Article

Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran

Alireza Arabameri [1],*[ID]**, Sunil Saha** [2][ID]**, Jagabandhu Roy** [2]**, Wei Chen** [3,4,5][ID]**, Thomas Blaschke** [6][ID] **and Dieu Tien Bui** [7],*[ID]

1 Department of Geomorphology, Tarbiat Modares University, Tehran 14117-13116, Iran
2 Department of Geography, University of Gour Banga, Malda 732101, West Bengal, India;
   sunilgeo.88@gmail.com (S.S.); jagabandhuroy1991@gmail.com (J.R.)
3 College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, China;
   chenwei0930@xust.edu.cn
4 Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Land and
   Resources, Xi'an 710021, China
5 Shaanxi Provincial Key Laboratory of Geological Support for Coal Green Exploitation, Xi'an 710054, China
6 Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria;
   thomas.blaschke@sbg.ac.at
7 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
* Correspondence: a.arabameri@modares.ac.ir (A.A.); buitiendieu@duytan.edu.vn (D.T.B.)

check for updates

Abstract: This analysis aims to generate landslide susceptibility maps (LSMs) using various machine learning methods, namely random forest (RF), alternative decision tree (ADTree) and Fisher's Linear Discriminant Function (FLDA). The results of the FLDA, RF and ADTree models were compared with regard to their applicability for creating an LSM of the Gallicash river watershed in the northern part of Iran close to the Caspian Sea. A landslide inventory map was created using GPS points obtained in a field analysis, high-resolution satellite images, topographic maps and historical records. A total of 249 landslide sites have been identified to date and were used in this study to model and validate the LSMs of the study region. Of the 249 landslide locations, 70% were used as training data and 30% for the validation of the resulting LSMs. Sixteen factors related to topographical, hydrological, soil type, geological and environmental conditions were used and a multi-collinearity test of the landslide conditioning factors (LCFs) was performed. Using the natural break method (NBM) in a geographic information system (GIS), the LSMs generated by the RF, FLDA, and ADTree models were categorized into five classes, namely very low, low, medium, high and very high landslide susceptibility (LS) zones. The very high susceptibility zones cover 15.37% (ADTree), 16.10% (FLDA) and 11.36% (RF) of the total catchment area. The results of the different models (FLDA, RF, and ADTree) were explained and compared using the area under receiver operating characteristics (AUROC) curve, seed cell area index (SCAI), efficiency and true skill statistic (TSS). The accuracy of models was calculated considering both the training and validation data. The results revealed that the AUROC success rates are 0.89 (ADTree), 0.92 (FLDA) and 0.97 (RF) and predication rates are 0.82 (ADTree), 0.79 (FLDA) and 0.98 (RF), which justifies the approach and indicates a reasonably good landslide prediction. The results of the SCAI, efficiency and TSS methods showed that all models have an excellent modeling capability. In a comparison of the models, the RF model outperforms the boosted regression tree (BRT) and ADTree models. The results of the landslide susceptibility modeling could be useful for land-use planning and decision-makers, for managing and controlling the current and future landslides, as well as for the protection of society and the ecosystem.

Keywords: landslide susceptibility (LSM); alternative decision tree (ADTree); Fisher's linear discriminant function (FLDA); machine learning approach; R program

1. Introduction

Landslides are among the most destructive natural hazards in the watershed of the Gorganround River, destroying human life and property [1]. A landslide can be described as earthen materials falling along a slope due to the force of gravity [2]. Landslides are a predominantly geological phenomenon that occurs when the material's force exceeds the resistance of the soil's shear force [3]. Of all the different natural hazards occurring worldwide, landslides are in 7th place in terms of destruction of human life and property [4]. In Iran, the mountainous regions are at the greatest risk of landslides. Physiogeographically, one-third of Iran is mountainous, while deserts and semi-deserts cover over half, and only a small area is fertile land, namely the southern plain of the Caspian Sea and the plain of Khuzestan [5]. The main causes of landslides in Iran are geomorphological circumstances, increased population and its burden on natural resources and a lack of scientific environmental management [6]. Particularly in the northern province of Iran, landslides are the predominant type of natural disasters, affecting the roads, power transmission, railway lines, mining and mineral facilities, irrigation and water supply, oil and gas refineries, industrial centers, artificial lakes, forests and pastures, farms and residential areas [7]. Landslides cause 500 billion Rial of damage in Iran each year, according to the Iranian Ministry's National Committee on Natural Disaster Reduction. This damage to the economy, however, not only results from the direct and indirect damage to the non-renewable resources, but also from the depletion of soil, which is the most important natural resource. Soil loss can increase the volume of sediment that affects the ecosystem, and therefore, the economic loss resulting from landslides in Iran is more than the stated amount [7].

Landslide susceptibility mapping can be used for efficient planning and management of natural resources [8] and can be a useful tool for a region's sustainable development. Landslide susceptibility mapping has been implemented in recent decades as the subject of research around the world [9–11]. The factors contributing to the occurrence of landslides include lithology, climatic, morphometric, and human factors, whereby road construction, settlements and land-use changes are the main anthropogenic factors for landslide [12]. Understanding the LCFs, which control the occurrence of landslides, and having access to accurate LSMs may provide useful information to various planners and decision-makers, aiding them in taking important steps in soil and natural resource management and protection, and to determine the areas of land appropriate for the expansion of cities and villages [13]. Landslide susceptibility maps, based on appropriate models, can be used to determine the likely locations of landslides and can thus be used to make the best possible use of the land while avoiding landslide-prone areas. The conservation and management of land is the main motivation behind the preparation of LSMs [14]. Both quantitative and qualitative approaches were used to create the LSMs [15]. Qualitative methods have a key role in preparation of landslide inventory mapping [16]. One of the essential and effective data sources used for the LSAM is a landslide inventory map (LIM). A LIM may be obtained from the interpretation of the aerial photograph, field investigation, and Google Earth images. The LIM indicates the spatial distribution of the landslides in the form of points and polygons [17]. A LIM is a basic precondition for landslide susceptibility mapping [18,19]. Researchers around the world have applied different kinds of methods and techniques, including knowledge-driven, probability-based, machine learning, data mining, and ensemble methods for landslide susceptible mapping. Examples of applications of these methods and approaches have been found in the literature, including analytic hierarchy process (AHP), frequency ratio (FR), statistical index (SI), evidential belief function (EBF), landslide nominal risk factor (LNRF), Dempster-Shafer (DS), logistic regression (LR), weights of evidence (WOE), fuzzy logic (FL), support vector machine (SVM), artificial neural network (ANN) boosted regression tree (BRT), least squares SVM (LSSVM), multivariate adaptive regression spline (MARS), adaptive neuro-fuzz inference system (ANFIS), random forest (RF) decision tree (DT), Fisher's Linear Discriminant Function (FLDA), J48, k-nearest neighbor (KNN), Bayesian logistic regression (BLR), logistic model tree (LMT) and alternate decision tree (ADTree) [20–39].

