

Article

# Landslide Susceptibility Mapping Using GIS-Based Data Mining Algorithms

Vali Vakhshoori <sup>1,\*</sup>, Hamid Reza Pourghasemi <sup>2</sup>, Mohammad Zare <sup>3</sup> and Thomas Blaschke <sup>4</sup>

<sup>1</sup> Department of Geography, Social Science Centre, Western University, London, ON N6A 5C2, Canada

<sup>2</sup> Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz 71441-65186, Iran; hr.pourghasemi@shirazu.ac.ir

<sup>3</sup> Department of Earth Sciences, College of Sciences, Shiraz University, Shiraz 71467-13565, Iran; zare@susc.ac.ir

<sup>4</sup> Department of Geoinformatics—Z\_GIS, University of Salzburg, 5020 Salzburg, Austria; thomas.blaschke@sbg.ac.at

\* Correspondence: svakhsho@uwo.ca; Tel.: +1-226-998-0101

Received: 23 August 2019; Accepted: 24 October 2019; Published: 1 November 2019



**Abstract:** The aim of this study was to apply data mining algorithms to produce a landslide susceptibility map of the national-scale catchment called Bandar Torkaman in northern Iran. As it was impossible to directly use the advanced data mining methods due to the volume of data at this scale, an intermediate approach, called normalized frequency-ratio unique condition units (NFUC), was devised to reduce the data volume. With the aid of this technique, different data mining algorithms such as fuzzy gamma (FG), binary logistic regression (BLR), backpropagation artificial neural network (BPANN), support vector machine (SVM), and C5 decision tree (C5DT) were employed. The success and prediction rates of the models, which were calculated by receiver operating characteristic curve, were 0.859 and 0.842 for FG, 0.887 and 0.855 for BLR, 0.893 and 0.856 for C5DT, 0.891 and 0.875 for SVM, and 0.896 and 0.872 for BPANN that showed the highest validation rates as compared with the other methods. The proposed approach of NFUC proved highly efficient in data volume reduction, and therefore the application of computationally demanding algorithms for large areas with voluminous data was feasible.

**Keywords:** landslide susceptibility; data mining methods; small scale; normalized frequency ratio; unique condition unites; spatial modeling; geographic information system

## 1. Introduction

Landslides frequently cause damage to properties and loss of lives in susceptible areas all over the world. Since the beginning of the 21st century, around 1.5 million people have been affected by landslides; moreover, landslides have caused financial losses above 875 million US dollars [1]. For instance, in the course of 12 years from 2004 to 2016, more than 55,000 lives worldwide have been lost due to landslides [2]. Therefore, it is obvious that the identification of landslide-prone zones and prevention of possible damages and fatalities is of a crucial importance. Landslide studies for a sensitive zone can be done in three consecutive levels such as landslide susceptibility, hazard, and risk mapping [3,4]. The first fundamental step in this regard is to produce landslide susceptibility maps (LSMs) of the sensitive areas [3,5–8]. LSMs show where future landslides may occur in a study area [3,7–13] and are created according to three fundamental assumptions summarized as follows:

1. The signs of landslides that have occurred can be recognized through filed investigation and remote sensing techniques [14–17];

2. Causative factors which affect landslides can be collected and analyzed to assess the probability of occurrence of the future landslides [3,4,18–20];
3. Past and present landslides are the keys for predicting future landslides [3,4,9,20–23].

On the basis of these assumptions, the relationship between the occurred landslides and the causative factors for creating an LSM can be analyzed through qualitative or quantitative methods [3,4,24]. Qualitative analyses of landslide susceptibility assess the effects of causative factors on landslides based on experts' opinions [3,4], and therefore the existence of a landslide inventory is not necessary, however, the application of quantitative methods necessarily depends on the presence of a landslide inventory map [3,4,23]. Using both types of methods, LSMs are produced at different scales [4] such as site-specific zones (from several hectares to dozens of square kilometers), local scale (10–1000 km<sup>2</sup>), regional scale (1000–10,000 km<sup>2</sup>), and national scale (>10,000 km<sup>2</sup>) and a direct connection can be found between the scale of mapping and the sophistication of the employed methods, that is to say, more sophisticated methods are used at larger scales (small study areas) [3]. For example, sophisticated deterministic and mathematical models are often used at site-specific zones to investigate the behavior of individual landslides [25]. At local and regional scales, numerous landslide susceptibility studies have used various types of qualitative and quantitative methods such as analytical hierarchy process (AHP) [26–29], weights-of-evidence (WofE) [30–33], frequency ratio (FR) [34–39], fuzzy logic [40–42], logistic regression (LR) [43–49], artificial neural networks (ANN) [50–55], and support vector machine (SVM) [56–59]. Often, at a national scale, LSMs have been created by employing only simple methods, although national-scale LSMs should be produced with the highest possible accuracy and reliability because these maps are used in preliminary assessments when more detailed landslide susceptibility, hazard, or risk studies are required and are even used directly in land use management and environmental impact assessment [3,4,6,11].

Most of the time, simple qualitative methods are applied to produce national-scale LSMs [60], but it should be noted that qualitative methods are, in general, subjective and not as reliable as quantitative methods for producing LSMs [61–63]. The lack of a landslide inventory map, which needs a considerable budget to be prepared for a large area, is the main reason why qualitative methods are often used at a national scale [60,64–67], however, nowadays advanced data and techniques available through the integration of remote sensing and geographic information systems can facilitate the process of providing the data needed to perform a landslide susceptibility assessment. For example, automatic image classification techniques which use deep learning methods have been shown to be very promising for detecting landslides that have occurred, as well as providing the inventory maps as a prerequisite for quantitative methods [68–70]. Therefore, when feasible, the use of quantitative methods is recommended [4].

Among the quantitative methods, the simple methods (e.g., bivariate statistical methods) are generally inferior to the advanced methods, for example, machine learning methods [71–75]. Even when the landslide inventory data are available, the quantitative methods used at a national scale are often the simple ones [76–79] because the implementation of advanced methods for very large areas with voluminous data is computationally demanding [53,72,80,81] and often impossible in practice. In this study, therefore, an intermediate approach, called normalized frequency-ratio unique condition units (hereinafter called NFUC), is introduced to reduce the volume of data and prepare the data for mapping the landslide susceptibility of a national scale study area using more advanced data mining algorithms, such as LR, ANN, SVM, fuzzy gamma (FG), and C5 decision tree (C5DT).

## 2. Study Area

A large proportion of the landslides in Iran occur on the northern slopes of the Alborz Mountain range opposite the southern shoreline of the Caspian Sea. In this region and the southeast part of the Caspian Sea, the Bandar Torkaman catchment at a national scale (Figure 1) is one of the most susceptible zones to landslides. This catchment covers a large area of 11,593 km<sup>2</sup> between latitudes of 36° 35' 49" to 37° 47' 37" N and longitudes of 53° 59' 59" to 56° 07' 06" E. On the basis of the ASTER

DEM (digital elevation model) of the study area, the altitude varies from  $-28$  m in the northwest lowlands to 3682 m in the mountainous belt which extends in the NE-SW direction. With reference to the 1:250,000 geology maps provided by the Geological Survey of Iran, the main geological structure zones are Gorgan-Rasht in the southern belt of the catchment and Koppedagh in the northern plain areas covered by young deposits. In total, the catchment includes 40 distinct lithological units listed in Table 1 and displayed in Figure 2. The climatic regime of the area changes from an arid to a very humid climate under the effect of an annual average rainfall of 150 mm to 1000 mm, and an annual average temperature that varies from  $4$  °C in the southern parts to  $18$  °C in the northern regions (according to the Forest, Range, and Watershed Management Organization, Iran).

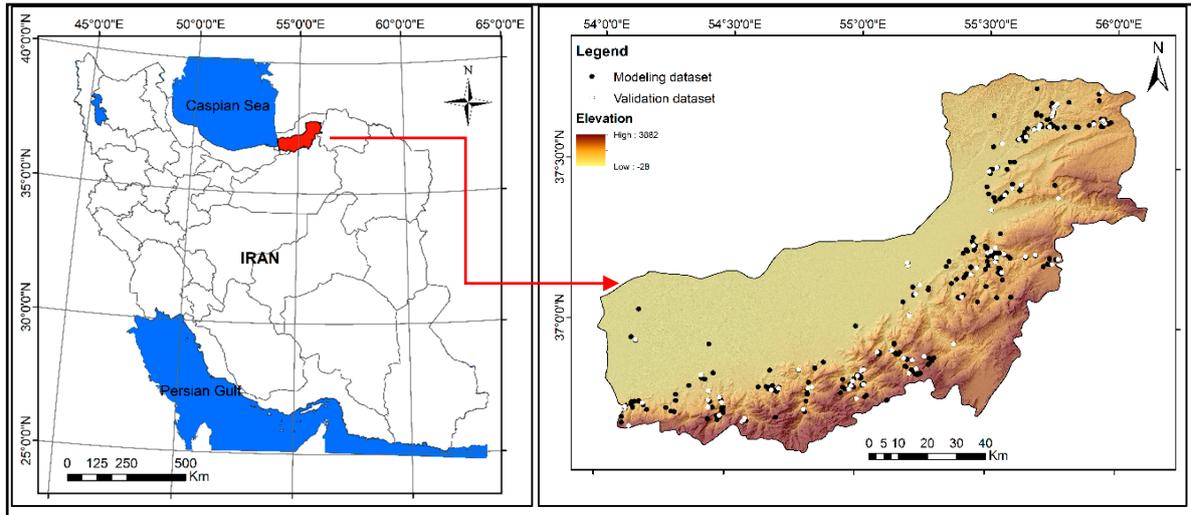


Figure 1. Location of the study area and the occurred landslides.

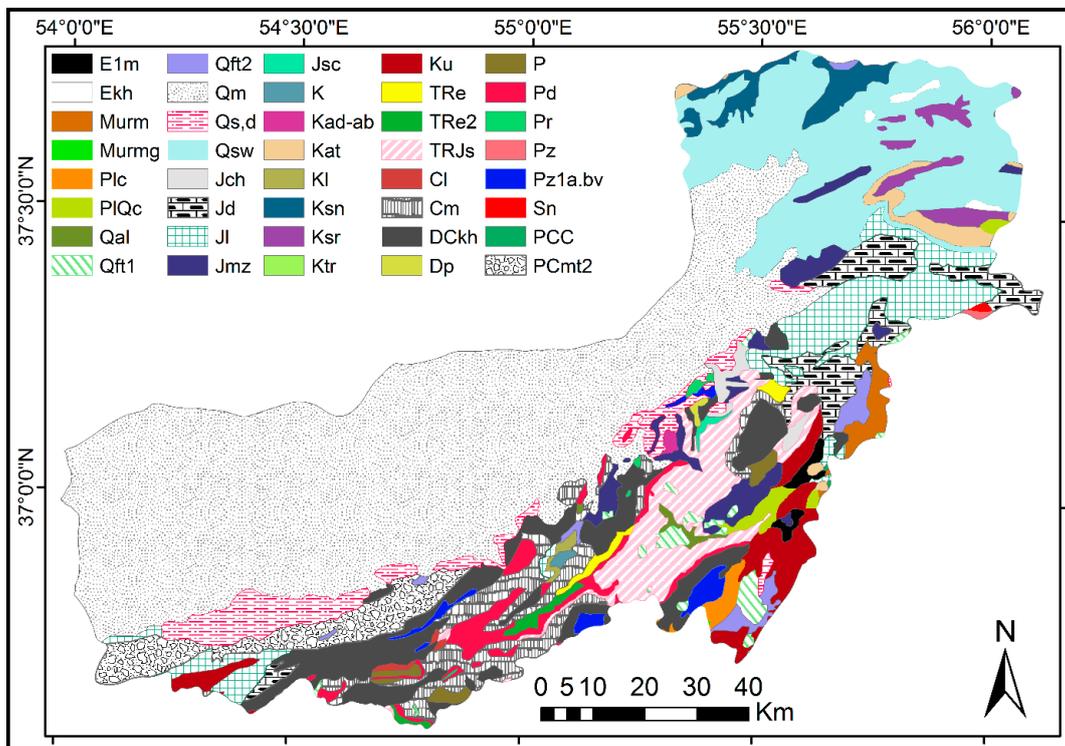


Figure 2. Lithological map of the study area.

Table 1. Details of the lithological units of the study area.

