also play an important role in landslide occurrence, such as the PL and PD types of forest having greater FR values of 7.36 and 5.22, respectively. In respect of forest diameter, the 2nd and 3rd classes of FR value are 2.12 and 1.45, and in respect to forest density class C has greater influence of landslide occurrence (FR = 1.44). In general, LULC types are very strong factors for landslide occurrence. We found especially high FR values in 400 (FR = 3.07) and 300 (FR = 1.23) classes. In addition, the highest FR value (2.10) represented in the Jbgr lithology class was followed by the Jjgr lithology class by FR value of 0.71. The FR value of 4.17 for a RC soil can be interpreted as a high landslide conditioning factor, while SmF2, and Muc FR values are 2.32 and 1.53, respectively in the area. Flow accumulation is one of the most important parameters that control the stability of a slope and the degree of saturation of the material on the slope [89]. There is a trend of inverse relationship between the flow accumulation and landslide occurrence. The highest FR value in this 4th factor class is 1.47.

6.3. Variable Contribution Analysis

The results of the LVQ technique is represented in Figure 4 and, according to that, slope (VI = 72.9%), soil (VI = 72.1%), TWI (VI = 72.0%), forest types (VI = 66.7%), and convexity (VI = 66.6%), are the most important landslide conditioning factors, followed by maximum curvature (VI = 63.9%), forest density (VI = 62.6%), forest diameter (VI = 62.4%), TPI (VI = 61.1%), profile of curvature (VI = 60.1%), elevation (VI = 59.7), lithology (VI = 59.1), flow accumulation (VI = 58.2%),

MSP (VI = 58.0%), LULC (VI = 57.3%), distance from fault (VI = 52.5%), and aspect (51.3%). Hence, all these conditioning factors were selected for landslide susceptibility mapping, due to them having a significant role in landslide occurrence in the study area.

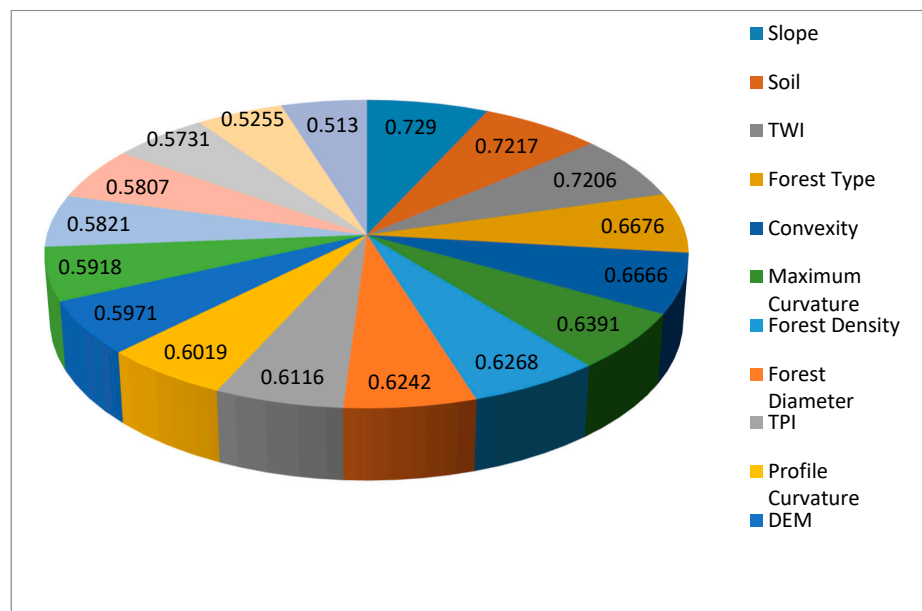


Figure 4. Variable contribution of factors according to Learning Vector Quantization (LVQ) method.

6.4. Landslide Susceptibility Models

Landslide predictions by the modeling and simulation are an important prospect for natural resources studies [9]. Therefore, the prepared landslide susceptible map can be used as a valuable tool for conservation and sustainable planning of landslide-prone areas. For this reason, landslide susceptibility assessment is required to be as accurate as possible. For the purpose of comparative visualization, the landslide susceptibility maps generated by applying three machine-learning algorithms (LR, LB and NB) are represented in Figures 5–7. The landslide susceptibility maps (LR, LB and NB models) represent the low to very high susceptibility of a landslide occurrence. A higher index represents a higher susceptibility of the area to landslides. Classification of the landslide susceptibility index into different classes is necessary for the description and easier interpretation of spatial landslide estimation [36]. For this reason, these landslide susceptibility values are divided into five classes based on the natural break [17,90] range representing five different zones in landslide susceptibility maps; very low, low, medium, high, and very high zones. It is clear in all landslide susceptibility maps that the maximum landslide susceptible areas are found middle part of this basin. Concurrent with this final model output, it was found that logistic regression and NaïveBayes generate LSMs that are spatially discontinuous, while the logitBoost (LB) model produce smoother patterns.