In the last three decades, machine learning ensemble approaches have been applied in different fields due to their superior performance capabilities and the ability to deal with complex and varying

data [40]. The geographic information system (GIS), combined with R programming, is a strong and effective tool for hazard mapping [41–44] that has provided appropriate and meaningful results in landslide susceptibility mapping [45]. The benefits and advantages of machine learning ensemble models are not only that these techniques provide transparent computations but they also lead to more accurate models [20]. Thus, the discovery of new landslide modeling technologies, processes, and models are essential [20,34–36].

Landslides are creating huge problems in the Gallicash drainage basin in Golestan Province, Iran [46]. We evaluated the literature and found that the area's vulnerability to landslides and resulting losses are due to a lack of appropriate guidelines in the construction of residential areas, roads and other facilities in potentially dangerous areas. It is necessary to recognize and map landslide-prone areas to achieve efficient and safe operations in the area. Informing decision-makers about potentially vulnerable areas would help them to avoid developing unsuitable structures in the area, avoiding the area altogether, or employing effective technological measures to overcome the potential threat. Therefore, given the importance of the issue, we used the ADTree, FLDA, and RF machine learning-based models in the Gallicash drainage basin in Golestan Province, Iran to obtain a landslide susceptibility map with high prediction accuracy. Models using the ADTree and FLDA algorithms are effective yet seldom used for spatial landslide modeling [47–49]. The ADTree integrates decision rules in the classification by boosting and decision tree algorithms, and the ADTree can, therefore, be used to construct a simpler layout, and its description of classification rules is also easier to understand and visualize [49]. FLDA is a popular statistical learning approach that produces promising results for classification issues [49]. The data is projected onto a line in the FLDA algorithm, and the classification takes place in one-dimensional space. Furthermore, the projection maximizes the difference between the means of two classes in the data, while the variance within each class is minimized [49]. In this analysis, we contrast the prediction accuracy of these two models with RF, which is a powerful model in natural hazard modeling [25]. These models were used to identify the landslide-prone areas that will support development planners and decision-makers without hampering nature to maintain environmental management.

2. Materials and Methods

2.1. Study Area

The Gallicash drainage basin, situated in northern Iran, is part of the Gorganroud basin, which encompasses an area of roughly 5368 km^2 between latitude 37°07′ and 37°43′N and longitude 54°58′ and 55°56′E and drains into the Caspian Sea (Figure 1). The elevation of the basin ranges from 13 m to 2870 m above sea level. While Iran is generally referred to as arid and semi-arid, in the eastern part of the Gallicash drainage basin the climate is semi-arid, and in the western part, it is wet. The mean basin temperature varies between 11 °C and 18.1 °C and the total annual precipitation of the watershed ranges from 195 to 946 mm [50]. About 36% of the total precipitation occurs between January and March. Generally, the topography is characterized by a complex mixture of mountains (46.1%), hills (9.6%), plateaus and higher terraces (4.6%), plains of Piedmont (15.5%), alluvial plains (16.3%) and lowlands (7.7%) [51]. The area is covered by numerous sedimentary rocks, including calcareous, sandstone, limestone, dolomite, and marl, along with conglomerates, loess sediments, and alluvium (Geological Department of Iran). The watershed comprises four soil classes, namely Entisols (25.6%), Alfisols (25.1%), Inceptisols (19.7%) and Mollisols (29.3%). The predominant land-use/landcover types in the basin are agriculture (10.44 %) with wheat, barley, sunflower and watermelon as the main crops, orchard (0.02%), afforest (25.88%), dry farming (47.1%), water (0.69%), agriculture-orchard (14.7%), rock (0.08%) and settlement (0.98%). The Gallicash basin is the population center of Golestan Province and home to around 1.2 million residents. At the river confluence, the Voshmgir dam serves the public water supply, flood control, hydroelectric power generation and irrigation. The basin maintains an agriculture-based economy along with manufacturing and mining. The rivers of this basin originate in

the southern and northeastern highlands and, from west to east, include Qazanabad, Taghi Abad and Mohammad Abad. The most significant structural elements of the region are the faults of the Caspian Sea and northern Alborz mountains. some faults extend from northeast to southwest and some from north to south in the eastern part [52]. Population growth in erodible soils has contributed to shifts in land-use in recent years, which has increased runoff and surface erosion [53]. Due to population growth and consequently land use change, the Gallicash River suffers from increased soil erosion, flash flooding, landslides, and a high yield of sediment [46].
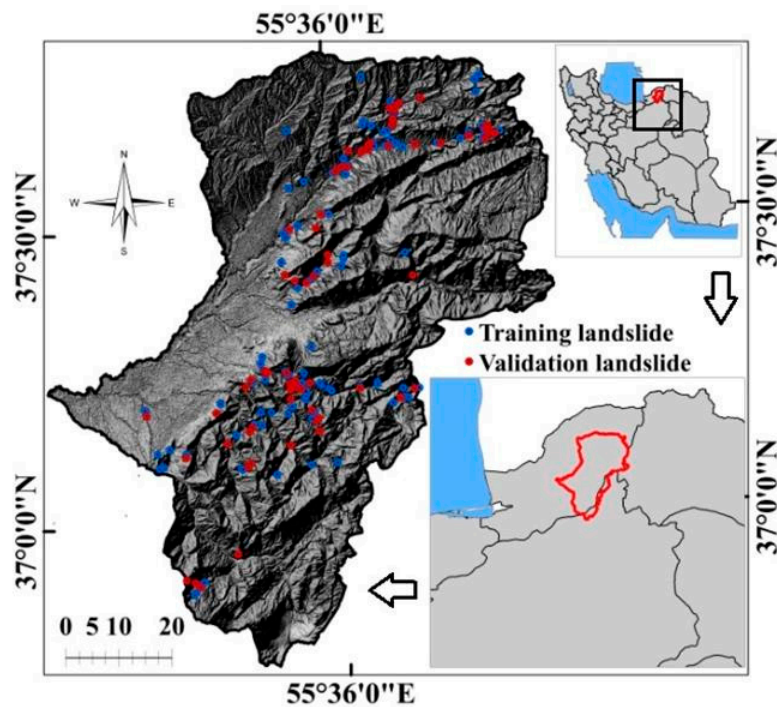

Figure 1. Location of the study area.