Age	Code	Formation Name	Lithology	
Cenozoic	E1m	–	Marl, gypsiferous marl, and limestone	
	Ekh	Khangiran	Olive-green shale and sandstone	
	Murm	–	Light red to brown marl and gypsiferous marl with sandstone intercalations	
	Murmg	–	Gypsiferous marl	
	Plc	–	Polymictic conglomerate and sandstone	
	PIQc	–	Fluvial conglomerate, piedmont conglomerate, and sandstone	
	Qal	–	Stream channel, braided channel, and floodplain deposits	
	Qft1	–	High-level piedmont fan and valley terrace deposits	
	Qft2	–	Low-level pediment fan and valley terrace deposits	
	Qm	–	Swamp and marsh	
	Qsd	–	Unconsolidated wind-blown sand deposit including sand dunes	
	Qsw	–	Swamp	
	Mesozoic	Jch	Chaman bid	Dark gray argillaceous limestone and marl
		Jd	Dalichai	Well-bedded to thin-bedded, greenish-gray argillaceous limestone with intercalations of calcareous shale
Jl		Lar	Light gray, thin bedded to massive limestone	
Jmz		Mozduran	Grey thick-bedded limestone and dolomite	
Jsc		–	Conglomerate	
K		–	Cretaceous rocks in general, include limestone, marly limestone, Inoceramus bearing	
Kad-ab		–	Undifferentiated unit including argillaceous limestone, marl, and shale	
Kat		Aitamir	Olive green glauconitic sandstone and shale	
Kl		–	Lower cretaceous undifferentiated rocks (Argillite, limestone, massive dolomite, sandstone)	
Ksn		Sanganeh	Grey to black shale and thin layers of siltstone and sandstone	
Ksr		Sarcheshmeh	Ammonite bearing shale with interaction of orbitolina limestone	
Ktr		Tirgan	Grey oolitic and bioclastic orbitolina limestone	
Ku		–	Upper cretaceous, undifferentiated rocks	
TRe		Elikah	Thick bedded gray oolitic limestone; thin-platy, yellow to pinkish shale-limestone with worm tracks and well to thick-bedded dolomite and dolomitic limestone	
TRe2		–	Thick bedded dolomite	
TRJs		Shemshak	Dark-gray shale and sandstone	
Paleozoic		Cl	Lalun	Dark red medium-grained arkosic to sub arkosic sandstone and micaceous siltstone
	Cm	Mobarak	Dark gray to black fossiliferous limestone with subordinate black shale	
	DCKh	–	Yellowish, thin to thick-bedded, fossiliferous argillaceous limestone, dark gray limestone, greenish marl, and shale, locally including gypsum	
	Dp	Pabdeh	Light red to white, thick bedded quartzarenite with dolomite intercalations and gypsum	
	P	–	Undifferentiated Permian rocks	
	Pd	Dorud	Red sandstone and shale with subordinate sandy limestone	
	Pr	Ruteh	Dark-gray medium-bedded to massive limestone	
	Pz	–	Undifferentiated lower Paleozoic rocks	
	Pz1a.bv	–	Andesitic basaltic volcanic	
	Sn	Niur	Greenish gray, shale, sandstone, sandy lime, coral limestone, and dolomite	
	Proterozoic	PCC	–	Late Proterozoic to early Cambrian undifferentiated rocks
PCmt2		Greenschist facies	Low-grade, regional metamorphic rocks	

### 3. Data and Information

The required datasets for landslide susceptibility mapping can be categorized into two groups, landslide inventory map and causative factors (including conditioning factors and triggering agents). The landslide inventory map is the most important factor in landslide susceptibility, hazard, and risk studies [23], and should be as complete and accurate as possible [23,82–84]. The inventory map of the study area was prepared by the Forest, Range, and Watershed Management Organization (Iran) through the interpretation of 1:25,000 aerial photographs and extensive field investigations. This inventory consists of 431 central points of landslides. Generally, archived landslide inventories are recorded as points [85], especially in small-scale areas [86] that are located in the center of the whole body [86] or the rupture zone [1] of landslides. Using a landslide inventory map which is recorded as points does not mean that the produced LSM is less reliable relative to a map produced by an entire area of the landslides [85], especially when statistical models are applied that considerably reduce the uncertainty of the inventory map [87]. The available landslide inventory map is split into two parts for modeling and validation [4,12]. There is no general rule to specify what percentage of landslides should be allocated for modeling and validation. Most of the researchers consider 70% and 30% of landslides for modeling and validation, respectively [35,49,56,57,75,88,89]. In this study, however, using the random procedure [90], 80% of landslides (344 points) were considered for modeling, and the remaining 20% (87 points) were considered for validation of the produced maps because an extra 20% of the modeling dataset was used for testing the data mining models in the implementation phase. This means that about 275 landslide points (64% of all the landslides in the study area) were in fact considered as the modeling dataset and the rest of the landslide points were used for testing and validation of the models in two separate phases (discussed in Section 4).

Concerning the conditioning factors, there is no specific rule for the selection of factors, and it depends on the scale and the geoenvironmental conditions of the study area, the type of the landslides considered in the analysis, and the availability of the data [23]. Taking these criteria into consideration, we provided 12 different conditioning factors which included elevation, slope degree, slope aspect, modified sediment transport index (STI-V), stream power index (SPI), lithology, land cover, distance to linear factors (rivers network, roads, and faults), climate type, and temperature. In addition, the annual average rainfall layer was provided as the main triggering agent.

Elevation, slope degree, slope aspect, and river network layers were all derived from the ASTER DEM with the spatial resolution of 30 m. In addition, we calculated the slope gradient ( $\beta$ ) and the contributing area ( $A$ ) using DEM, and the factors of stream power index (SPI), and sediment transport index (STI) [41,91] were created by means of the Raster Calculator tool of ArcGIS® 10 using the following equations:

$$SPI = A \times \tan\beta, \quad (1)$$

$$STI = \left(\frac{A}{22.13}\right)^{0.6} \left(\frac{\sin\beta}{0.0896}\right)^{1.3}. \quad (2)$$

The SPI index is a useful indicator for erosion caused by surface runoff [38,73] that can contribute to land sliding and the STI factor is used as an indicator for the power of erosive flows [92], and hence the occurrence of landslides [93]. In this study, however, a new factor of STI variations, called STI-V, was produced by modifying the STI factor. This modification was done so that the highest STI-V values were assigned to the pixels on the adjacent belt of the most powerful erosive flows rather than to the flows themselves, because landslides occur adjacent to these flows not inside them. The STI modification was made using ArcGIS® software by entering the STI raster instead of a DEM in Slope Tool, which calculates the STI-V values based on the magnitude of changes of the STI values per the distance unit. The highest STI-V values were assigned to the pixels of slopes near the most erosive flows.

The lithological data for the study area were digitized from 1:250,000 geology maps provided by the Geological Survey of Iran (Figure 2). In addition, the layer of faults was extracted from the geology

maps. The roads were mapped using the topographical maps of the study area at 1:100,000 scale provided by the Iran National Cartographic Center. Afterwards, the layers of faults, roads, and river networks were classified into several classes based on the proximity of the lines ready for the modeling process. The land cover map (Figure 3) was digitized from the national map of soil, pasture, and forest potentialities created by the Iranian Center for National Spatial Planning, with some modification using DigitalGlobe satellite data.

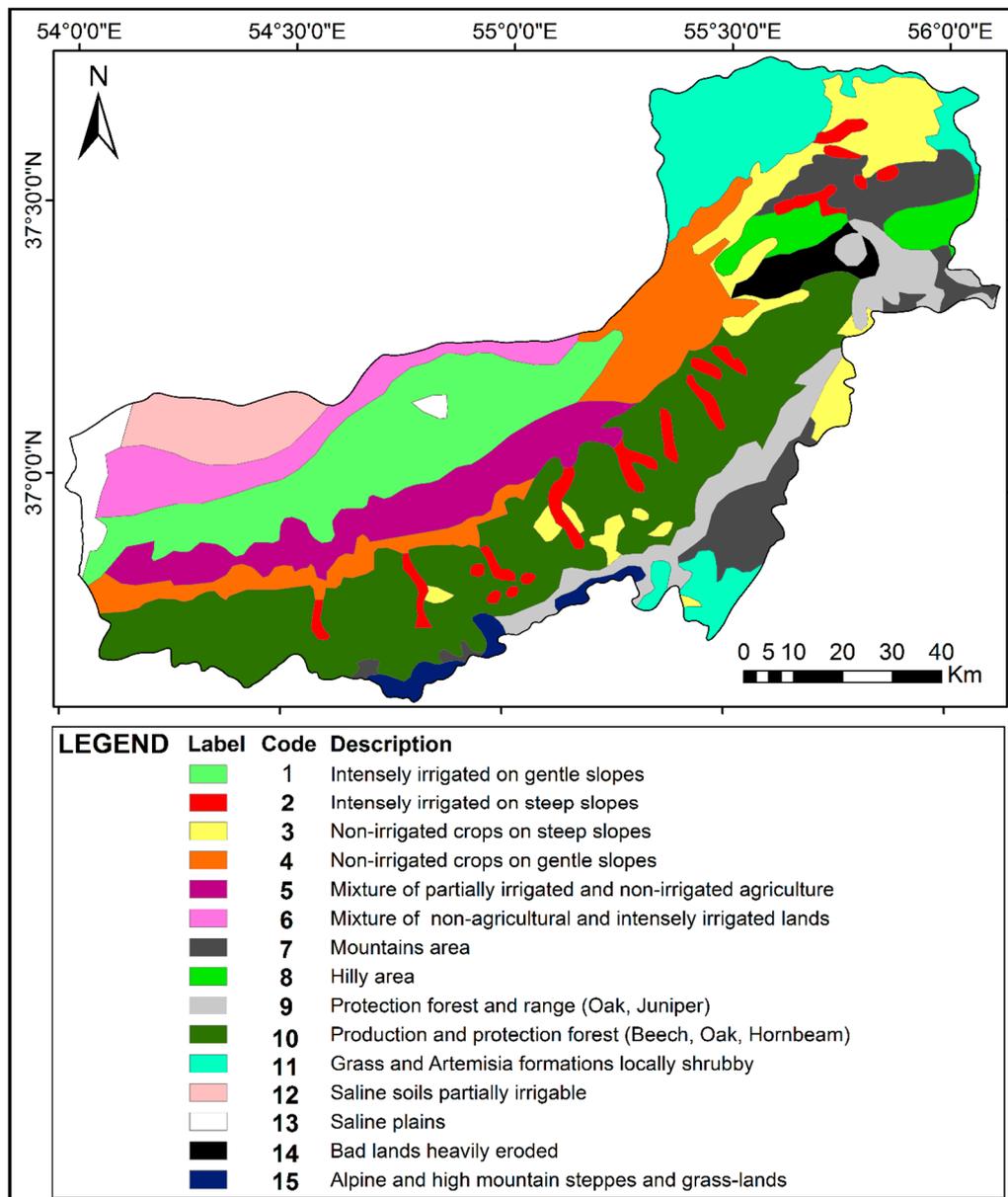


Figure 3. Land cover map of the study area.

The maps of the climate conditions, temperature, and rainfall used in this study were provided by the Forest, Range and Watershed Organization of Iran. The layer of climate conditions was digitized into six classes of humidity, from arid to very humid regions. The digital layer of temperature degree (annual average) consisted of eight classes with a 2 °C increase in each class from 4 °C to 18 °C. In large-scale studies, the layers of climate conditions and temperature degree are counted homogeneous, and hence are unsuitable, but in this research, because of the small scale of the study area, those layers were heterogeneous enough for evaluation of their possible effects on landslide occurrence. Both factors

have been shown to impact landslide occurrence by affecting the soil–atmosphere relationship, that is to say, they can change the soil moisture which in turn can change the pore water pressure, and therefore shear strength of the slopes [94,95]. In addition, an increase of humidity and temperature can both considerably influence the weathering process of rocks and soils and facilitate the production of basic materials for land sliding [96].

The rainfall map, as a triggering factor, was classified into 11 classes based on the annual average values. For LSMs that are produced at a small scale (the area is very large), even if the number of meteorological stations is low or the data does not exist, the amount of rainfall can be assessed by utilizing satellite data [23] because water plays a very important role in landslide triggering, for example, by increasing pore water pressure and lubrication in materials [31].

As mentioned before, in this study, machine learning methods are integrated with the bivariate statistical method of FR through the NFUC approach; in fact, the FR method is supposed to provide the initial weights for running the machine learning models. To apply the FR method, it is necessary to categorize the continuous factors (e.g., elevation and slope degree) into discrete classes. Classification is mainly done based on experts' opinions [56,97]. In this study, the continuous factors were classified considering the geoenvironmental conditions of the study area and the potential effects of each factor on the landslides. Because of the very large area of the catchment, we considered the maximum possible number of classes for each factor to prevent homogeneity of the data, thus, increasing the accuracy of the final LSMs. After classification, all factors were converted to raster format with  $30 \times 30$  m pixel size (equal to the resolution of the ASTER DEM, employed in this study) according to the method of pixel-based modeling [7,12]. The causative factors are shown in Figure 4a–k, except for the lithology and the land cover maps that are shown separately.