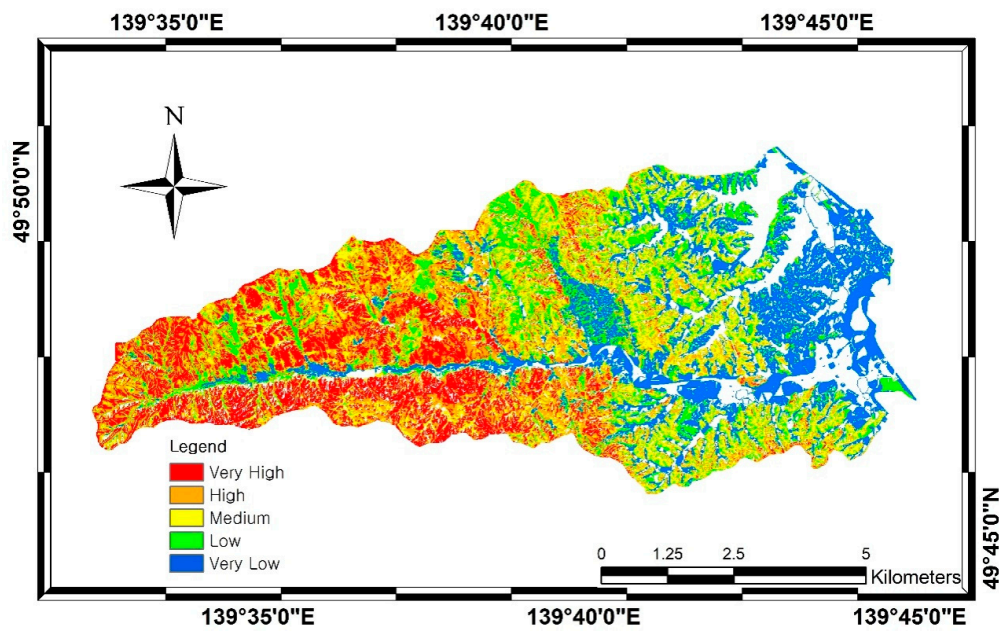


Figure 5. Landslide susceptibility map produced by Logistic regression.

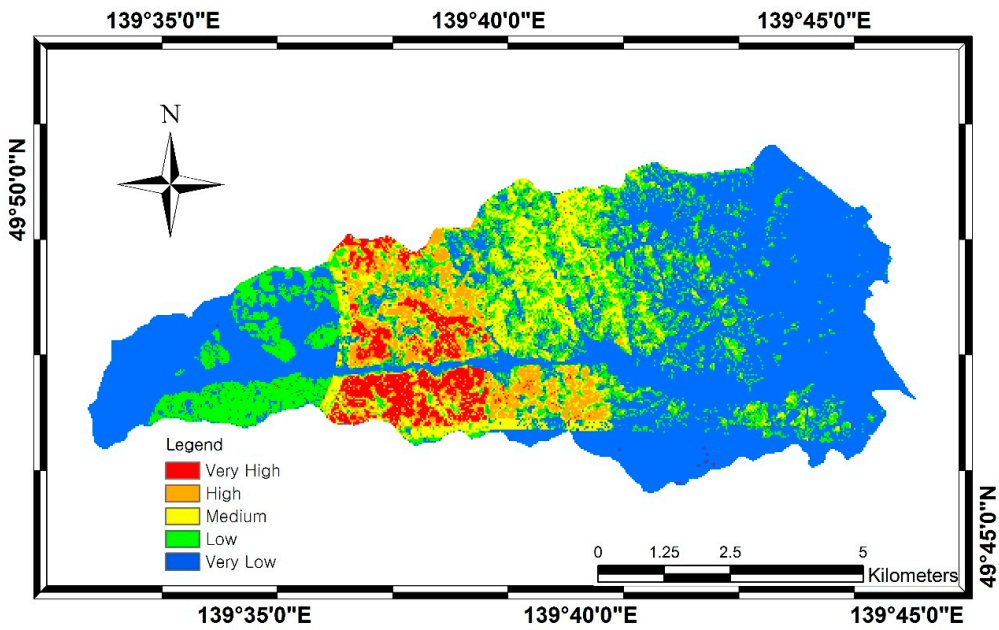


Figure 6. Landslide susceptibility map produced by LogitBoost.