2.2. Methodology

The current research demonstrates the application of different machine learning ensemble techniques in Landslide susceptibility mapping. Preparing the Landslide susceptibility map in the present study involved data collection, inventory map preparation, LCFs preparation, collinearity test, factors importance analysis using the RF and BRT models, LS mapping using RF, ADtree, and FLDA and validation of the models (Figure 2). Both primary and secondary data were used in this work. The LIM was prepared based on the field investigation in a different time (2018/05/12; 2018/08/10; and 2018/10/14) and Google Earth imagery. The topographic parameters of altitude, slope aspect convergence index (CI), stream power index (SPI) and topographical positioning index (TPI) were obtained from the 12.5 m × 12.5 m resolution ALOS PALSAR DEM that was downloaded from the Alaska Satellite Facility homepage. The enhanced thematic mapper plus (ETM+) satellite imagery of 15 m × 15 m resolution was obtained from USGS. The topographic map (1:50,000) and geological map (1:1,000,000) were obtained from the Cartographic and Geological Department, Iran. Rainfall data of different rain gauge stations was also collected from the Metrological Department of Iran. The LSMs were classified into five classes using the natural break method of Jenks. The Jenks method of optimization, also called the natural breaks classification method, is a method of data segmentation designed to determine the best value arrangement in different classes. This method of classification aims to decrease the average deviation from the mean class value while increasing the deviation from the mean of other classes. The method reduces the intra-class variance and maximizes the inter-class variance [54,55]. The natural breaks classification method is an iterative process. That is, it repeats calculations using different breaks in the dataset to decide which set of the breaks have the smallest

variance in the class. There are three steps that need to be followed for this method: (1) calculate the sum of squared deviations from the class mean (SDCM); (2) calculate the sum of squared deviations from the array mean (SDAM); (3) after each SDCM is checked, decided to move one unit from a class with a larger SDCM to an adjacent class with a lower SDCM. New class deviations are then measured, and the process continues until a minimum value is reached by the sum of the deviations within the class [54]. The whole process was carried out using R studio, ArcGIS, ENVI and Weka.
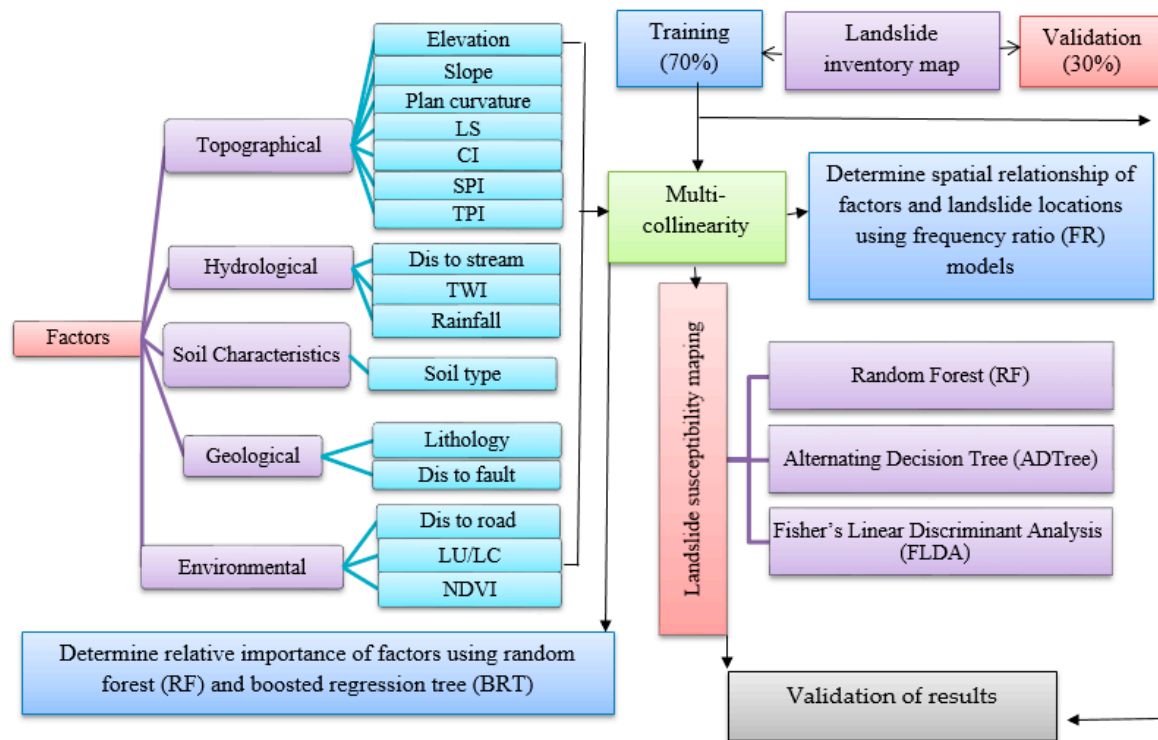

Figure 2. Flowchart of research in the study area.

2.3. LIM

The LIM can be defined as the total set of landslide locations as it includes both past and present landslide points. The LIM is an essential part of the landslide susceptibility mapping. The relationship between the landslide inventory datasets and landslide conditioning factors (LCFs) is the most important for landslide susceptibility mapping [56]. The LIM contains both the old and new landslides that have occurred in the study area. A LIM can be prepared based on a combination of several data sources, including previous records, local field examinations, the perception survey with inhabitants and interpretation of satellite images [57]. Some researchers have used geomorphologic features visible in satellite imagery to detect landslides [58].