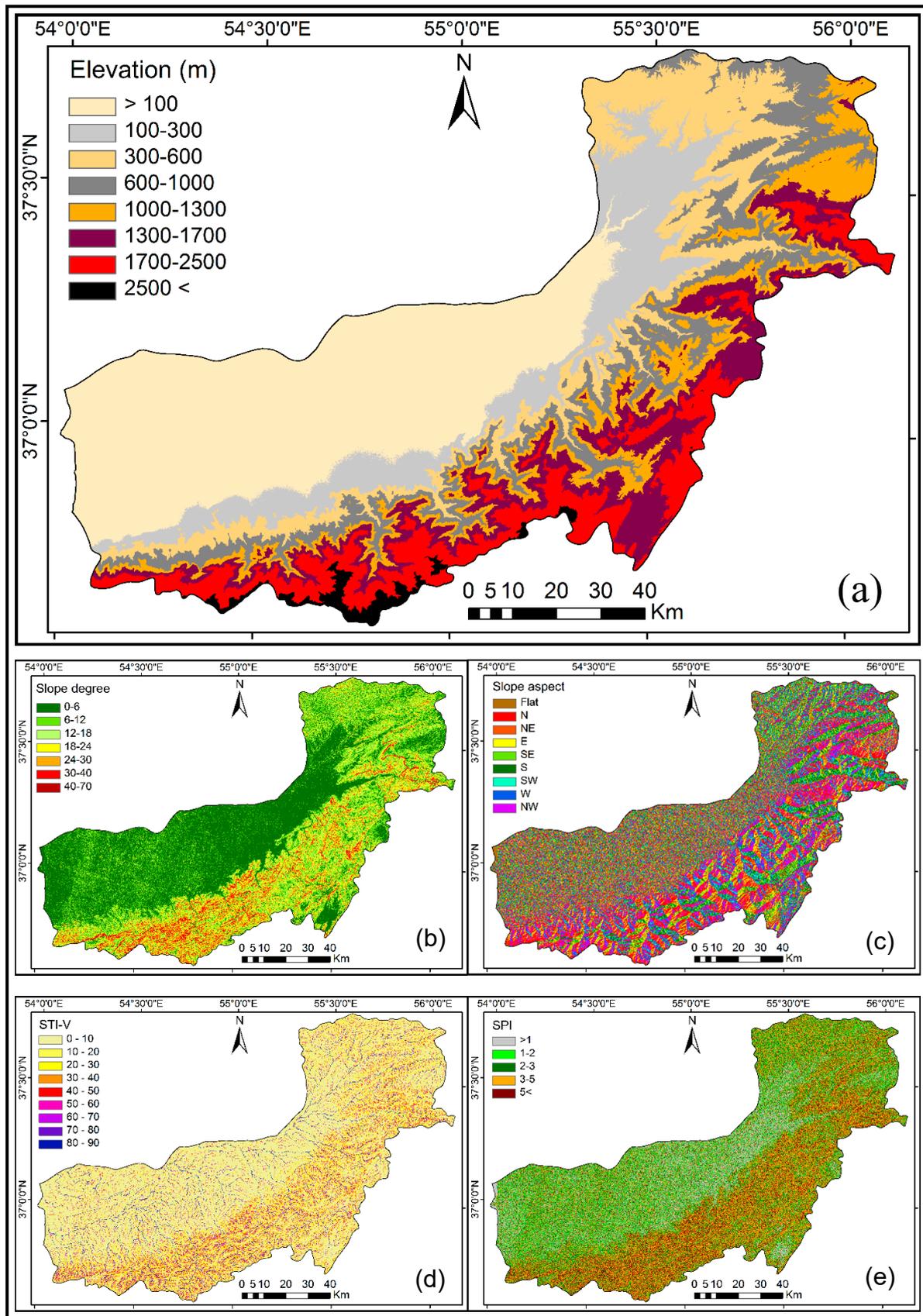
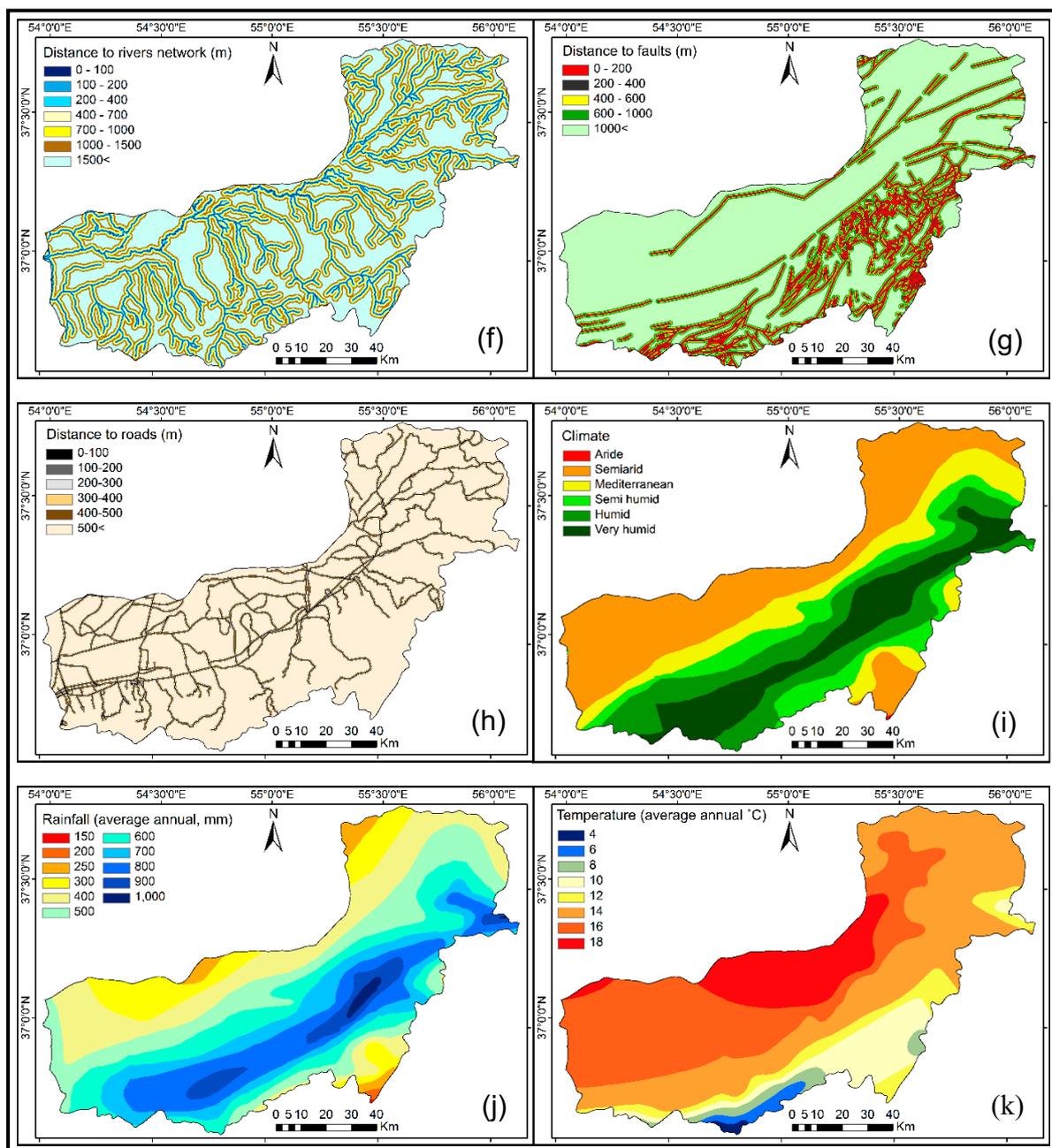


Figure 4. Cont.



**Figure 4.** Prepared causative factors including elevation (a), slope degree (b), slope aspect (c), modified sediment transport index (STI-V) (d), stream power index (SPI) (e), distance to rivers network (f), distance to faults (g), distance to roads (h), climate (i), annual average rainfall (j), and annual average temperature (k).

#### 4. Methods

In this study, each of the five data mining models, namely FG, binary logistic regression (BLR), backpropagation artificial neural network (BPANN), SVM, and C5DT, were separately integrated with the FR model through an intermediate approach called NFUC to produce the LSMs of the study area at a national scale. As mentioned before, the process of running the machine learning methods requires intensive computer processing for landslide susceptibility mapping at a national scale (a large area with big data), which is very time-consuming and sometimes practically impossible [53,72,80,81]. For example, to apply the machine learning methods of BPANN and SVM in this study, it was required to convert the data from raster format to a text file readable by statistical software. When visualizing

the scored text file to the first raster format using the Lookup tool in ArcGIS software, this process took a very long time and was practically impossible using standard computers due to the huge number of pixels (around 12,881,000 pixels). To overcome this problem, the number of pixels can be reduced by increasing the size of pixels. A pixel size from  $26 \times 26$  m to  $1000 \times 1000$  m has been tried for very big areas (at national to global scales) and has been reported to have satisfying results depending on different conditions [60,62,65–67,77,79,98,99], however, it should be noted that increasing the size of pixels is done at the expense of reducing the models' accuracy, i.e., the larger the pixel size, the lower the spatial accuracy of the produced map. In this study, a pixel size of  $30 \times 30$  m was considered to fully exploit the accuracy of available data, for example, the available DEM. Therefore, it resulted in a huge number of pixels which made the process, especially for machine learning methods, practically impossible. Therefore, the NFUC approach was applied to facilitate the application of these methods at this scale by reducing the volume of the data and integrating the models as follows:

- At the first stage, to find the correlation of the landslides with the causative factors and calculate the initial weights [9,100], the bivariate statistical method of FR is applied. In this method, the weight ( $Fr_i$ ) of each class ( $i = 1, 2, 3, \dots, n$ ) of a factor is equal to the percentage of its landslides divided by the percentage of its area as a ratio of the whole map;
- At the second stage, the  $Fr_i$ s (the sixth column of Table 2) should be normalized in the standard interval of (0.1, 0.9) that results in the  $\mu_i$  values (the last column of Table 2) as follows [101,102]:

$$\mu_i = 0.8 \frac{Fr_i - Fr_{min}}{Fr_{max} - Fr_{min}} + 0.1, \quad (3)$$

where,  $Fr_{min}$  and  $Fr_{max}$  are the minimum and maximum observed FR weights among the classes of a given factor. In this step, because the pixels with an equal  $\mu_i$  in a factor layer are equally important in terms of affecting the occurrence of landslides, they can be merged together, which results in separate units. This act reduces the number of computations in the process of converting the data and employing the machine learning models. Additionally, the separate units with very close  $\mu_i$  values (some units showed negligible differences in terms of the  $\mu_i$  value) can also be integrated into single units. Apart from reducing the number of pixels, another advantage of the second stage is that the pixels' values fall in a standard continuous range which means there is no unknown value in relation to the categorical factors and it helps to apply the machine learning models;

- At the third stage, the unique condition units of the study area are created by overlaying all the factors with the  $\mu_i$  values in GIS software (e.g., using the Combine tool in ArcGIS® 10);
- The last stage involves creating a calibration dataset which is comprised of the  $\mu_i$  values of both landslide and stable pixels extracted from the unique condition units. Both landslide and stable pixels are necessary for training some of the data mining models [57,81,93,103,104], such as BLR, BPANN, SVM, and C5DT (except for FG which was applied directly using the unique condition raster). The calibration dataset consisted of 80% landslide pixels (344 pixels in the modeling dataset) and 344 randomly selected stable pixels. A buffer distance of 100 m around the landslides was considered when randomly picking out the stable pixels to provide relative assurance of the insensitivity of these stable pixels, which in turn helped to increase the accuracy of the models. A low-volume text file (such as DBF, database file) of the calibration dataset was used in the training process of the machine learning models in statistical software (SPSS® Statistics 19 and SPSS® Modeler 18 in this study).

**Table 2.** Correlation between the landslide modeling dataset and the classes of causative factors.

Factor	Class	Class Area %	No. of Landslide	Landslide %	$Fr_i$	$\mu_i$
Elevation (m above m.s.l.)	<100	32.29	6	1.74	0.050	0.118
	100–300	12.16	63	18.31	1.500	0.648
	300–600	13.01	98	28.49	2.190	0.900
	600–1000	11.82	89	25.87	2.190	0.900
	1000–1300	9.33	49	14.24	1.520	0.655
	1300–1700	10.34	32	9.30	0.900	0.429
	1700–2500	9.68	7	2.03	0.210	0.177
	2500<	1.33	0	0.00	0.000	0.100
Slope degree	0–6	40.83	89	25.87	0.630	0.226
	6–12	22.55	76	22.09	0.980	0.457
	12–18	15.40	76	22.09	1.430	0.755
	18–24	10.35	59	17.15	1.650	0.900
	24–30	6.17	27	7.85	1.270	0.649
	30–40	4.03	16	4.65	1.150	0.569
	40<	0.65	1	0.29	0.440	0.100
Slope aspect	Flat	0.23	0	0.00	0.000	0.100
	North	17.42	60	17.44	1.000	0.740
	Northeast	12.00	32	9.30	0.770	0.593
	East	8.61	33	9.59	1.110	0.810
	Southeast	10.51	29	8.43	0.800	0.612
	South	13.32	52	15.11	1.130	0.823
	Southwest	11.18	48	13.95	1.250	0.900
	West	10.59	38	11.04	1.040	0.766
Northwest	16.13	52	15.11	0.930	0.695	
STI-V	0–10	59.43	160	46.51	0.780	0.108
	10–20	17.07	68	19.76	1.150	0.185
	20–30	7.41	19	5.52	0.740	0.100
	30–40	4.35	21	6.10	1.400	0.236
	40–50	3.06	10	2.90	0.950	0.143
	50–60	2.41	10	2.90	1.200	0.195
	60–70	2.10	8	2.32	1.100	0.174
	70–80	2.01	14	4.06	2.020	0.365
80–90	2.14	34	9.88	4.610	0.900	
SPI	<1	15.24	22	6.39	0.420	0.100
	1–2	19.29	49	14.24	0.740	0.388
	2–3	21.29	90	26.16	1.230	0.828
	3–5	30.43	121	35.17	1.150	0.756
	5<	13.73	62	18.02	1.310	0.900
Lithology (Code)	E1m	0.35	0	0.00	0.000	0.100
	Ekh	0.22	0	0.00	0.000	0.100
	Murm	0.70	3	0.87	1.250	0.239
	Murmg	0.00	0	0.00	0.000	0.100
	Plc	0.37	0	0.00	0.000	0.100
	PlQc	0.50	0	0.00	0.000	0.100
	Qal	0.21	0	0.00	0.000	0.100
	Qft1	0.89	0	0.00	0.000	0.100
	Qft2	0.87	6	1.74	1.990	0.321
	Qm	40.80	14	4.07	0.100	0.111
	Qsd	3.44	31	9.01	2.620	0.391
	Qsw	12.35	100	29.06	2.350	0.361
	Jch	0.51	6	1.74	3.400	0.477
	Jd	3.52	10	2.90	0.820	0.191
Jl	5.19	20	5.81	1.120	0.224	

Table 2. Cont.