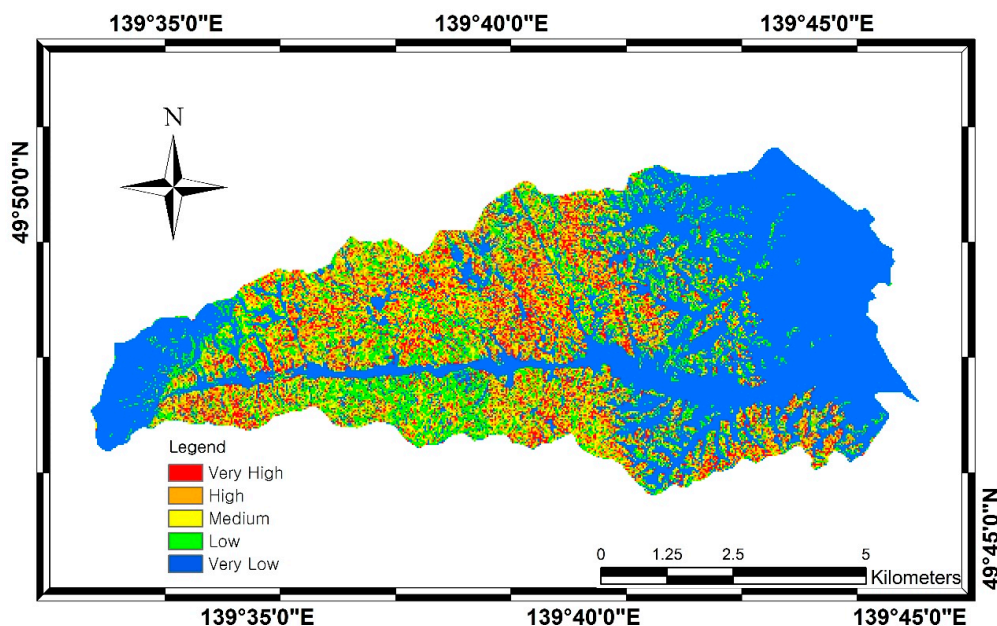


Figure 7. Landslide susceptibility map produced by NaïveBayes.

6.5. Accuracy Assessment and Their Comparison

The ROC curve was generated to compare the machine-learning models for validation dataset [86]. In the case of the prediction rate curve (using validation data set), the LogitBoost displayed the highest AUC value of 85.2% followed by logistic regression (AUC = 84.2%, and NaïveBayes (AUC = 70.7%) models. The result of prediction curve reveals that the LB model is relatively better than the LR and NB models. The models LB and LR were reasonably satisfactory for the prediction of the landslide susceptibility map in the study area.

Although, different machine-learning algorithms for landslide susceptibility assessment have been implemented, the prediction accuracy of these models is also debated [9,26]. Previous research works revealed that some of the machine-learning techniques had equal model accuracy, although they are unique in respect to individual approaches for landslide susceptibility assessment and identifying the relevant spatial relationship between the landslide factors and landslide initiation [38]. Exploring and gathering knowledge of these differences is necessary to implement a suitable model for a significant study goal for a proposed study area [91,92]. The use of machine-learning algorithms is relatively easier than statistical model for identifying interactions between dependent and independent variables [24]. The analyzed result shows a reasonable performance regarding predicting landslide susceptibility mappings as receiver operating characteristic (ROC) curve in Figure 8.