In this study, a LIM was prepared using historical records [59,60], aerial photos, field inquiry with GPS, Google Earth imagery and the PALSAR DEM. The PALSAR DEM is more accurate for detecting landslides in forested areas [61]. The resolution of the PALSAR DEM is 12.5 m × 12.5 m. Many landslides, which due to their remote locations aren't even detectable through field surveys, were extracted with acceptable resolution using Google Earth's optical images. A total of 249 landslides were randomly selected in the basin. Of the 249 landslides, 70% were randomly selected to be used for training, and 30% were used for validation. The research area's landslide identification results show that landslides in the study area are mainly translational earth flow slides, rotational slides and debris flows (Figure 3a). Of the 249 identified landslides 75 (30%) are creeps, 47 (19%) are rotational slides, 27 (11%) are earth flows, 87 (35%) are translational slides, and 13 (5%) are debris flows. The combined

area covered by all the landslides in the study area is 16,213,651 m^2. Of the identified landslides, the smallest was 191.76 m^2 and the largest was 1,254,925 m^2, and the mean size was 562,782 m^2.

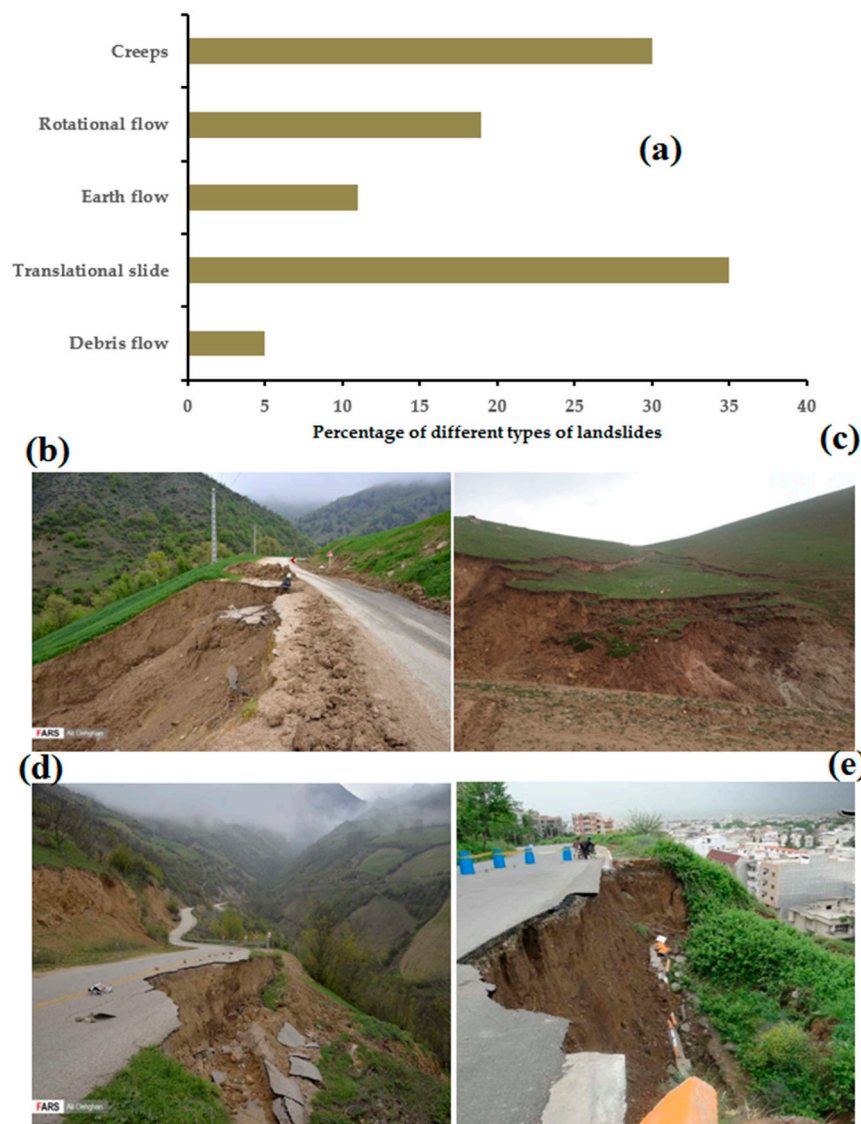**Figure 3.** (**a**) Distribution of different landslide types; (**b**–**e**) examples of landslides in the study area (Farsnews, 2019 and field survey).

Very little research is available on analyzing areas where the chance of landslides occurring is virtually zero, which is the definition of non-landslide areas [62]. Marchesini et al. [63] recommended a non-linear quantile model for identifying non-susceptible landslide areas based on a morphometric evaluation. However, non-landslide areas were delimited in the current study based on the availability of good inventory maps. It was necessary to generate non-landslide evidence for modeling as the landslide susceptibility analysis is based on a binary classification [64]. In the present research, we also used non-landslide data for generating the training and validation datasets by identifying non-landslide areas on Google Earth images where no building and slope failures were found. For both training and validation datasets, the number of non-landslide data equals the number of landslide data. Landslide data was assigned a value of "1," while non-landslide data was assigned a value of "0" for spatial landslide susceptibility modeling. Figure 1 shows the landslide locations of the training and validation datasets. The percentages of different landslide types along with some field photographs, are shown in Figure 3a–e.