Factor	Class	Class Area %	No. of Landslide	Landslide %	$Fr_i$	$\mu_i$
	Jmz	2.17	6	1.74	0.800	0.189
	Jsc	0.09	0	0.00	0.000	0.100
	K	0.08	2	0.58	7.210	0.900
	Kad-ab	0.09	0	0.00	0.000	0.100
	Kat	0.99	0	0.00	0.000	0.100
	Kl	0.14	3	0.87	6.170	0.785
	Ksn	1.51	1	0.29	0.190	0.121
	Ksr	1.08	1	0.29	0.260	0.129
	Ktr	0.00	0	0.00	0.000	0.100
	Ku	2.20	3	0.87	0.390	0.143
	TRe	0.33	7	2.03	6.130	0.780
	TRe2	0.30	0	0.00	0.000	0.100
	TRJs	4.12	24	6.97	1.690	0.288
	Cl	0.13	1	0.29	2.130	0.336
	Cm	3.89	35	10.17	2.610	0.390
	DCKh	6.24	31	9.01	1.440	0.260
	Dp	0.07	0	0.00	0.000	0.100
	P	0.47	1	0.29	0.610	0.168
	Pd	1.93	7	2.03	1.050	0.217
	Pr	0.12	0	0.00	0.000	0.100
	Pz	0.05	0	0.00	0.000	0.100
	Pz1a.bv	0.68	2	0.58	0.850	0.194
	Sn	0.04	0	0.00	0.000	0.100
	PCC	0.01	0	0.00	0.000	0.100
	PCmt2	3.19	30	8.72	2.730	0.403
	1	12.90	2	0.58	0.040	0.107
	2	3.12	52	15.11	4.840	0.900
	3	6.69	93	27.03	4.040	0.768
	4	8.45	9	2.61	0.300	0.150
	5	7.48	6	1.74	0.230	0.138
	6	5.16	3	0.87	0.170	0.128
	7	6.22	2	0.58	0.090	0.115
	8	2.62	4	1.16	0.440	0.173
	9	5.73	12	3.49	0.600	0.199
	10	24.16	147	42.73	1.770	0.393
	11	9.20	13	3.78	0.410	0.168
	12	3.74	0	0.00	0.000	0.100
	13	1.71	0	0.00	0.000	0.100
	14	1.52	1	0.29	0.190	0.131
	15	1.25	0	0.00	0.000	0.100
	0–100	4.51	72	20.93	4.640	0.900
	100–200	3.76	38	11.04	2.930	0.568
	200–300	4.06	30	8.72	2.140	0.415
	300–400	3.40	29	8.43	2.470	0.479
	400–500	3.58	31	9.01	2.510	0.486
	500<	80.67	144	41.86	0.520	0.100
	0–100	5.41	83	24.12	4.450	0.900
	100–200	4.59	45	13.08	2.850	0.549
	200–400	9.38	37	10.75	1.140	0.175
	400–700	13.08	36	10.46	0.800	0.100
	700–1000	11.74	39	11.33	0.960	0.135
	1000–1500	16.85	54	15.69	0.930	0.128
	1500<	38.93	50	14.53	0.370	0.100

Table 2. Cont.

Factor	Class	Class Area %	No. of Landslide	Landslide %	$Fr_i$	$\mu_i$
Distance to faults (m)	0–200	9.30	61	17.73	1.900	0.900
	200–400	7.85	43	12.50	1.590	0.738
	400–600	6.81	47	13.66	2.000	0.952
	600–1000	10.54	55	15.10	1.510	0.696
	1000<	65.49	138	40.11	0.610	0.225
Climate (type)	Very humid	15.32	97	28.19	1.840	0.900
	Humid	19.95	83	24.12	1.210	0.626
	Semi humid	13.37	25	7.26	0.540	0.335
	Mediterranean	15.24	85	24.71	1.620	0.804
	Semiarid	35.55	54	15.69	0.440	0.291
	Arid	0.02	0	0.00	0.000	0.100
Annual average rainfall (mm)	150	0.01	0	0.00	0.000	0.100
	200	0.15	0	0.00	0.000	0.100
	250	2.11	0	0.00	0.000	0.100
	300	9.73	1	0.29	0.030	0.109
	400	20.36	31	9.01	0.440	0.225
	500	19.40	99	28.78	1.480	0.521
	600	16.48	31	9.01	0.540	0.254
	700	13.03	41	11.92	0.910	0.359
	800	13.92	97	28.19	2.020	0.675
	900	4.13	40	11.62	2.810	0.900
Annual average temperature (°C)	1000	0.64	4	1.16	1.820	0.618
	4	0.34	0	0.00	0.000	0.100
	6	1.42	2	0.58	0.400	0.269
	8	2.36	3	0.87	0.370	0.257
	10	8.88	23	6.68	0.750	0.417
	12	7.14	21	6.10	0.850	0.460
	14	34.92	227	65.99	1.890	0.900
	16	32.01	64	18.60	0.580	0.346
18	12.90	4	1.16	0.090	0.138	

When using the calibration dataset in the statistical software, 80% of data was engaged in training and 20% in testing the models. Apart from the validation dataset that had been preserved for the final validation of the models, applying this 20% testing proportion provided a preliminary performance evaluation when the BLR, BPANN, SVM, and C5DT models were executed.

#### 4.1. FG

The fuzzy set theory was introduced by Zadeh [105]. The term “set” in this method refers to the range of values, between 0.1 and 0.9 in this study, which can be assigned to different members. This method can be used qualitatively or quantitatively as a flexible method, depending upon the source of the fuzzy membership values, which was the quantitative  $\mu_i$  values in this study. There is a direct relationship between the membership value and susceptibility to landslides, that is to say, a minimum membership value shows the minimum susceptibility of a pixel to landslides, and vice versa. After assigning the relevant  $\mu_i$  values to the pixels, the layers of the causative factors can be combined by using one of the fuzzy functions (OR, AND, SUM, PRODUCT, and GAMMA) to calculate the probability of landslide occurrence (P) for each output pixel (x). The GAMMA function that is reported to have the best results among all fuzzy functions [42,101,102,106–108] is used here with the equation of

$$P_{\text{GAMMA}}(x) = [1 - \prod_{i=1}^n \mu_i(x)]^\gamma \times [\prod_{i=1}^n \mu_i(x)]^{1-\gamma}, \tag{4}$$

where,  $\gamma$  can range from 0 to 1. The  $\gamma = 0$  and the  $\gamma = 1$ , respectively, results in the minimum and maximum possible P of the pixels [108,109]. In this study, different  $\gamma$  values of 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, and 0.975 are tested to see which one would produce a more reliable LSM.

#### 4.2. BLR

As a multivariate statistical method, logistic regression is reported to construct one of the most reliable landslide susceptibility models [35,43,48,110,111]. The BLR method calculates the probability of a two-category dependent variable, such as landslide occurrence, regarding its relationship with causative factors as the independent variables [112,113]. The relationship between the dependent variable and the independent variables is a nonlinear form of correlation. The probability of occurrence of a dependent variable (P) based on a set of given independent factors in the LR model is expressed as follows:

$$P = 1/(1 + e^{-z}), \quad (5)$$

where, P, on an S-shaped curve, ranges from 0 to 1 in a direct relationship with the variation of the parameter Z from  $-\infty$  to  $+\infty$ . And the parameter Z is specified as

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n, \quad (6)$$

where  $\beta_0$  is the model intercept, and  $\beta_i$  ( $i = 1, 2, 3, \dots, n$ ) is the coefficient of each given factor ( $X_i$ ) [114]. By assigning the probability values calculated in this model to the related pixels, the LSM of the area is produced.

To determine the accuracy of the model, two types of R-squared can be considered, the Cox and Snell R-squared and the Nagelkerke R-squared. The first can vary from zero to a maximum of less than one for a perfect model [115], and the second, which is an adjusted version of the first one, falls within the range of zero to one [116]. The higher values for the above two indices show the better results.

#### 4.3. BPANN

Artificial neural networks benefit from nonlinear mathematical algorithms to mimic the learning process of the human brain in dealing with complex issues [50,117–119]. The main advantage of this model is that there is no need to necessarily engage a specific statistical variable because, in fact, this model is independent of the data statistical distribution [53,120]. The LSMs are produced by a trained ANN as a feedforward structure [121,122]. The BPANN model, constructed in this study, was a multilayer perceptron network trained by the backpropagation algorithm [123] using the optimization algorithm of gradient descent with a momentum parameter [124].

Structurally, the network is comprised of an input layer, hidden layer(s) with different numbers of neurons, and an output layer. Neurons of the input layer can be scaled, categorical, or binary data [125]. As mentioned before, the prepared calibration dataset consisted of the  $\mu_i$  values (the normalized FR weights) of both the landslide pixels and the stable pixels as the necessary input data for training the model [93,103]. The calibration points were divided into two parts, training and test dataset, that were very important in modification of the weights within the network and evaluation of the network prediction power, respectively [54,126]. Because there is no general rule [93], 80% of the calibration points were assigned to the learning dataset and 20% to the test dataset, as suggested by Swingler [127]. The ideal number of hidden layers and associated neurons for enhancing the network performance was selected through trial and error [122,128]. After running the model, the training process was iterated to modify the weights in the input layer until one of the specified stopping rules occurred. As the stopping rules in this study, the total number of iterations was set at 2000 and the maximum iterations without an error reduction at 10.

The other important parameters of the network are the initial learning rate and the momentum factor, which were set at 0.01 and 0.9, respectively, after considering the values given in the literature and testing different values. A high learning rate leads to a high-speed learning process, however, with

a higher degree of uncertainty [50,103]. In addition, momentum value is defined to prevent the possible network instability originated from a high-speed learning process [50,103].

#### 4.4. SVM

SVM is a machine learning method that has been recently used in landslide susceptibility analysis [57,58,81,129]. It classifies the input training points (calibration dataset) of the landslide causative factors ( $F_i = F_1, F_2, \dots, F_n$ ) into two classes of stable ( $P_i = -1$ ) and unstable ( $P_i = 1$ ) pixels using the optimal hyperplane (an n-dimensional surface) with the widest possible space between the margins of the nearest points. In linear form, the equation of a hyperplane can be written as follows:

$$P_i(W \cdot F_i + b) \geq 1, \quad (7)$$

where,  $b$ , as a constant value, shows the offset of the hyperplane from the origin. The Euclidean length between the hyperplane and each of the margins is  $\frac{1}{2}\|W\|^2$  [56,59] that is used in the Lagrangian equation to define the optimal hyperplane [57] as follows:

$$L = \frac{1}{2}\|W\|^2 - \sum_{i=1}^n \lambda_i(P_i(W \cdot F_i + b) - 1). \quad (8)$$

In the above equation,  $\lambda_i$  is the Lagrange multiplier. More information and the detailed equations of the SVM method can be found in Hong et al. (2016). Four different types of kernel functions can be used for an SVM model such as linear, polynomial, sigmoid, and radial basis function (RBF). The last one is frequently reported as the best function for landslide susceptibility mapping studies [56–58]. The balance between accuracy and overfitting of the model can be adjusted by the regularization (C) and gamma parameters. The C is usually set between one and 10; the higher the C is, the more accurate the model would become, but it may cause overfitting of the model. The same trend is seen in terms of the RBF gamma parameter; nevertheless, values in the range of three to six divided by the number of the input factors are worth trying [130].

#### 4.5. C5DT

C5 is one of the most powerful algorithms used in decision tree models that has recently been employed in landslide studies [59,131]. This algorithm is the new version of the old C4.5, which had been reported to be the fastest machine learning method [132]. C5 is even faster than the traditional C4.5 algorithm, it is efficient in terms of both memory usage and weighting process, and it benefits from two options of boosting and winnowing [133,134]. Selection of the boosting option helps to significantly increase the accuracy of the model by building a number of consecutive models that focus on the misclassified records of the preceding model, and the winnowing option helps to prune ineffective factors before construction of the model, thus, increasing the speed of assessment, which is a considerable advantage, notably, in dealing with big data [135–138].