The result reveals that the LB model was the most accurate model in comparison to the LR and NB models in this study area. The better results of the LB model due to it solving the data over fitting problems [76] and also this model uses maximum likelihood to find the smallest probable difference between predicted and observed values [93]. Tsangaratos and Ilia (2018) [79] compared a logistic regression and NaïveBayes model in landslide susceptibility assessment. The result reveals that the NB model had great predication accuracy [79]. But in this study area the NB model did not predict a better result in respect to LR and LB models because the NB classifier converges more accelerated than LB and LR models, meaning that it can run with the help of less training data set and still have lesser predictive power in this study area. However, if one assesses the performance of the two models i.e., NB and LR, the LR approaches provides a more appropriate result. When applying NaïveBayes to models with the help of small number of variables the increase of the training data size but still it's not affected the training accuracy, but this pattern is not shown in the cases of LR models. LR had provided a positive relationship with training data a predictive accuracy. It can be concluded that the

LB model had a great performance because it provides a better discrimination between presence and absences of landslide occurrence. With respect to landslide susceptibility assessment, the NB model does not perform better than the LR and LB models in this study area.

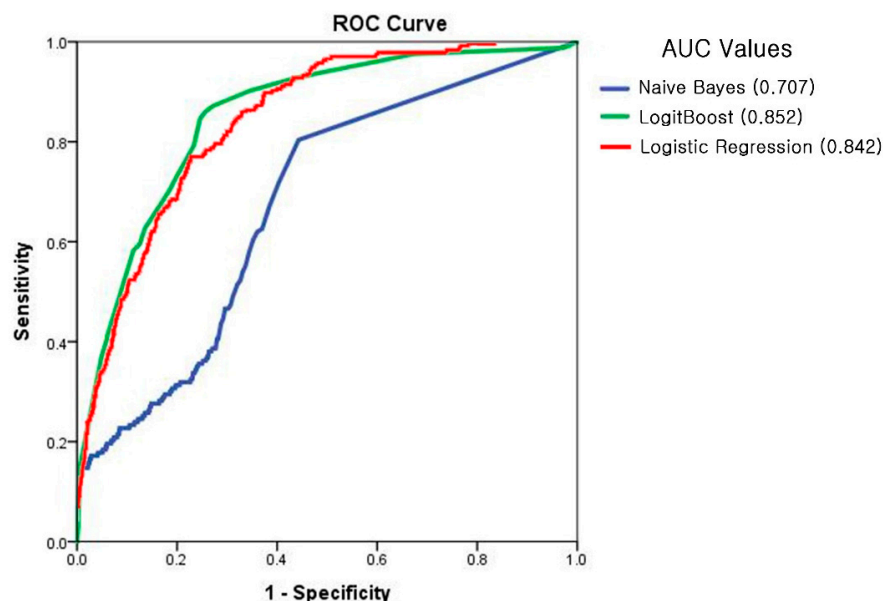


Figure 8. Receiver operating characteristic (ROC) curve of the landslide susceptibility maps using validation data set.

7. Conclusions

Mitigation landslide hazard and management planning plays a crucial role in sustaining biotic society and the economic world. In this research, three machine-learning models (LR, LB and NB) were used to identify the possible landslide-prone areas at a regional scale. The landslide susceptibility maps can be used to (i) most accurately identify areas that in the present time are more vulnerable to landslide occurrence; (ii) identify the most important factors for landslide occurrence through the LVQ method which can be useful for tools developing land use management strategies; (iii) finally, LSMs provide more sustainable landslide planning with an emphasis on land degradation management.

In order to validation process to prove the prediction ability of three proposed models, the prediction rate curve of the ROC was used. The AUC was calculated based on the validation data set. The AUC results showed that the prediction rates are 85.2%, 84.2% and 70.7%, respectively, for LB, LR and NB models. In addition, the LB was estimated to have prediction ability in respect to other machine-learning models in this study area. The logitBoost algorithm is generally used for best fitting the linear logistic regression functions at the tree node and solved the overfitting problems by the use of logitBoost algorithm. Finally, these landslide susceptibility maps can be supportive for planners and policy makers in hazard and disaster management planning.

Author Contributions: Writing-original draft, H.R.P. and A.G.; Methodology, S.P.; Supervision, C.-W.L.; Conceptualization, S.L.

Funding: This research was part of a 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. 2016K1A3A1A09915721 and 2017R1A2B4003258).