2.4. Preparing Landslide Conditioning Factors (LCFs)

A landslide susceptibility map cannot be accurate without understanding the effective LCFs. The topographical factors, i.e., elevation, slope, slope length (LS), plan curvature, convergence index (CI), stream power index (SPI), topographic position index (TPI), geomorphologic, i.e., distance to stream, topographic wetness index (TWI), geological, i.e., lithology, distance to fault, and environmental, i.e., land-use/land cover (LU/LC), soil type, normalized differential vegetation index (NDVI) were considered based on existing literature [19–25] and multi-collinearity analysis for the prediction of the landslide risk zones in the Gallicash river watershed. Determining the LCFs is the first step in mapping the LS. The Phased Array type L-band synthetic aperture radar (PALSAR) DEM is more reliable and accurate than the ASTER and SRTM DEM. So, in this study, we used the PALSAR DEM with 12.5 m × 12.5 m resolution. The topographical parameters of elevation (Figure 4a), slope (Figure 4b), LS (Figure 4g), plan curvature (Figure 4c), convergence index (Figure 4h), stream power index (Figure 4d), TWI (Figure 4f), TPI (Figure 4e), and stream network (Figure 4k) were extracted from the PALSAR DEM in GIS using suitable techniques, which are presented in Table 1. The geological map of the study area was prepared from the 1:100,000 geological map of Iran, (Figure 4m). The eight geological units in this watershed were identified (Table 1). Faults were extracted from the ETM+2002 satellite imagery using ENVI 4.7 software. The distance to the fault map was prepared using the Euclidian distance tool in ArcGIS (Figure 4i). The Sentinal-2 images with 30 m × 30 m resolution were used to construct the LULC (Figure 4n) and NDVI maps (Figure 4o) using the supervised classification (maximum likelihood). For the assessment of the classification accuracy of the LULC map, 523 ground control points were used. The overall accuracy of the LULC map is 97.1% (Kappa coefficient = 0.971), which is acceptable. Different types of land-use, namely Afforest, Agriculture, Orchard, Dry farming, Water, Agriculture-Orchard, Rock and Urban area (Figure 4n) have been identified in this watershed. The soil map was prepared from a projected soil map (scale 1: 50,000), which was obtained from the Agriculture and Natural Resources Center using the digitized process in the GIS environment (Figure 4p). The study area is composed of different soil types, namely Rock Outcrops/Entisols, Inceptisols, Mollisols, and Alfisols (Figure 4p and Table 1). The elevation, slope, LS, convergence index (CI), SPI, TPI, TWI were classified into five sub-layers with the aid of the natural break method (NBM) of Jenk's in GIS, as shown in Figure 4 and Table 1. The NDVI was categorized into the three sub-layers of <−0.201, −0.201 −0.36 and >0.36 using the NBM (Figure 4o and Table 1). The soil type, lithology and LU/LC are the categorical variables, with detailed description given in Figure 4 and Table 1.
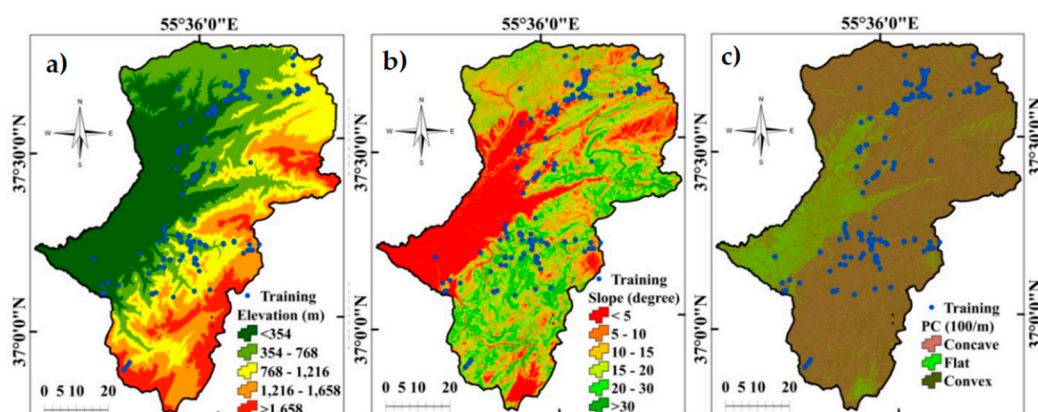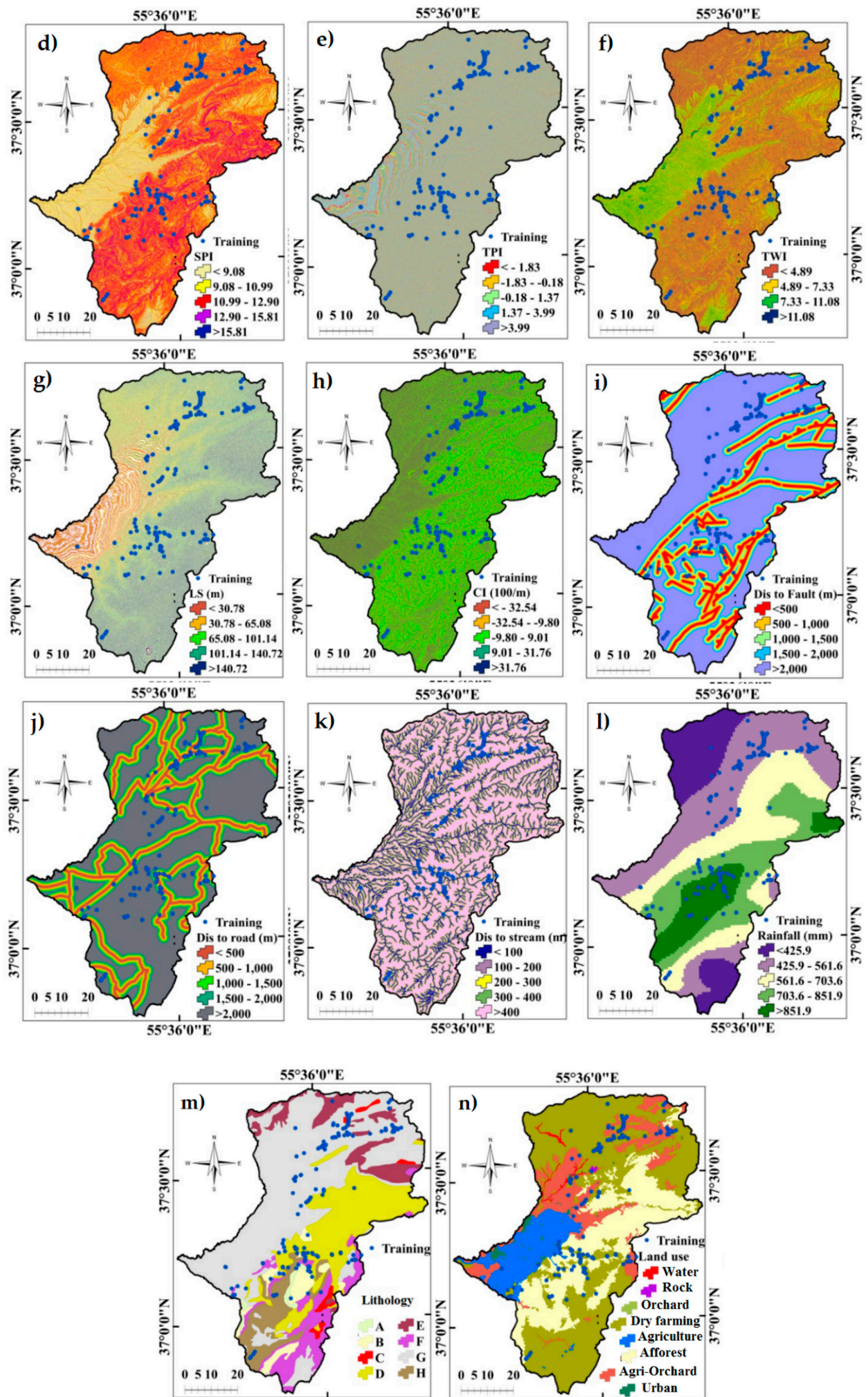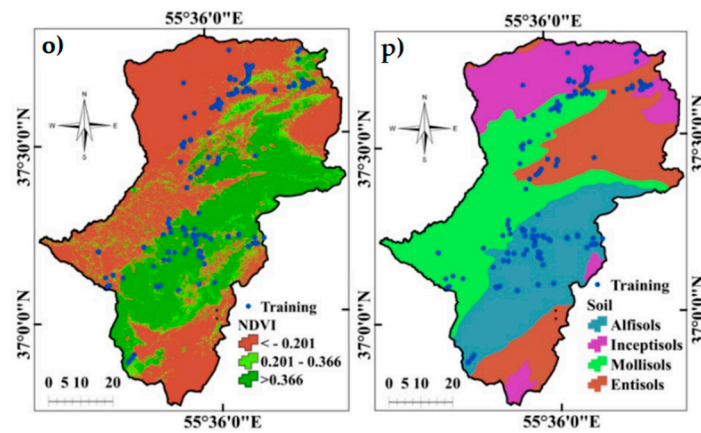