Depending on the amount of information each causative factor reflects about the landslide occurrence, the C5 algorithm separates the input training points of the factors. This process begins with the factor that reveals the maximum information about the occurrence of landslides. Then, each of the created subgroups are split up again based on another important factor, and this process continues by taking all other factors into account, one by one. The lowest-level branches of the decision tree are then pruned if they do not enhance the results of the model significantly. The process of pruning the tree is done in two stages, local pruning and global pruning [59]. Local pruning evaluates the subtrees and prunes the branches, and global pruning treats the tree as a whole with some weak subtrees that should be collapsed. To control the severity of the pruning in the local stage, the pruning severity parameter of the model is set between zero and 100 in SPSS® Modeler software. The higher the parameter, the smaller the tree would be and it can prevent the model from overfitting. Finally, the decision tree created is used to score the whole pixels of the study area.

#### 4.6. Validation of the Built Models

To assess the reliability of landslide susceptibility models, the receiver operating characteristic (ROC) method is recommended [3,139,140]. The area under the curve (AUC) of the ROC graph is used as a scalar statistic [141,142] to indicate the validation rates of the models. To calculate the AUC, several thresholds ( $i = 1, 2, 3, \dots, n + 1$ ) are defined for each of which the sensitivity and specificity statistics are calculated as follows [42,139,141]:

$$\text{Sensitivity}_i = \frac{L_{>i}}{TL}, \quad (9)$$

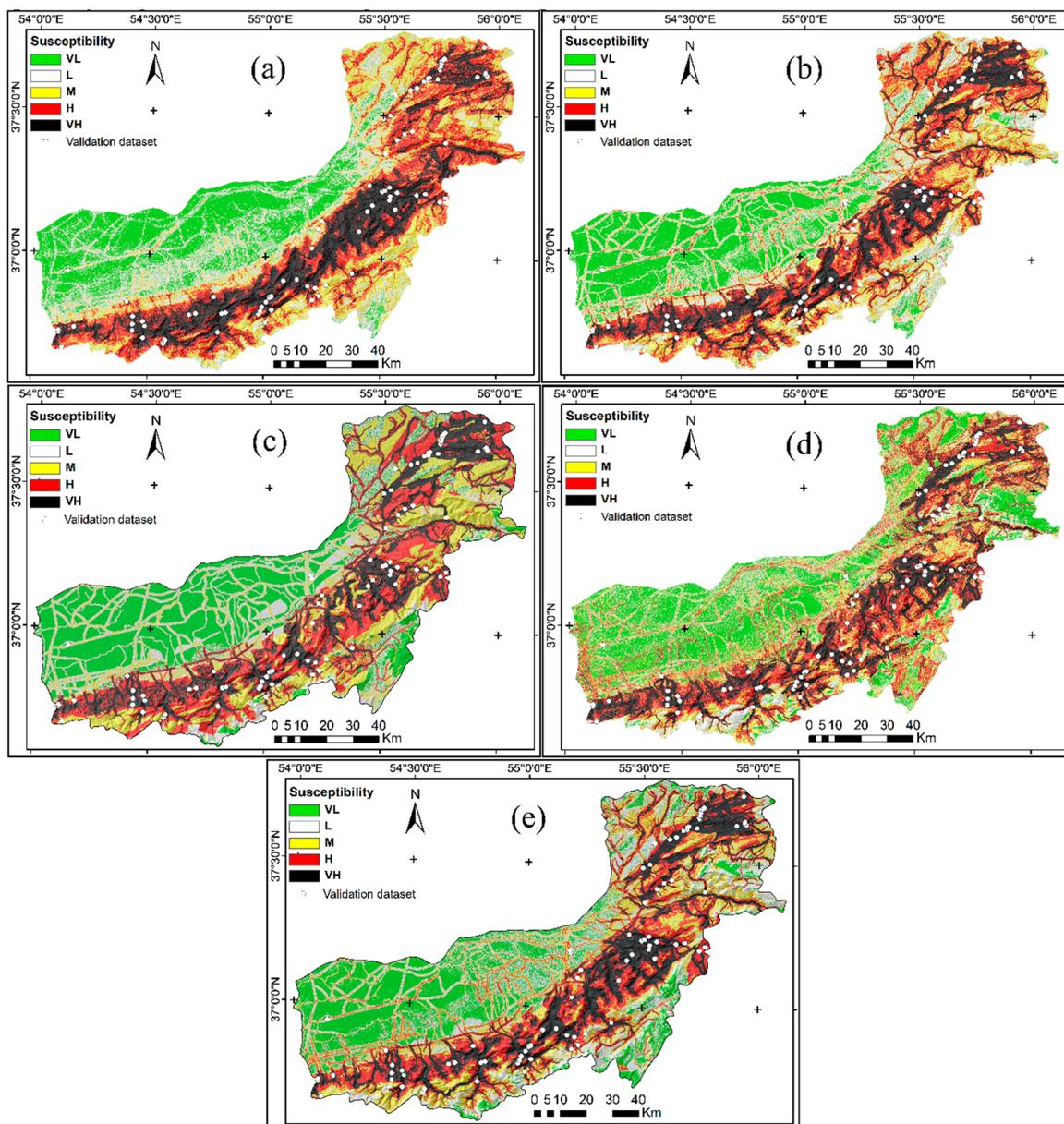
$$\text{Specificity}_i = \frac{S_{>i}}{TS}, \quad (10)$$

where,  $L_{>i}$  is the number of landslide pixels with a value higher than that of the threshold, and  $S_{>i}$  is the number of stable pixels with a value lower than that of the threshold. TL and TS are the total number of the landslide pixels and the stable pixels in the map, respectively. Plotting the sensitivity and 1-specificity of each threshold on the  $y$ -axis and  $x$ -axis, respectively, the AUC of the ROC graph is calculated by the following equation [80,139]:

$$\text{AUC} = \sum_{i=1}^{n+1} \frac{1}{2} \sqrt{(x_i - x_{i+1})^2} \times (y_i + y_{i+1}). \quad (11)$$

The calculated AUC shows the success rate of the model if the modeling dataset is engaged in Equations (9) and (10), and the prediction rate of the validation dataset. Both rates are required to be evaluated [90,143]. The success rate shows how well the model classifies the areas of existing landslides but not future landslides, therefore, the prediction rate is calculated [41,72,90,144].

After the validation process is done, the outputs of the models that are continuous numerical values should be visualized, preferably not in more than five zones for clarity [3]. In this regard, five main techniques can be used such as simple ranking, natural breaks, the mean value and standard deviation intervals, equal interval classes, and equal area classes [90,145,146]. The equal area classes technique is more suitable than the others for comparison of the maps [42,53,90]. Using this technique and considering the geoenvironmental situations of the study area, an equal proportion of 20% of the area was assigned to each of the zones of the maps, very low, low, moderate, high, and very high susceptibility (Figure 5a–e) for a straightforward comparison.



**Figure 5.** Landslide susceptibility maps produced by the models of fuzzy gamma (FG) (a), binary logistic regression (BLR) (b), C5 decision tree (C5DT) (c), backpropagation artificial neural network (BPANN) (d), and support vector machine (SVM) (e).

## 5. Results and Discussion

### 5.1. The Relationship between the Landslides and Causative Factors

By applying the FR method, the correlation between the landslides and the causative factors was examined. The higher the  $Fr_i$  of a class is (Table 2), the stronger the correlation between that class and the landslides would be, and vice versa. In the case of altitude, the middle classes of altitude (300–600 and 600–1000 m) with equal  $Fr_i$ s (2.19) were the most sensitive zones. After these, altitudes of 100–300 and 1000–1300 m by the same weight of 1.5 also showed a meaningful relationship with landslide occurrence, but other classes of altitude were not susceptible ( $Fr_i < 1$ ). In all probability, the main reason for the higher weights of the middle altitude classes was the interaction effect of other important causative factors such as the high amount of precipitation, the potential slope degrees, and the existence of loose soils and stones that accompanied these classes.

In terms of slope degree, the areas with a slope degree from 12 to 40, which covered about 36% of the study area, were susceptible to landslides. On the contrary, the two classes of zero to six degree and six to 12 degree (owing to their low shear stress) and the class  $40 < \text{degree}$  (due to gradual fall of the unstable materials and hence existence of the weather-resistant rocks) comprised the low-risk regions.

The slope aspect factor as a geomorphological attribute can influence the occurrence of landslides [147], however, similar to some other studies [43,80,148], in this study, the  $Fr_i$ s of the classes did not reveal any clear correlation between the slope aspects and occurrence of landslides. Nevertheless, the flat areas ( $Fr_i = 0$ ) and the northeast and southeast aspects (with weights of well below one) showed an inverse correlation with the occurrence of landslides.

The two last classes STI-V factor had a strong correlation with the occurrence of landslides. The last class of this factor, 80–90, showed the highest weight, 4.6, with a covering just above 2% of the whole area encompassing about 10% of the landslides. These landslides were those which were highly affected by the toe erosion process of the powerful rivers. Therefore, STI-V (the modified version of STI factor) can have a high density of landslides in its classes with the highest values, however, the other classes with lower values did not show a clear relationship with the landslides because the main cause of the landslides occurring in these classes was the effect of other factors rather than the rivers.

When examining the weights of the classes of the SPI factor, the last three classes with the highest weights showed a relatively high susceptibility to landslides, although not exponentially, as seen in the case of STI-V. They had similar weights (just over one). This shows that this factor is not as good as the STI-V factor for indicating landslides affected by rivers.

In the case of lithology, the most sensitive lithological unit was K (described in Table 1), with a weight of 7.21, followed by the K1 and TRe units with weights of 6.17 and 6.13, respectively. Other very susceptible units were consecutively Jch, PCmt2, Qsd, Cm, Qsw, and Cl, which all had a weight over two. By considering the lithological combination and the spatial range of units, it was observed that most of the susceptible units had some sensitive materials inside (often limestone and marl) and were located in the regions with potential conditions for land sliding (e.g., on steep slopes with a high amount of rainfall and high density of rivers, roads, and faults).

Concerning the land cover factor, the two most susceptible classes (2 and 3) both consisted of lands on steep slopes which were deforested for intensely irrigated to non-irrigated farming. The weights of these two classes (about 4.8 and 4, respectively) were about double the weight of their surrounding dense forest (i.e., class 10 which was the only other landslide-prone class). Not surprisingly, the landslides in dense forest often happened very close to the roads, notably those constructed on the steep slopes and alongside the rivers. The above results reveal that landslides can be significantly affected by human activities such as deforestation for cultivation and construction of roads in forests, for example, for carrying wood [149,150]. The nearest buffer zone from roads (0–100 m) had the highest weights (about 4.6) which strongly supports the assumption about the profound influence of roads on landslides. The four next classes covering a distance of 100 to 500 m from roads were also very susceptible to landslides, all having a weight above two; the only stable buffer zone had a distance of more than 500 m from roads.

For rivers, similar to roads, the highest weights belonged to the classes with the closest distance to the network of rivers; the classes of 0 to 100 m and 100 to 200 m were much more liable to landslides, and the only other susceptible class was (200 to 300 m). This confirms the significant influence rivers have on the landslides; rivers can promote the occurrence of landslides, for example, by eroding the toe of slopes and affecting the groundwater table.

In the case of distance to faults, however, no certain relationship can be seen between the distance of the classes and the landslides; except the class of  $>1000$  m that was relatively stable. Other classes with different ranges of distance to faults were all similarly susceptible to landslides with some differences in terms of weight.

In relation to climate, as expected, the very humid areas showed the highest susceptibility to landslides ( $Fr_i = 1.84$ ). Surprisingly, after that class, it was the Mediterranean class that was the second

susceptible class with an  $Fr_i = 1.62$ , and not the humid areas ( $Fr_i = 1.2$ ) because climate factor, like any other causative factor, is not the absolute predictor of landslides and may show unexpected weights under the effect of other overlapping factors.

Consecutively, the most susceptible rainfall classes are those with an annual average of 900, 800, and 1000 mm, whereas the four initial classes with the lowest amount of rainfall (150, 200, 250, 300 mm) showed a  $Fr_i$  of about zero which is considered normal. The higher weight of the rainfall class with 500 mm ( $Fr_i = 1.48$ ) in comparison to its two upper classes (with higher precipitation) indicated that, although water plays important roles such as the lubrication effect [151] in triggering landslides, the interaction between the causative factors still plays an influential role in the determination of the weights of classes. This may also be true in the case of the annual average temperature factor, for which the weight of greater than one, for the only susceptible class (14 °C), appears to be largely due to the effect of other accompanying important factors.

### 5.2. Application of the NFUC Method

As mentioned before, it was impossible to apply most of the data mining models to the raw pixel-by-pixel rasters of the study area because of the high volume of data. Therefore, the intermediate approach of NFUC was designed and employed to reduce the volume of data. Results showed that, with the aid of the NFUC approach, the number of raw pixels (30 × 30 m) in the whole study area was considerably reduced to about one-fourth, that is to say, the very large initial number of around 12,881,000 pixels of the study area decreased to about 3,385,000 unique condition units (each of which comprising the pixels with the same  $\mu_i$  values). When the NFUC approach is applied, a decrease in data volume mainly depends on the number of classes considered for each causative factor, i.e., the lower the number of classes, the more the reduction of data volume. In this study, since the highest possible number of classes for each factor was considered to prevent the homogeneity of data, the data volume reduction was finally about 75%, which is still a considerable percentage. The NFUC approach is expected to have the potential to reduce data volume even more in other studies where the number of classes for each factor is usually considered relatively lower. A decrease of about 75% in the volume of data made the implementation of all the considered data mining methods in this research possible.