Acknowledgments: The third 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

1. Lee, S.; Choi, J.; Min, K. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *Int. J. Remote Sens.* **2004**, *25*, 2037–2052. [[CrossRef](#)]
2. Lee, S.; Pradhan, B. Probabilistic Landslide Hazards and Risk Mapping on Penang Island, Malaysia. *J. Earth Syst. Sci.* **2006**, *115*, 661–672. [[CrossRef](#)]
3. Varnes, D.J. Slope movement types and processes. In *Landslides Analysis and Control*; Schuster, R.L., Krizek, R.J., Eds.; Transportation Research Board National Academy of Sciences: Washington, DC, USA, 1978; pp. 11–33. ISBN 0-309-02804-3.
4. Smith, K. *Environmental Hazards: Assessing Risk and Reducing Disaster*, 6th ed.; Routledge: Abingdon, UK, 2013; pp. 205–217. ISBN 978-0415681063.
5. Wang, L.; Sawada, K.; Moriguchi, S. Landslide Susceptibility Mapping by Using Logistic Regression Model with Neighborhood Analysis: A Case Study in Mizunami City. *Int. J. GEOMATE* **2011**, *1*, 99–104. [[CrossRef](#)]
6. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C.; Moezzi, K.D. Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz watershed, Iran. In *Terrigenous Mass Movements*; Pradhan, B., Buchroithner, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 23–49. ISBN 978-3-642-25495-6.
7. Basu, T.; Pal, S. Identification of landslide susceptibility zones in Gish River basin, West Bengal, India. *Georisk Assess. Manag. Risk Eng. Syst. Geohazards* **2018**, *12*, 14–28. [[CrossRef](#)]
8. Betts, H.; Basher, L.; Dymond, J.; Herzig, A.; Marden, M.; Phillips, C. Development of a landslide component for a sediment budget model. *Environ. Model. Softw.* **2017**, *92*, 28–39. [[CrossRef](#)]
9. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192. [[CrossRef](#)]
10. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
11. Zare, M.; Pourghasemi, H.R.; Vafakhah, M.; Pradhan, B. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arab. J. Geosci.* **2013**, *6*, 2873–2888. [[CrossRef](#)]
12. Lee, E.M.; Jones, D.K.C. *Landslide Risk Assessment*; Thomas Telford: London, UK, 2004; p. 454. ISBN 0-7277-3171-8.
13. Pourghasemi, H.R.; Moradi, H.R.; Aghda, S.F.; Gokceoglu, C.; Pradhan, B. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arab. J. Geosci.* **2014**, *7*, 1857–1878. [[CrossRef](#)]
14. Hasekiogullari, G.D.; Ercanoglu, M.A. A new approach to use AHP in landslide susceptibility mapping: A case study at Yenice (Karabuk, NW Turkey). *Nat. Hazards* **2012**, *63*, 1157–1179. [[CrossRef](#)]
15. Bijukchhen, S.M.; Kayastha, P.; Dhital, M.R. A Comparative Evaluation of Heuristic and Bivariate Statistical Modeling for Landslide Susceptibility Mappings in Ghurmi-Dhad Khola, East Nepal. *Arab. J. Geosci.* **2013**, *6*, 2727–2743. [[CrossRef](#)]
16. Kayastha, P.; Dhital, M.R.; Smedt, F.D. Evaluation and Comparison of GIS Based Landslide Susceptibility Mapping Procedures in Kulekhani Watershed, Nepal. *J. Geol. Soc. India* **2013**, *81*, 219–231. [[CrossRef](#)]
17. 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](#)]
18. Akgun, A. A Comparison of Landslide Susceptibility Maps Produced by Logistic Regression, Multi-criteria Decision, and Likelihood Ratio Methods: A Case Study at İzmir, Turkey. *Landslides* **2012**, *9*, 93–106. [[CrossRef](#)]
19. Eker, R.; Aydın, A. Assessment of Forest Road Conditions in Terms of Landslide Susceptibility: A Case Study in Yığılca Forest Directorate (Turkey). *Turk. J. Agric. For.* **2014**, *38*, 281–290. [[CrossRef](#)]
20. Nandi, A.; Shakoor, A.A. GIS-based Landslide Susceptibility Evaluation Using Bivariate and Multivariate Statistical Analyses. *Eng. Geol.* **2010**, *110*, 11–20. [[CrossRef](#)]
21. Pradhan, B.; Mansor, S.; Pirasteh, S.; Buchroithner, M.F. Landslide Hazard and Risk Analyses at a Landslide Prone Catchment Area Using Statistical Based Geospatial Model. *Int. J. Remote Sens.* **2011**, *32*, 4075–4087. [[CrossRef](#)]
22. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [[CrossRef](#)]