Figure 4. *Cont.*

Figure 4. *Cont.*

Figure 4. Landslide conditioning factors—(**a**) Elevation, (**b**) Slope, (**c**) Plan curvature (PC), (**d**) Stream power index (SPI), (**e**) Topographic position index (TPI), (**f**) Topographic wetness index (TWI), (**g**) Slope length (LS), (**h**) Convergence index (CI), (**i**) Distance to fault, (**j**) Distance to road, (k) Distance to stream, (**l**) Rainfall, (**m**) Lithology, (**n**) LULC, (**o**) NDVI, (**p**) Soil type.

Table 1. Suitable techniques, data sources, and classification methods of landslide conditioning factors.

Factors	Class	Data Used/Resolution/Scales	Data Sources	Techniques	Classification Methods	Ref.
Elevation (m)	1. <354, 2. 354–768, 3. 768–1216, 4. 1216–658, 5. >1658	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	12.5 m × 12.5 m DEM	Natural break (Jenks)	[65]
Slope (^{0})	1. <5, 2. 5–10, 3. 10–15, 4. 20–30, 5. >30	ALOS DEM (12.5 m × 15.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$Tan\theta = \frac{N \times i}{636.6}$ where N = No. of Contour Cuttings; I = Contour Interval (12.5 m × 15.5 m resolution)	Natural break (Jenks)	[66]
PC	1. Concave, 2. Flat, 3. Convex	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	12.5 m × 12.5 m DEM	Natural break (Jenks)	[67]
SPI	1. <9.08, 2. 9.08–10.99, 3. 10.99–12.9, 4. 12.9–15.8, 5. >15.8	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$SPI = As \times \tan \beta$ where A_S is the upstream contributing area and β is the slope gradient (in degrees)	Natural break (Jenks)	[68]
TPI	1. <−1.83, 2. −1.83−−0.18, 3. −0.18–1.37, 4. 1.37–3.99, 5. >3.99	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$TPI = Z_0 - \overline{Z}$ $\overline{Z} = \frac{1}{n_R}\sum_{i \in R} Z_i$ Center point elevation (Z_0) and the average elevation ($\overline{Z}$) around it within a predetermined radius (R)	Natural break (Jenks)	[69]
TWI	1. <4.89, 2. 4.89–7.33, 3. 7.33–11.08, 4. >11.08	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$TWI = In\left(\frac{A_s}{\tan \theta}\right)$ where α not in formula is a total upsloped area that drains through a point (per unit contour length), β not in formula is a gradient of the slope (in degree).	Natural break (Jenks)	[68]
LS (m)	1. <30.7, 2. 30.7–65.08, 3. 65.08–101.1, 4. 101.1–140.7, 5. >140.7	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$LS = \left(\frac{A_s}{22.13}\right)^{0.6} \times \left(\frac{sin\beta}{0.0896}\right)^{1.3}$ where AS is the particular area of the basin (m2 m−1) and β is the slope in degrees.	Natural break (Jenks)	[70]
CI (100/M)	1. <−32.54, 2. −32.54−−9.8, 3. −9.8–9.01, 4. 9.01–31.76, 5. >31.76	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	$CI = \left(\frac{1}{8}\sum_{i=1}^{8} \theta_i\right) - 90°$ where θ indicates the average angle between the aspect of adjacent cells and the direction to the central cell.	Natural break (Jenks)	[71]