Although the application of the NFUC approach and transfer of the data between the statistical and GIS software was relatively time-consuming, it possessed the advantage of employing more advanced models at a national scale that were impossible to use with conventional computers due to the high volume of information at this scale. In addition, one could question the way the NFUC approach introduces the predefined pixel weights ( $\mu_i$ ) as the raw input data to the data mining models, however, it should be considered that these weights are not defined arbitrarily but through a statistically significant approach. A  $\mu_i$  value is representative of the common feature of a bunch of pixels, which is their relationship with the occurred landslides (modeling dataset). Therefore,  $\mu_i$  values can be used in other data mining methods in order to further process the weights. Using the NFUC approach, models were successfully constructed and reliable LSMs were produced, which are discussed in the next sections.

### 5.3. Application of Different Data Mining Methods

Among all the  $\gamma$  values applied in this study (0.5, 0.6, 0.7, 0.8, 0.9, 0.95, and 0.975) for the FG model, the value of 0.9 produced an output layer with the widest range of susceptibility degrees, from 0.056 to 0.873 (Table 3). By applying a higher  $\gamma$  value, the output map did not contain the very low susceptibility degrees. On the contrary, using a lower  $\gamma$  value, the generated map did not include the high susceptibility degrees. A similar result has also been reported by Tangestani [108]. Therefore, the map of  $\gamma = 0.9$  was compared with the maps of other methods.

**Table 3.** Range of the output susceptibility index of different  $\gamma$  values in the FG method.

$\gamma$ Value	The Lowest Output Susceptibility	The Highest Output Susceptibility
0.5	0	0.294
0.6	0	0.376
0.7	0	0.480
0.8	0.003	0.613
0.9	0.056	0.873
0.95	0.218	0.885
0.975	0.429	0.940

To construct the BLR, BPANN, SVM, and C5DT models, the calibration dataset with a text format was imported to SPSS<sup>®</sup> Statistics 19 and SPSS<sup>®</sup> Modeler 18. The summary of the BLR model shows that the Cox and Snell R-square and Nagelkerke R-square values were 0.444 and 0.592, respectively, testifying to the good results of the model. With regards to the BLR model, the significance probability values of all factors were lower than 0.05, except for STI-V, SPI, climate, rainfall, and temperature. A value of <0.05 for the significance probability implied that the factor statistically affects the occurrence of landslides.

With respect to the BPANN model, the best result was achieved when one hidden layer with eight units was applied, the initial learning rate was 0.01, and the momentum factor was set at 0.9. However, the mentioned network was executed ten times with different random seeds at a training and test ratio of 80:20 to obtain the best possible results. The results showed that the training and test accuracies of the best model (Table 4) are 91.4% and 88%, respectively, with an overall incorrect prediction of only 9.28%, and therefore this model was selected to produce the related LSM.

**Table 4.** The accuracy of the BPANN models performed using different random seeds.

BPANN Models	Training Accuracy (%)	Test Accuracy (%)	Overall Incorrect Predictions (%)
1	85.7	81.8	15.08
2	86.4	84.3	14.02
3	84.9	85.5	14.98
4	87.7	84.8	12.88
5	80.7	82.1	19.02
6	85.9	80.9	15.01
7	87.9	83.3	13.02
8	91.4	88.0	9.28
9	82.6	85.1	16.09
10	87.2	80.7	14.10

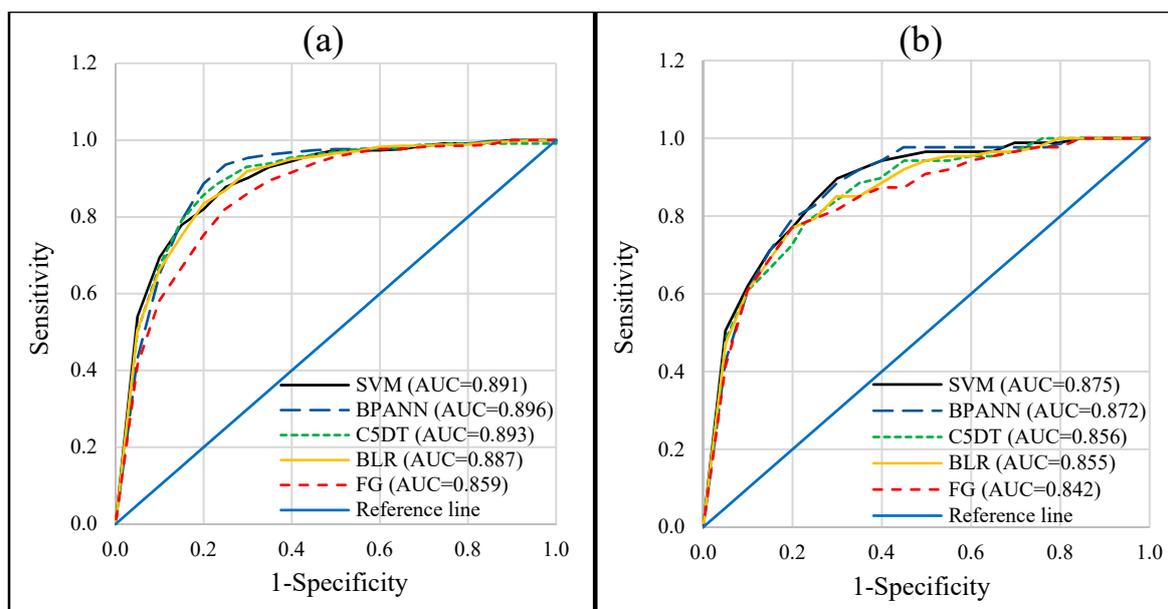
The highest accuracy without being overfit for the SVM model were achieved when the parameter C was set on one and the RBF gamma was 0.23. The training and test accuracies of this model were 84.26% and 79.73%, respectively. Therefore, this model was used to create the LSM of the area. Evaluation of the posterior probability histogram and the training and test accuracies of the predictions for the input calibration dataset showed that increasing the parameters C and RBF gamma led to overfitting of the model. On the one hand, by increasing these parameters, posterior probability of most of the predictions tended to be very close to zero or one, which means if an LSM had been produced, it was unable to predict the middle range of susceptibility values. Likewise, in that case, the training accuracy was higher than the test accuracy, which could show the models were overfitted to the existing landslides and had a low capability to predict future events. On the other hand, with an RBF gamma lower than 0.23, both training and test accuracies of the model declined.

In the case of the C5DT model, the best results were obtained when the boosting and winnowing options were activated, the pruning severity was 100 (the highest possible value), and the number of trials for the boosting method was 14. Under these conditions, the model considered the four causative

factors of slope degree, temperature, STI-V, and SPI not to be significantly effective, and therefore dropped them from the process. With a pruning severity lower than 100, the model showed a high propensity to get overfitted.

#### 5.4. Validation of the Data Mining Models

The ROC graphs in Figure 6a,b illustrate the success and prediction curves of the models, respectively.



**Figure 6.** Success (a) and prediction (b) rate curves of the models based on the receiver operating characteristic (ROC) method.

Overall, although a small-scale landslide susceptibility map is generally less reliable than a larger scale map [4,35], the LSMs produced in this study were all satisfactorily reliable and showed validation rates between 0.8 and 0.9, which can be categorized as a good accuracy [152].

In ascending order, the success rates of the models were 0.859 for FG, 0.887 for BLR, 0.891 for SVM, 0.893 for C5DT, and 0.896 for BPANN, and the prediction rates were 0.842 for FG, 0.855 for BLR, 0.856 for C5DT, 0.872 for BPANN, and 0.875 for SVM. Notwithstanding the use of the same normalized FR weights, the performance of the FG model was comparatively low where its success rate was about 3% lower than that of four other models. For the other models, there was less than a 1% difference in the success rates (BPANN was the best). In terms of predicting future landslides, however, the prediction rates of the models BPANN and SVM were similarly better (about 2% to 3% higher) than those of the FG, BLR, and C5DT.

It should be noted that in the case of the C5DT model, despite controlling the parameters to prevent overfitting, it was the most overfitted model by showing a relatively large difference between its success and prediction rates (about 4%). Generally, the higher this difference, the more overfitted the model, in other words, the model performs well in zoning the area based on the current landslides (modeling dataset), but it is not as successful for predicting future landslides (validation dataset) as well. Considering this criterion, the SVM model had the best results because the difference between its success rate and prediction rate was only about 1.5 percent. Nevertheless, because C5DT was the fastest machine learning model, uses fewer factors as compared with other models (nine of the 13 prepared factors in this study), and it has been proven to be a reliable model in other studies [131,138], it is worth trying this model in other studies especially for large areas with a large amount of data where the speed of model is important.

The percentage of landslides in each susceptible zone is given in Table 5. Because the area percentages of the zones were the same (20%), the percentage of landslides that fell in the zones can be compared directly. Generally, the greater the percentage of landslides inside a very high susceptible zone, the better the map would be. Regarding this matter, the BPANN model was the best model because the percentage of landslides in a very high susceptible zone of its LSM was 87% (about 6% higher than that of the SVM, C5DT, and BLR). The worst results were obtained for the FG LSM model that encircled only 75% of the landslides in its very high susceptible zone. By considering both high and very high susceptible zones together, the BPANN LSM, again, had the best results because it could encompass more than 96% of the landslides in the two mentioned zones, whereas SVM, C5DT, and BLR encompassed about 94%, and FG encompassed about 90%.

**Table 5.** The percentage of landslides in each zone of the landslide susceptibility maps (LSMs) produced by different data mining methods.

Susceptibility Zone	Zones Area %	Landslides %				
		FG	BLR	C5DT	SVM	BPANN
Very low	20	1.62	0.93	0.70	0.93	1.39
Low	20	1.39	1.39	1.86	1.86	0.93
Medium	20	6.26	3.94	3.25	2.78	1.16
High	20	14.85	11.60	11.37	13.46	9.51
Very high	20	75.87	82.13	82.83	80.97	87.01

All things considered, BPANN followed by SVM are the best models, although it should be noted that all other employed models also produced a reliable LSM at a national scale.

The results of the models in this study can be compared to the same models employed in other studies, however, at a national scale (like the scale of this research), only some of the data mining methods applied in this research have been used in other studies. For example, the BPANN model has been employed with satisfactory results producing a national-scale LSM of China [46]. Some studies have reported reasonable accuracy of small-scale LSMs produced by the logistic regression method in different conditions [78,134,135]. For further comparison, the results could be compared with studies at scales different than the scale of this research, such as regional and local scales. At these different scales, SVM with RBF function has high reliability and has been shown to be better than models such as decision tree and Bayesian network [50], neuro-fuzzy inference system, and generalized additive model [36], and C5DT model [39]. The C5DT model has been proven, however, to be a reliable model for predicting landslide-prone areas in two other studies [111,118], hence, once more as a suggestion, it is worth comparing this model with different models in future studies because of the high learning speed of this model. In line with this study, but at different scales, some studies have also shown better results for the ANN models as compared with the LR model [33,49,56], however, there could be cases where the LR model has been more reliable than the ANN and SVM models depending on the study conditions [133]. Overall, the ANN and SVM models have often shown better results in comparison to other data mining models, which is consistent with the findings of this research. As both models are similarly and highly reliable, their output maps can be combined to produce the best possible result. Combining the output of landslide susceptibility models is a recommended way to reduce the uncertainty of the final map [136–138], however, the ways the maps can be combined are various and depend on other criteria, and therefore it needs to be addressed in future studies.

All in all, it was observed that nowadays it is possible to employ advanced machine learning methods for small-scale landslide susceptibility mapping with the aid of remote sensing and GIS techniques and their combination. To deal with the problem of the lack of landslide inventory data for very large areas, advanced remote sensing techniques can be applied. The inventory data used in this study was produced as part of a preplanned national project and through visual interpretation of aerial photographs and field investigations which are very time-consuming and costly.

To facilitate the detection of landslides, future studies could utilize modern remote sensing techniques in combination with GIS, for example, the technique of automatic image classification using deep learning methods [139–141]. In addition, with regards to the difficulties of using advanced machine learning methods to create the LSMs of very large areas, intermediate approaches such as NFUC could be adopted to reduce the data volume and make the application of these methods possible.

## 6. Conclusions

The aim of this research was to apply and compare different data mining methods in small-scale landslide susceptibility zoning. It is often impossible to use most data mining models, including advanced machine learning methods, for small-scale analyses because the data at this scale are voluminous and the mentioned methods are computer intensive. Therefore, in this study, an intermediate approach, called NFUC, was designed to reduce the volume of the related data. One of the biggest, most susceptible catchments, in northern Iran, was selected as the study area. The LSMs of the area were produced by employing the data mining methods of FG, LR, C5DT, BPANN, and SVM. To enhance the speed of training the models and make the implementation of the models feasible at this scale, the relatively big data of the 13 selected causative factors were converted to a low-volume format of continuous variables using the NFUC approach. The NFUC approach showed significant capability of reducing the volume of data up to one-fourth, and therefore it can be used as an effective approach for dealing with voluminous data in small-scale landslide susceptibility assessments. Considering the validation rates of the models determined by the ROC method and the percentage of landslides in susceptible areas of their maps, BPANN, followed by SVM, were the most reliable models. However, the C5DT, BLR, and FG models could also be considered reliable enough for a small-scale study. For very large areas (at continental and global scales) where the balance between the reliability and the speed of training is even more important, the C5DT model, as the fastest model, could be more helpful.