23. 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](#)]
24. Yilmaz, I. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability (CP) and artificial neural network (ANN). *Environ. Earth Sci.* **2010**, *60*, 505–519. [[CrossRef](#)]
25. Arnone, E.; Francipane, A.; Scarbaci, A.; Puglisi, C.; Noto, L.V. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environ. Model. Softw.* **2016**, *84*, 467–481. [[CrossRef](#)]
26. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes models. *Math. Probl. Eng.* **2012**, *2012*, 974638. [[CrossRef](#)]
27. 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](#)]
28. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* **2012**, *239*, 27–39. [[CrossRef](#)]
29. 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](#)]
30. Oh, H.J.; Pradhan, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput. Geosci.* **2011**, *37*, 1264–1276. [[CrossRef](#)]
31. Sezer, E.A.; Pradhan, B.; Gokceoglu, C. Manifestation of an adaptive neurofuzzy model on landslide susceptibility mapping: Klang valley, Malaysia. *Expert Syst. Appl.* **2011**, *38*, 8208–8219. [[CrossRef](#)]
32. Ghorbanzadeh, O.; Rostamzadeh, H.; Blaschke, T.; Gholaminia, K.; Aryal, J. A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping. *Nat. Hazards* **2018**, *93*, 1–21. [[CrossRef](#)]
33. Ghorbanzadeh, O.; Blaschke, T.; Aryal, J.; Gholaminia, K. A new GIS-based technique using an adaptive neuro-fuzzy inference system for land subsidence susceptibility mapping. *J. Spat. Sci.* **2018**. [[CrossRef](#)]
34. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
35. Dickson, M.E.; Perry, G.L.W. Identifying the controls on coastal cliff landslides using machine-learning approaches. *Environ. Model. Softw.* **2016**, *76*, 117–127. [[CrossRef](#)]
36. Trigila, A.; Iadanza, C.; Esposito, C.; Mugnozza, G.S. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampiliera (NE Sicily, Italy). *Geomorphology* **2015**, *249*, 119–136. [[CrossRef](#)]
37. Pham, B.T.; Pradhan, B.; Bui, D.T.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
38. Roodposhti, M.J.; Shahabi, H.; Safarrad, T. Fuzzy Shannon entropy: A hybrid GIS-based landslide susceptibility mapping method. *Entropy* **2016**, *18*, 343. [[CrossRef](#)]
39. Pourghasemi, H.R.; Moradi, H.R.; Aghda, S.F. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* **2013**, *69*, 749–779. [[CrossRef](#)]
40. Pourghasemi, H.; Pradhan, B.; Gokceoglu, C.; Moezzi, K.D. A comparative assessment of prediction capabilities of Dempster-Shafer and weights-of-evidence models in landslide susceptibility mapping using GIS. *Geomat. Nat. Hazards Risk* **2013**, *4*, 93–118. [[CrossRef](#)]
41. Lee, S.; Seong, W.J.; Oh, K.Y.; Lee, M.J. The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models; a case study of Inje, Korea. *Open Geosci.* **2016**, *8*, 117–132. [[CrossRef](#)]
42. Lee, S.; Hong, S.M.; Jung, H.S. A support vector machine for landslide susceptibility mapping in Gangwon province, Korea. *Sustainability* **2017**, *9*, 48. [[CrossRef](#)]
43. Kalantar, B.; Pradhan, B.; Naghibi, A.S.; Motevalli, A.; Mansor, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomat. Nat. Hazards Risk* **2018**, *9*, 49–69. [[CrossRef](#)]