Table 1. *Cont.*

Factors	Class	Data Used/Resolution/Scales	Data Sources	Techniques	Classification Methods	Ref.
Dis to fault (m)	1. <500, 2. 500–1000, 3. 1000–1500, 4. 1500–2000, 5. >2000	ETM+2002 satellite data (15 m × 15 m)	Soil Conservation Section of the Agricultural and Natural Resources Research Center	Euclidian Distance Buffering	Natural break (Jenks)	[72]
Dis to road (m)	1. <500, 2. 500–1000, 3. 1000–1500, 4. 1500–2000, 5. >2000	Topographical map (scale 1: 25,000) and Google Earth image	Cartography department of Iran	Digitization process	Natural break (Jenks)	[73]
Dis to stream (m)	1. <100, 2. 100–200, 3. 200–300, 4. 300–400, 5. >400	ALOS DEM (12.5 m × 12.5 m resolution)	https//vertex.daac.asf.alaska.edu (Alaska Satellite Facility)	Euclidian Distance Buffering	Natural break (Jenks)	[72]
Rainfall (mm)	1. <425.9, 2. 425.9–561.6, 3. 561.6–703.6, 4. 703.6–851.9, 5. >851.9	30 years average rainfall data of different stations	Islamic Republic of Iran metrological Organization	Kriging Interpolation method	Natural break (Jenks)	[74]
Lithology	1. A, 2. B, 3. C, 4. D, 5. E, 6. F, 7.G, 8. H	Projected geological map (scale 1:100,000)	Geological survey of Iran	Digitization process	Equal interval	[73]
LU/LC	1. Afforest, 2. Agriculture, 3. Orchard, 4. Dry farming, 5. Water, 6. Agri-Orch, 7. Rock, 8. Urban	Landsat 8 OLI/TIRS (Path 162/Row 34) (30 m × 30 m resolution)	U.S Geological Survey	Supervised classification (Maximum likelihood)	Maximum likelihood	[75]
NDVI	1. <−0.201, 2. −0.201–0.36, 3. >0.36	Landsat 8 OLI/TIRS (Path 162/Row 34) (30 m × 30 m resolution)	U.S Geological Survey	$NDVI = \frac{NIR-IR}{NIR+IR}$ where NIR is near inferred band or band 4 and IR is the infrared band or band 3.	Natural break (Jenks)	[76]
Soil type	1. Rock Outcrops/Entisols, 2. Inceptisols, 3. Mollisols, 4. Alfisols	Projected soil map (Scale 1:25,000)	Agricultural and Natural Resources Research Center in Iran	Digitization process	Equal interval	[73]

In the lithology map, A = Cm unit, B = Dp, and DCkh units, C = E1c, and E1m units, D = Jsc, Jd, Jl, Jmz, and Jch units, E = Kat, Ksn, Ksr, Ku, and Kad-ab units, F = Pz, Pz1av, pC-C, Pr, Pz1a.bv, Pd, Plc, and P units, G = Qsw, Qft2, Qm, Qft1, Qs,d, and Qal, H = Tre, and TRJs units.

2.5. Testing Multi-Collinearity Problems

The multi-collinearity test may enhance the results of the models by helping in the selection of ideal factors for hazard mapping [77]. Multi-collinearity is the linear dependency that implies that there are two or more related variables in a dataset. The multi-collinearity test has been used for several purposes, such as landslide susceptibility mapping, soil and gully erosion susceptibility, groundwater potentiality mapping, etc. [1,20–38]. The multi-collinearity was tested by applying the variance inflation factor (VIF) and tolerance [78]. The VIF is reciprocal to the tolerance values, calculated using the following Equations (1) and (2):

$$Tolerance = 1 - R_J^2 \tag{1}$$

$$VIF = \left[\frac{1}{Tolerance}\right] \tag{2}$$

where R_J^2 is the coefficient of determination of a regression of explanator *j* on all the other explanators. The threshold value of tolerance and VIF is more than 0.1 and less than 10 [79]. Arabameri et al. [1] and Roy and Saha [33] used the multi-collinearity analysis to test the topographical, geomorphological, and environmental factors for LS mapping. In the current research, the multi-collinearity of the sixteen LCFs was tested using the SPSS software.

2.6. Landslide Susceptibility Modeling

2.6.1. Applying Random Forest (RF)

RF is one of the main machine learning approaches [80]. The RF method generates multiple classification trees and aggregates them to compute a classification [81]. Hansen and Salamon [82] point out that ensemble classification trees are much more precise than any of their individual members. RF adds diversity among the classification trees by changing the data with alternate and modifying the explanatory factor sets arbitrarily over the different processes of tree induction. The two user-defined variables that are needed to grow a random forest are the number of trees (k) and the number of predictive factors used to divide the nodes (m). The numerical and categorical variables are used for conditioning factors. The out-of-bag (OOB) error is characterized as the percentage of the total number of out-of-bag items that are misclassified. The OOB error is a rational generalization error estimation. The OOB error is estimated at the time of construction of the RF model. Breiman [80] mentioned that the RF produces a limiting value of the error of generalization. The generalization errors often fall together as the number of trees grows. The k has to be set high enough to allow this convergence. The RF method calculates the value of a predictor variable by examining how much the OOB error decreases as OOB data is permuted for that variable while being constant for all other variables. The rise in the error of OOB corresponds to the value of the explanatory variable [83]. One of the major advantages of RF is the resistance to overtraining and the development of a large number of random forest trees where it doesn't really pose a risk of overfitting (e.g., each tree is a random experiment completed independently). There is no need to rescale, transform, or change the RF algorithm. For predictors, RF has resistance to outliers and manages the missing values automatically [76]. In this work, the RF model was computed using the "Randomforest" package in R 3.8, whereby, after a primary discursion, the number of trees was set to 1000 and the m samples at each node was chosen to be 3 to calculate the mutual contribution of attribute subsets while preserving gradual convergence during iterations. No training set is required to monitor the parameters [84,85]. The mean decrease in Gini error and accuracy were analyzed.

2.6.2. Applying Alternating Decision Tree (ADTree)

ADTree is one of the machine learning ensemble classifiers that incorporates the decision tree and thus improves the algorithm [49]. ADTree is related to direct decision tree generalizations in which each node is separated into two nodes, including a node of prediction and a node of splitter. For a mapping of the base ruler, the actual number consists of two conditions, namely a precondition (c_1) and a base condition (c_2). ADTree was detected the two real numbers a and b. The prediction is if $c_1 \cap c_2$ or b if $c_1 \cap -c_2$ – identified as negation (NOT). The two real numbers a and b are derived using Equations (3) and (4):

$$a = \frac{1}{2} In \frac{W_+(c_1 \cap c_2)}{W_-(c_1 \cap c_2)} \right) \tag{3}$$

$$b = \frac{1}{2} In \frac{W_+(c_1 \cap^- c_2)}{W_-(c_1 \cap^- c_2)} \tag{4}$$

where $W(p)$ is the total weight of the p predicate training instances. The best c_1 and c_2 are calculated by minimizing the $Z_t(c_1,c_2)$ calculated using Equation (5):

$$Z_t(c_1,c_2) = 2 \sqrt{W_+(c_1 \cap c_2)W_-(c_1 \cap c_2)} + \sqrt{W_+(c_1 \cap -c_2)W_-(c_1 \cap -c_2)} + W(-c_2) \tag{5}$$

When R symbol is believed to be a set of basic rules, a new base rule will be defined as $R_{t+1} = R_t + r_t.r_t(x)$ showing the two predictive values of *a* and *b* at each predictive layer of T. *X* is a set of instances. An instance classification is the symbol of the sum of all predictive values in R_{t+1}:

$$Class(x) = sign\left(\sum_{t=1}^{T} r_t(x)\right) \tag{6}$$

The algorithm begins by choosing the best constant forecasting for the whole dataset [49]. Typically, cross-validation is useful to build the selection [86].