To summarize, we conclude that advanced methods, such as ANN or SVM, can reliably be employed with the aid of the NFUC approach to enhance the reliability of LSMs for large areas. Additionally, the best LSMs produced in this study (BPANN and SVM, or preferably a combination of them depending on different criteria) can be engaged in national land use management plans and as a guide for detailed mapping. Finally, to produce more reliable small-scale LSMs, it is suggested that more advanced methods, such as ANN and SVM, can be used with the help of the NFUC approach. In addition, future studies could benefit considerably from more advanced remote sensing and GIS techniques to prepare the required data for very large areas and also implement the machine learning models at these scales for landslide susceptibility assessment.

**Author Contributions:** The first author, V.V., contributed to the article by conceptualization, literature review, investigation, data collection, modeling, validation, visualization, and writing the original draft and the revised versions. H.R.P. and M.Z., supervised the research and helped to write and revise the article. T.B., obtained the research funding and assisted the research team to revise the article.

**Funding:** This study was supported by the Austrian Science Fund FWF through the GIScience Doctoral College (DK W 1237-N23) at the University of Salzburg.

**Acknowledgments:** We would like to thank Open Access Funding by the Austrian Science Fund (FWF) and the anonymous reviewers for their helpful comments.

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

## References

1. Sterlacchini, S.; Ballabio, C.; Blahut, J.; Masetti, M.; Sorichetta, A. Spatial agreement of predicted patterns in landslide susceptibility maps. *Geomorphology* **2011**, *125*, 51–61. [[CrossRef](#)]
2. Froude, M.J.; Petley, D. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazard Earth Sys.* **2018**, *18*, 2161–2181. [[CrossRef](#)]

3. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.-P.; Fotopoulou, S.; Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [[CrossRef](#)]
4. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng. Geol.* **2008**, *102*, 85–98. [[CrossRef](#)]
5. Dahal, R.K.; Hasegawa, S.; Bhandary, N.P.; Poudel, P.P.; Nonomura, A.; Yatabe, R. A replication of landslide hazard mapping at catchment scale. *Geomatics, Nat. Hazards Risk* **2012**, *3*, 161–192. [[CrossRef](#)]
6. Glade, T.; Anderson, M.G.; Crozier, M.J. *Landslide Hazard and Risk*; John Wiley & Sons: Chichester, UK, 2005.
7. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [[CrossRef](#)]
8. Pradhan, A.; Kim, Y. Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *Catena* **2016**, *140*, 125–139. [[CrossRef](#)]
9. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Int. Assoc. Eng. Geol.* **1999**, *58*, 21–44. [[CrossRef](#)]
10. 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. Goetz, J.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
12. Guzzetti, F. Landslide Hazard and Risk Assessment. Ph.D. Thesis, Universitäts- und Landesbibliothek Bonn, Bonn, Germany, 2006.
13. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
14. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
15. Hansen, A. Landslide Hazard Analysis. In *Slope Instability*; Brunsdon, D., Prior, D.B., Eds.; Wiley & Sons: New York, NY, USA, 1984; pp. 523–602.
16. Mondini, A.; Guzzetti, F.; Reichenbach, P.; Rossi, M.; Cardinali, M.; Ardizzone, F. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens. Environ.* **2011**, *115*, 1743–1757. [[CrossRef](#)]
17. Varnes, D.J. Slope movement types and processes. *Spec. Rep.* **1978**, *176*, 11–33.
18. Crozier, M.J. *Landslides: Causes, Consequences & Environment*; Croom Helm Pub: London, UK, 1989.
19. Dietrich, W.E.; Reiss, R.; Hsu, M.L.; Montgomery, D.R. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hydrol. Process.* **1995**, *9*, 383–400. [[CrossRef](#)]
20. Highland, L.; Bobrowsky, P.T. *The Landslide Handbook: A Guide to Understanding Landslides*; US Geological Survey: Reston, VA, USA, 2008.
21. Carrara, A.; Cardinali, M.; Detti, R.; Guzzetti, F.; Pasqui, V.; Reichenbach, P. GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf. Process. Landf.* **1991**, *16*, 427–445. [[CrossRef](#)]
22. Harp, E.L.; Keefer, D.K.; Sato, H.P.; Yagi, H. Landslide inventories: The essential part of seismic landslide hazard analyses. *Eng. Geol.* **2011**, *122*, 9–21. [[CrossRef](#)]
23. Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [[CrossRef](#)]
24. Keefer, D.K.; Larsen, M.C. Assessing landslide hazards. *Sci.* **2007**, *316*, 1136–1138. [[CrossRef](#)]
25. Cascini, L. Applicability of landslide susceptibility and hazard zoning at different scales. *Eng. Geol.* **2008**, *102*, 164–177. [[CrossRef](#)]
26. Fan, W.; Wei, X.S.; Cao, Y.B.; Zheng, B. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. *J. Mt. Sci.* **2017**, *14*, 906–925. [[CrossRef](#)]
27. Myronidis, D.; Papageorgiou, C.; Theophanous, S. Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Nat. Hazards* **2016**, *81*, 245–263. [[CrossRef](#)]
28. Yalcin, A.; 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](#)]

29. Yoshimatsu, H.; Abe, S. A review of landslide hazards in Japan and assessment of their susceptibility using an analytical hierarchic process (AHP) method. *Landslides* **2006**, *3*, 149–158. [[CrossRef](#)]
30. Cui, K.; Lu, D.; Li, W. Comparison of landslide susceptibility mapping based on statistical index, certainty factors, weights of evidence and evidential belief function models. *Geocarto Int.* **2017**, *32*, 935–955. [[CrossRef](#)]
31. 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](#)]
32. Ilia, I.; Tsangaratos, P. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. *Landslides* **2016**, *13*, 379–397. [[CrossRef](#)]
33. Pradhan, B.; Oh, H.J.; Buchroithner, M. Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics, Nat. Hazards Risk* **2010**, *1*, 199–223. [[CrossRef](#)]
34. Chen, W.; Chai, H.; Zhao, Z.; Wang, Q.; Hong, H. Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China. *Environ. Earth Sci.* **2016**, *75*, 474. [[CrossRef](#)]
35. Choi, J.; Oh, H.J.; Lee, H.J.; Lee, C.; Lee, S. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Eng. Geol.* **2012**, *124*, 12–23. [[CrossRef](#)]
36. Lee, S.; Talib, J.A. Probabilistic landslide susceptibility and factor effect analysis. *Environ. Earth Sci.* **2005**, *47*, 982–990. [[CrossRef](#)]
37. Regmi, A.D.; Devkota, K.C.; Yoshida, K.; Pradhan, B.; Pourghasemi, H.R.; Kumamoto, T.; Akgun, A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **2014**, *7*, 725–742. [[CrossRef](#)]
38. Regmi, A.D.; Yoshida, K.; Pourghasemi, H.R.; Dhital, M.R.; Pradhan, B. Landslide susceptibility mapping along Bhalubang—Shiwapur area of mid-Western Nepal using frequency ratio and conditional probability models. *J. Mt. Sci.* **2014**, *11*, 1266–1285. [[CrossRef](#)]
39. Reis, S.; Yalcin, A.; Atasoy, M.; Nisanci, R.; Bayrak, T.; Erduran, M.; Sancar, C.; Ekerin, S. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio and analytical hierarchy methods in Rize province (NE Turkey). *Environ. Earth Sci.* **2012**, *66*, 2063–2073. [[CrossRef](#)]
40. 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](#)]
41. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat. Hazards* **2012**, *63*, 965–996. [[CrossRef](#)]
42. Vakhshoori, V.; Zare, M. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1731–1752. [[CrossRef](#)]
43. 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](#)]
44. Bai, S.B.; Lu, P.; Wang, J. Landslide susceptibility assessment of the Youfang catchment using logistic regression. *J. Mt. Sci.* **2015**, *12*, 816–827. [[CrossRef](#)]
45. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **2016**, *118*, 53–64. [[CrossRef](#)]
46. Du, G.-L.; Zhang, Y.-S.; Iqbal, J.; Yang, Z.-H.; Yao, X. Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *J. Mt. Sci.* **2017**, *14*, 249–268. [[CrossRef](#)]
47. Gorsevski, P.V.; Gessler, P.E.; Foltz, R.B.; Elliot, W.J. Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis. *Trans. GIS* **2006**, *10*, 395–415. [[CrossRef](#)]
48. Mancini, F.; Ceppi, C.; Ritrovato, G. GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 1851–1864. [[CrossRef](#)]
49. Raja, N.B.; Çiçek, I.; Türkoğlu, N.; Aydın, O.; Kawasaki, A.; Aydın, O. Correction to: Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat. Hazards* **2017**, *91*, 1423. [[CrossRef](#)]
50. Chauhan, S.; Sharma, M.; Arora, M.; Gupta, N. Landslide Susceptibility Zonation through ratings derived from Artificial Neural Network. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 340–350. [[CrossRef](#)]

51. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385. [[CrossRef](#)]
52. Gorsevski, P.V.; Brown, M.K.; Panter, K.; Onasch, C.M.; Simic, A.; Snyder, J. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: A case study in the Cuyahoga Valley National Park, Ohio. *Landslides* **2016**, *13*, 467–484. [[CrossRef](#)]
53. 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](#)]
54. 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](#)]
55. Zeng-Wang, X. GIS and ANN model for landslide susceptibility mapping. *J. Geogr. Sci.* **2001**, *11*, 374–381. [[CrossRef](#)]
56. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [[CrossRef](#)]
57. Hong, H.; Pradhan, B.; Bui, D.T.; Xu, C.; Youssef, A.M.; Chen, W. Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: A case study at Suichuan area (China). *Geomat Nat. Hazards Risk* **2017**, *8*, 544–569. [[CrossRef](#)]
58. 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](#)]
59. Wu, X.; Ren, F.; Niu, R. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ. Earth Sci.* **2014**, *71*, 4725–4738. [[CrossRef](#)]
60. Günther, A.; Eeckhaut, M.V.D.; Malet, J.-P.; Reichenbach, P.; Hervás, J. Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology* **2014**, *224*, 69–85. [[CrossRef](#)]
61. Thiery, Y.; Malet, J.-P.; Sterlacchini, S.; Puissant, A.; Maquaire, O. Landslide susceptibility assessment by bivariate methods at large scales: Application to a complex mountainous environment. *Geomorphology* **2007**, *92*, 38–59. [[CrossRef](#)]
62. Trigila, A.; Frattini, P.; Casagli, N.; Catani, F.; Crosta, G.; Esposito, C.; Iadanza, C.; Lagomarsino, D.; Mugnoz, G.S.; Segoni, S.; et al. Landslide Susceptibility Mapping at National Scale: The Italian Case Study. In *Landslide Science and Practice*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 287–295.
63. Van Westen, C.; Van Asch, T.W.; Soeters, R. Landslide hazard and risk zonation—Why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [[CrossRef](#)]
64. Gaprindashvili, G.; Van Westen, C.J. Generation of a national landslide hazard and risk map for the country of Georgia. *Nat. Hazards* **2016**, *80*, 69–101. [[CrossRef](#)]
65. Günther, A.; Reichenbach, P.; Malet, J.-P.; Van Den Eeckhaut, M.; Hervás, J.; Dashwood, C.; Guzzetti, F. Tier-based approaches for landslide susceptibility assessment in Europe. *Landslides* **2013**, *10*, 529–546. [[CrossRef](#)]
66. 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](#)]
67. Van Den Eeckhaut, M.; Hervás, J.; Jaedicke, C.; Malet, J.P.; Montanarella, L.; Nadim, F. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* **2012**, *9*, 357–369. [[CrossRef](#)]
68. Li, Y.; Zhang, Y.; Huang, X.; Yuille, A.L. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 182–196. [[CrossRef](#)]
69. Li, Y.; Zhang, Y.; Huang, X.; Zhu, H.; Ma, J. Large-Scale Remote Sensing Image Retrieval by Deep Hashing Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 950–965. [[CrossRef](#)]

70. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens.* **2019**, *11*, 196. [[CrossRef](#)]
71. Lee, S.; Kim, Y.; Min, K. Development of spatial landslide information system and application of spatial landslide information. *J. Gis Assoc. Korea* **2000**, *8*, 141–153.
72. Lee, S.; Ryu, J.-H.; Kim, I.-S. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: Case study of Youngin, Korea. *Landslides* **2007**, *4*, 327–338. [[CrossRef](#)]
73. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [[CrossRef](#)]
74. Shahabi, H.; Hashim, M.; Bin Ahmad, B. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environ. Earth Sci.* **2015**, *73*, 8647–8668. [[CrossRef](#)]
75. Sahana, M.; Sajjad, H. Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: A case from Rudraprayag district, India. *J. Mt. Sci.* **2017**, *14*, 2150–2167. [[CrossRef](#)]
76. Balteanu, D.; Chendeş, V.; Sima, M.; Enciu, P. A country-wide spatial assessment of landslide susceptibility in Romania. *Geomorphology* **2010**, *124*, 102–112. [[CrossRef](#)]
77. Holec, J.; Bednárík, M.; Šabo, M.; Minár, J.; Yilmaz, I.; Marschalko, M.; Yilmaz, I. A small-scale landslide susceptibility assessment for the territory of Western Carpathians. *Nat. Hazards* **2013**, *69*, 1081–1107. [[CrossRef](#)]
78. Komac, B.; Zorn, M. Statistical landslide susceptibility modeling on a national scale: The example of Slovenia. *Rev. Roum. Géogr.* **2009**, *53*, 179–195.
79. Sabatakakis, N.; Koukis, G.; Vassiliades, E.; Lainas, S. Landslide susceptibility zonation in Greece. *Nat. Hazards* **2013**, *65*, 523–543. [[CrossRef](#)]
80. 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](#)]
81. Bui, D.T.; Pradhan, B.; Löfman, O.; Revhaug, I. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. *Math. Probl. Eng.* **2012**, *2012*, 1–26.
82. Glade, T. Landslide Hazard Assessment and Historical Landslide Data—An Inseparable Couple? In *The Use of Historical Data in Natural Hazard Assessments*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 153–168.
83. Ibsen, M.-L.; Brunsden, D. The nature, use and problems of historical archives for the temporal occurrence of landslides, with specific reference to the south coast of Britain, Ventnor, Isle of Wight. *Geomorphology* **1996**, *15*, 241–258. [[CrossRef](#)]
84. Lang, A.; Moya, J.; Corominas, J.; Schrott, L.; Dikau, R. Classic and new dating methods for assessing the temporal occurrence of mass movements. *Geomorphology* **1999**, *30*, 33–52. [[CrossRef](#)]
85. Zêzere, J.; Pereira, S.; Melo, R.; Oliveira, S.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. [[CrossRef](#)]
86. Wang, X.; Zhang, L.; Wang, S.; Lari, S. Regional landslide susceptibility zoning with considering the aggregation of landslide points and the weights of factors. *Landslides* **2014**, *11*, 399–409. [[CrossRef](#)]
87. Ardizzone, F.; Cardinali, M.; Carrara, A.; Guzzetti, F.; Reichenbach, P. Impact of mapping errors on the reliability of landslide hazard maps. *Nat. Hazard Earth Sys.* **2002**, *2*, 3–14. [[CrossRef](#)]
88. 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](#)]
89. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Xu, C.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [[CrossRef](#)]
90. Chung, C.J.F.; Fabbri, A.G. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat. Hazards* **2003**, *30*, 451–472. [[CrossRef](#)]

91. Gritzner, M.L.; Marcus, W.A.; Aspinall, R.; Custer, S.G. Assessing landslide potential using GIS, soil wetness modeling and topographic attributes, Payette River, Idaho. *Geomorphology* **2001**, *37*, 149–165. [[CrossRef](#)]
92. Moore, I.D.; Wilson, J.P. Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation. *J. Soil Water Conserv.* **1992**, *47*, 423–428.
93. Nefeslioglu, H.; Gokceoglu, C.; Sonmez, H. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng. Geol.* **2008**, *97*, 171–191. [[CrossRef](#)]
94. Krzeminska, D.M.; Steele-Dunne, S.C.; Bogaard, T.A.; Rutten, M.M.; Sailhac, P.; Geraud, Y. High-resolution temperature observations to monitor soil thermal properties as a proxy for soil moisture condition in clay-shale landslide. *Hydrol. Process.* **2012**, *26*, 2143–2156. [[CrossRef](#)]
95. Rianna, G.; Comegna, L.; Mercogliano, P.; Picarelli, L. Potential effects of climate changes on soil–atmosphere interaction and landslide hazard. *Nat. Hazards* **2016**, *84*, 1487–1499. [[CrossRef](#)]
96. Daneshvar, M.R.M. Landslide susceptibility zonation using analytical hierarchy process and GIS for the Bojnurd region, northeast of Iran. *Landslides* **2014**, *11*, 1079–1091. [[CrossRef](#)]
97. Süzen, M.L.; Doyuran, V. Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchment, Turkey. *Eng. Geol.* **2004**, *71*, 303–321. [[CrossRef](#)]
98. Ferentinou, M.; Chalkias, C. Mapping Mass Movement Susceptibility across Greece with GIS, ANN and Statistical Methods. In *Landslide Science and Practice*; Springer: Berlin, Germany, 2013; pp. 321–327.
99. Lima, P.; Steger, S.; Glade, T.; Tilch, N.; Schwarz, L.; Kociu, A. Landslide Susceptibility Mapping at National Scale: A First Attempt for Austria. In *Workshop on World Landslide Forum*; Springer: Cham, Switzerland, 2017; pp. 943–951.
100. 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](#)]
101. Lee, S. Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. *Environ. Geol.* **2007**, *52*, 615–623. [[CrossRef](#)]
102. Pradhan, B.; Lee, S.; Buchroithner, M.F. Use of geospatial data and fuzzy algebraic operators to landslide-hazard mapping. *Appl. Geomat.* **2009**, *1*, 3–15. [[CrossRef](#)]
103. Caniani, D.; Pascale, S.; Sdao, F.; Sole, A. Neural networks and landslide susceptibility: A case study of the urban area of Potenza. *Nat. Hazards* **2008**, *45*, 55–72. [[CrossRef](#)]
104. Feizizadeh, B.; Roodposhti, M.S.; Blaschke, T.; Aryal, J. Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. *Arab. J. Geosci.* **2017**, *10*, 122. [[CrossRef](#)]
105. Zadeh, L.A. Fuzzy sets. *Inf. Cont.* **1965**, *8*, 338–353. [[CrossRef](#)]
106. Ercanoglu, M.; Temiz, F.A. Application of logistic regression and fuzzy operators to landslide susceptibility assessment in Azdavay (Kastamonu, Turkey). *Environ. Earth Sci.* **2011**, *64*, 949–964. [[CrossRef](#)]
107. Sema, H.V.; Guru, B.; Veerappan, R. Fuzzy gamma operator model for preparing landslide susceptibility zonation mapping in parts of Kohima Town, Nagaland, India. *Model. Earth Sys. Environ.* **2017**, *3*, 499–514. [[CrossRef](#)]
108. Tangestani, M.H. A comparative study of Dempster–Shafer and fuzzy models for landslide susceptibility mapping using a GIS: An experience from Zagros Mountains, SW Iran. *J. Asian Earth Sci.* **2009**, *35*, 66–73. [[CrossRef](#)]
109. 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](#)]
110. Chau, K.T.; Chan, J.E. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. *Landslides* **2005**, *2*, 280–290. [[CrossRef](#)]
111. Yang, I.T.; Chun, K.S.; Park, J.H. The effect of landslide factor and determination of landslide vulnerable area using GIS and AHP. *J. Korean Soc. Geos. Inf. Sys.* **2006**, *14*, 3–12.
112. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [[CrossRef](#)]
113. Zhu, L.; Huang, J.F. GIS-based logistic regression method for landslide susceptibility mapping in regional scale. *J. Zhejiang Univ. A* **2006**, *7*, 2007–2017. [[CrossRef](#)]
114. Lee, S.; Sambath, T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ. Earth Sci.* **2006**, *50*, 847–855. [[CrossRef](#)]

115. Cox, D.R.; Snell, E.J. *Analysis of Binary Data*, 2nd ed.; Chapman and Hall: London, UK, 1989.
116. Nagelkerke, N.J.D. A note on a general definition of the coefficient of determination. *Biometrika* **1991**, *78*, 691–692. [[CrossRef](#)]
117. Fuchu, D.; Lee, C.; Sijing, W. Analysis of rainstorm-induced slide-debris flows on natural terrain of Lantau Island, Hong Kong. *Eng. Geol.* **1999**, *51*, 279–290. [[CrossRef](#)]
118. Kawabata, D.; Bandibas, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology* **2009**, *113*, 97–109. [[CrossRef](#)]
119. Palani, S.; Liong, S.Y.; Tkalich, P. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **2008**, *56*, 1586–1597. [[CrossRef](#)]
120. 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](#)]
121. Ermini, L.; Catani, F.; Casagli, N. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* **2005**, *66*, 327–343. [[CrossRef](#)]
122. Paola, J.D.; Schowengerdt, R.A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **1995**, *16*, 3033–3058. [[CrossRef](#)]
123. Negnevitsky, M. *Artificial Intelligence: A Guide to Intelligent Systems*; Pearson education: London, UK, 2005.
124. Kanungo, D.P.; Arora, M.; Sarkar, S.; Gupta, R. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng. Geol.* **2006**, *85*, 347–366. [[CrossRef](#)]
125. Garrett, J. Where and why artificial neural networks are applicable in civil engineering. *J. Comput. Civil Eng.* **1994**, *8*, 129–130. [[CrossRef](#)]
126. Congalton, R.G. Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues. *Photogramm. Eng. Rem. Sens.* **1991**, *57*, 677–687.
127. Swingler, K. *Applying Neural Networks: A Practical Guide*; Academic Press: New York, NY, USA, 1996.
128. Gong, P. Integrated Analysis of Spatial Data from Multiple Sources: An Overview. *Can. J. Remote Sens.* **1994**, *20*, 349–359. [[CrossRef](#)]
129. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; 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](#)]
130. IBM. Knowledge Center. Available online: [https://www.ibm.com/support/knowledgecenter/en/SS3RA7\\_17.1.0/modeler\\_mainhelp\\_client\\_ddita/clementine/svm\\_node\\_experttab.html](https://www.ibm.com/support/knowledgecenter/en/SS3RA7_17.1.0/modeler_mainhelp_client_ddita/clementine/svm_node_experttab.html) (accessed on 23 August 2019).
131. Schlögel, R.; Braun, A.; Torgoev, A.; Fernández-Steeger, T.M.; Havenith, H.-B. Assessment of Landslides Activity in Maily-Say Valley, Kyrgyz Tien Shan. In *Landslide Science and Practice*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 111–117.
132. Lim, T.S.; Loh, W.Y.; Shih, Y.S. A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. *Mach. Learn.* **2000**, *40*, 203–228. [[CrossRef](#)]
133. Cho, J.H.; Kurup, P.U. Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors Actuators B Chem.* **2011**, *160*, 542–548. [[CrossRef](#)]
134. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013; Volume 26.
135. Aquino, L.D.; Tullis, J.A.; Stephen, F.M. Modeling red oak borer, *Enaphalodes rufulus* (Haldeman), damage using in situ and ancillary landscape data. *For. Ecol. Manag.* **2008**, *255*, 931–939. [[CrossRef](#)]
136. Fakhr, M.; Elsayad, A.M. Steel Plates Faults Diagnosis with Data Mining Models. *J. Comput. Sci.* **2012**, *8*, 506–514.
137. Nisbet, R.; Elder, J.; Miner, G. *Handbook of Statistical Analysis and Data Mining Applications*; Academic Press: London, UK, 2009.
138. Wang, X.; Niu, R. Landslide intelligent prediction using object-oriented method. *Soil Dyn. Earthq. Eng.* **2010**, *30*, 1478–1486. [[CrossRef](#)]
139. Begueria, S. Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management. *Nat. Hazards* **2006**, *37*, 315–329. [[CrossRef](#)]
140. Fielding, A.H.; Bell, J.F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **1997**, *24*, 38–49. [[CrossRef](#)]
141. Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]

142. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [[CrossRef](#)]
143. Neuhäuser, B.; Damm, B.; Terhorst, B. GIS-based assessment of landslide susceptibility on the base of the weights-of-evidence model. *Landslides* **2012**, *9*, 511–528. [[CrossRef](#)]
144. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192. [[CrossRef](#)]
145. Chung, C.-J.F.; Fabbri, A.G. Probabilistic prediction models for landslide hazard mapping. *Photogramm. Eng. Rem.* **1999**, *65*, 1389–1399.
146. Fabbri, A.G.; Chung, C.J. On Blind Tests and Spatial Prediction Models. *Nat. Resour. Res.* **2008**, *17*, 107–118. [[CrossRef](#)]
147. Dai, F.; Lee, C.; Ngai, Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [[CrossRef](#)]
148. Akgun, A.; Dag, S.; Bulut, F. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ. Geol.* **2008**, *54*, 1127–1143. [[CrossRef](#)]
149. Cannon, S. Debris-Flow Response of Southern California Watersheds Burned by Wildfire. In Proceedings of the Second International Conference on Debris-Flow Hazards Mitigation, Taipei, Taiwan, 16 August 2000; A.A. Balkema: Brookfield, WI, USA, 2000.
150. Glade, T. Landslide occurrence as a response to land use change: A review of evidence from New Zealand. *Catena* **2003**, *51*, 297–314. [[CrossRef](#)]
151. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; Unesco: Paris, France, 1984.
152. Hasanat, M.H.A.; Ramachandram, D.; Mandava, R. Bayesian belief network learning algorithms for modeling contextual relationships in natural imagery: A comparative study. *Artif. Intell. Rev.* **2010**, *34*, 291–308. [[CrossRef](#)]



© 2019 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/>).