44. Dai, F.C.; Lee, C.F.; Li, J.; Xu, Z.W. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environ. Geol.* **2001**, *40*, 381–391. [[CrossRef](#)]
45. Ercanoglu, M.; Gokceoglu, C. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Eng. Geol.* **2004**, *75*, 229–250. [[CrossRef](#)]
46. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [[CrossRef](#)]
47. Bednarik, R.G.; Khan, M. The rock art of southern Arabia reconsidered. *Adumatu J.* **2009**, *20*, 7–20.
48. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [[CrossRef](#)]
49. Ayalew, L.; Yamagishi, H.; Ugawa, N. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides* **2004**, *1*, 73–81. [[CrossRef](#)]
50. Gritzner, M.L.; Marcus, W.A.; Aspinall, R.; Custer, S.G. Assessing landslide potential using GIS, soil wetness modelling and topographic attributes, Payette River, Idaho. *Geomorphology* **2001**, *37*, 149–165. [[CrossRef](#)]
51. Haigh, M.; Rawat, J.S. Landslide Disasters: Seeking Causes—A Case Study from Uttarakhand, India. In *Management of Mountain Watersheds*; Springer: Dordrecht, The Netherlands, 2012; pp. 218–253. [[CrossRef](#)]
52. Talebi, A.; Uijlenhoet, R.; Troch, P.A. Soil moisture storage and hillslope stability. *Nat. Hazards Earth Syst. Sci.* **2007**, *7*, 523–534. [[CrossRef](#)]
53. Gomez, H.; Kavzoglu, T. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa river basin, Venezuela. *Eng. Geol.* **2005**, *78*, 11–27. [[CrossRef](#)]
54. Yilmaz, I. A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. *Bull. Eng. Geol. Environ.* **2009**, *68*, 297–306. [[CrossRef](#)]
55. Lee, S.; Chwae, U.; Min, K. Landslide susceptibility mapping by correlation between topography and geological structure: The Janghung area, Korea. *Geomorphology* **2002**, *46*, 149–162. [[CrossRef](#)]
56. Bucci, F.; Santangelo, M.; Cardinali, M.; Fiorucci, F.; Guzzetti, F. Landslide distribution and size in response to Quaternary fault activity: The Peloritani Range, NE Sicily, Italy. *Earth Surf. Process. Landf.* **2016**, *41*, 711–720. [[CrossRef](#)]
57. Liu, C.; Li, W.; Wu, H.; Lu, P.; Sang, K.; Sun, W.; Chen, W.; Hong, Y.; Li, R. Susceptibility evaluation and mapping of China's landslides based on multi-source data. *Nat. Hazards* **2013**, *69*, 1477–1495. [[CrossRef](#)]
58. Restrepo, C.; Vitousek, P.; Neville, P. Landslides significantly alter land cover and the distribution of biomass: An example from the Ninole ridges of Hawaii. *Plant Ecol.* **2003**, *166*, 131–143. [[CrossRef](#)]
59. Rahmati, O.; Haghighizadeh, A.; Pourghasemi, H.R.; Noormohamadi, F. Gully erosion susceptibility mapping: The role of GIS-based bivariate statistical models and their comparison. *Nat. Hazards* **2016**, *82*, 1231–1258. [[CrossRef](#)]
60. 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](#)]
61. Henriques, C.; Zêzere, J.L.; Marques, F. The role of the lithological setting on the landslide pattern and distribution. *Eng. Geol.* **2015**, *189*, 17–31. [[CrossRef](#)]
62. Sharma, L.P.; Patel, N.; Debnath, P.; Ghose, M.K. Assessing Landslide Vulnerability from Soil Characteristics—A GIS-based Analysis. *Arab. J. Geosci.* **2012**, *5*, 789–796. [[CrossRef](#)]
63. Kazakis, N.; Kougiass, I.; Patsilis, T. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Sci. Total Environ.* **2015**, *538*, 555–563. [[CrossRef](#)] [[PubMed](#)]
64. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 311–324. [[CrossRef](#)]
65. Böhner, J.; AntoniĆ, O. Land-surface parameters specific to topo-climatology. In *Geomorphometry: Concepts, Software, Applications. Development Sin Soil Science*; Hengl, T., Reuter, H.I., Eds.; Elsevier Science & Technology: Oxford, UK, 2008; Volume 33, pp. 195–226. ISBN 978-0-12-374345-9.
66. Häring, T.; Dietz, E.; Osenstetter, S.; Koschitzki, T.; Schröder, B. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* **2012**, *185*, 37–47. [[CrossRef](#)]

67. Saha, S. Groundwater potential mapping using analytical hierarchical process: A study on Md. Bazar Block of Birbhum District, West Bengal. *Spat. Inf. Res.* **2017**, *25*, 615–626. [[CrossRef](#)]
68. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [[CrossRef](#)] [[PubMed](#)]
69. Hosmer, D.W.; Lemeshow, S. *Applied Regression Analysis*; John Wiley and Sons: New York, NY, USA, 1989; ISBN 978-0-470-58247-3.
70. Menard, S. *Applied Logistic Regression Analysis*, 2nd ed.; Sage Publication: Thousand Oaks, CA, USA, 2001; pp. 1–101. ISBN 0-7619-2208-3.
71. Atkinson, P.M.; Massari, R. Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Comput. Geosci.* **1998**, *24*, 373–385. [[CrossRef](#)]
72. Gayen, A.; Saha, S. Deforestation probable area predicted by logistic regression in Pathro river basin: A tributary of Ajay River. *Spat. Inf. Res.* **2018**, *26*, 1–9. [[CrossRef](#)]
73. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Mach. Learn.* **2005**, *59*, 161–205. [[CrossRef](#)]
74. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**, *13*, 361–378. [[CrossRef](#)]
75. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall: New York, NY, USA, 1984; p. 337. ISBN 978-0-412-04841-8.
76. Doetsch, P.; Buck, C.; Golik, P.; Hoppe, N.; Kramp, M.; Laudenberg, J.; Oberdörfer, C.; Steingrube, P.; Forster, J.; Mauser, A. Logistic model trees with AUC split criterion for the KDD Cup 2009 Small Challenge. In Proceedings of the 2009 International Conference on KDD-Cup 2009, Paris, France, 28 June–1 July 2009; pp. 77–88.
77. Soria, D.; Garibaldi, J.M.; Ambrogi, F.; Biganzoli, E.M.; Ellis, I.O. A “non-parametric” version of the naive Bayes classifier. *Knowl. Based Syst.* **2011**, *24*, 775–784. [[CrossRef](#)]
78. Wu, X.; Kumar, V.; Ross, Q.J. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
79. Tsangaratos, P.; Ilia, I. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* **2016**, *145*, 164–179. [[CrossRef](#)]
80. Cortez, P. *Package ‘Rminer’*; Teaching Report; Department of Information System, ALGORITMI Research Centre, Engineering School: Guimares, Portugal, 2016; p. 59.
81. Termeh, S.V.; Kornejady, A.; Pourghasemi, H.R.; Keesstra, S. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci. Total Environ.* **2018**, *615*, 438–451. [[CrossRef](#)] [[PubMed](#)]
82. Pavel, M.E.; Hainsworth, J.D.; Baudin, E.; Peeters, M.; Hörsch, D.; Winkler, R.E.; Klimovsky, J.; Lebwohl, D.; Jehl, V.; Wolin, E.M.; et al. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): A randomised, placebo-controlled, phase 3 study. *Lancet* **2011**, *378*, 2005–2012. [[CrossRef](#)]
83. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and randomforest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [[CrossRef](#)] [[PubMed](#)]
84. Tayebi, M.H.; Tangestani, M.H. Sub pixel mapping of alteration minerals using SOM neural network model and Hyperion data. *Earth Sci. Inform.* **2015**, *8*, 279–291. [[CrossRef](#)]
85. Williams, R.N.; de Souza, P.A., Jr.; Jones, E.M. Analysing coastal ocean model outputs using competitive-learning pattern recognition techniques. *Environ. Model. Softw.* **2014**, *57*, 165–176. [[CrossRef](#)]
86. Kohonen, T.; Hynninen, J.; Kangas, J.; Laaksonen, J.; Torkkola, K. *LVQPAK: The Learning Vector Quantization Program Package*; Technical Report; Laboratory of Computer and Information Science Rakentajanaukio 2 C: Espoo, Finland, 1996; pp. 1991–1992. ISBN 951-22-2948-X.
87. Kohonen, T. Learning Vector Quantization. In *Self-Organizing Maps Springer Series in Information Sciences*; Huang, T.S., Kohonen, T., Schroeder, M.R., Eds.; Springer: Berlin, Germany, 1995; pp. 175–189. ISBN 978-3-642-97610-0.

88. Foumelis, M.; Lekkas, E.; Parcharidis, I. Landslide susceptibility mapping by GIS-based qualitative weighting procedure in Corinth area. In Proceedings of the 10th International Congress, Thessaloniki, Greece, 14–20 April 2004; pp. 904–912.
89. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena* **2008**, *72*, 1–12. [[CrossRef](#)]
90. Pourghasemi, H.R.; Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ. Earth Sci.* **2016**, *75*, 1–17. [[CrossRef](#)]
91. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modelling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
92. Brenning, A. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In *SAGA—Seconds Out*; Böhner, J., Blaschke, T., Montanarella, L., Eds.; Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie 19; Hamburg, Germany, 2008; pp. 23–32.
93. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).