2.6.3. Applying Fisher's Linear Discrimination Analysis (FLDA)

FLDA is another important machine learning approach. FLDA is a feature recognition method used in several fields [87]. Theoretically, *m* is used for the method as the recognize classes. $X_j^{(i)}$ indicates the *j*-th class *i* training trial. X^i corresponds to the mean of class *i* training trials, while *X* refers to complete training trials. The supposed learning trials M_b and M_w show a scatter matrix for class and with-class and can be determined using Equations (7) and (8) as well:

$$M_b = \frac{1}{N} \sum_{i=1}^{m} N_i (X^i - X)(X^i - X)^T \tag{7}$$

$$M_w = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{Ni} (X_j^{(i)} - X^i)(X_j^{(i)} - X^i)^T \tag{8}$$

where N_i indicates the number of training trials in class $i (\sum_{i=1}^{m} S_i = S)$ and *N* is the total number of training trials carried out. FLDA aims at obtaining an optimal set of defining vectors for computing $\Phi_d = (\varphi_1, \varphi_2, \ldots \ldots \varphi_d)$ transformation by optimizing the Fisher criterion, as indicated [88]:

$$J(\Phi) \stackrel{\text{def}}{=} \frac{tr(\Phi_d^T M_b \Phi_d)}{tr(\Phi_d^T M_w \Phi_d)} \tag{9}$$

where *T* is the matrix's transpose. Because benchmark maximization allows the optimal discriminating vectors, a vector can be derived for one input instance, and the following vector can be used to categorize the subsequent instance [89].

2.7. Considering the Contribution of Landslide Conditioning Factors

2.7.1. Boosted Regression Tree (BRT) Model

BRT is the non-parametric machine learning method which establishes a relationship between dependent and independent variables [90]. It is a crucial means of measuring the importance of individual variables and to overcome grouping and forecasting problems [91]. The response to a parameter depends on the higher node values in the regression tree's hierarchical structure, and then the relationships between the factors are automatically built [92]. At first, the divisor variable is used to split the target variable in the BRT modeling. To select the best divisor variable, the data was classified into two groups [93]. The recursive aspect point segmentation splits each individual node into two nodes and ultimately separates these divided nodes. This differentiation ends when nodes become homogeneous [91]. The point of the regression analysis is to make a decision during postpone to the divisions. The tree algorithm splits all the data and can lead to overfitting if the divisions do not

stop [94]. Trees are more complex and fragile, i.e., they have more errors in the new observations. BRT is calculated using the following Equation (10) [95]:

$$F\left(X; [\beta_m \alpha_m]^{m_0}\right) = \sum_{m=0}^{m} \beta_m h(x; \alpha_m) \tag{10}$$

where $h(x; m)$ is described as a basic function of classification with parameter a and variables x, m is the stage of the model. BRT was computed using R 3.3.1 and the "Rattle" package [96].

2.7.2. Application of Frequency Ratio Model

The frequency ratio (FR) is a geospatial assessment tool and one of the important bivariate statistical methods [97]. The FR values are determined from the relationship between landslide locations and LCFs. The result of FR is useful and appropriate [98]. The FR can be described by Equation (11) [99]:

$$FR = \frac{(A/B)}{(C/D)} \tag{11}$$

where A is the number of landslide pixels in each class of LCFs, B is the total number of landslide pixels in the watershed, C is the total number of pixels of each sub-class of the LCFs, and D is the total number of pixels of the watershed.

2.8. Methods for Validating the Models

Validation is necessary for the justification of the LSMs created using different models. Without validation, the LSMs have no value or meaning [100]. The area under receiver operating characteristics (AUROC) curve was used to test the various models. The AUROC curve is a threshold-independent tool for the measurement of predictive performance [101–106]. The AUROC indicates the model's predictive accuracy. The AUROC is expressed as the percentage of prediction values [107]. Generally, the value of AUROC ranges from 0.5 to 1. The closer an AUROC value is to 1, the better the performance of the method [9,10]. The AUROC values were classified by Yeashiar and Topal [108] into five classes: low (0.5–0.6), moderate (0.6–0.7), good (0.7–0.8), very good (0.8–0.9) and excellent (0.9–1). Using all the learning and testing datasets, this approach was used to construct a curve, and subsequently, it was used to calculate the goodness of fit and model prediction accuracy [109]. The goodness of fit is superior in measuring the predictive performance of the models using the training and testing datasets [89]. In this work, the AUROC curve was constructed based on the training and testing landslide datasets. The AUROC curve plot consists of two variables, namely the true positive rate (TPR) and the false positive rate (FPR), called the sensitivity and specificity [110]. The TPR shows the percentage of landslide pixels accurately categorized as landslide occurrences, and the FPR indicates the percentage of non-landslide pixels wrongly identified as landslides [111]. The Formulas (12) and (13) were used to calculate the *TPR* and *FPR*:

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

where *TP* (true positive) and *TN* (true negative) are accurately classified pixel numbers, while *FP* (false positive) and *FN* (false negative) are falsely classified pixel numbers. The AUROC reflects the reliability of simulations in correctly predicting landslide occurrence or non-occurrence [108]. The SCAI validation technique used in this study was developed by Süzen and Doyuran [112]. The SCAI is defined as the ratio between the percentage of pixels of the specific landslide susceptibility class and the percentage of the pixels of existing landslides in this specific landslide susceptibility zone. Two threshold-dependent measurement metrics, namely the efficiency (E) and true skill statistic (TSS), were

also used to analyze the performance of the models [113]. The E and TSS can be calculated using Equations (14) and (15):

$$E = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$TSS = TPR - FPR \tag{15}$$

3. Results

3.1. Considering Multi-Collinearity of Factors Contributing to Landslide Susceptibility

A multi-collinearity test was performed on the LCFs. Chen et al. [14] mentioned that the linear relationship between the parameters would minimize the model's predictability. The tolerance and variance inflation factors (VIF) were utilized to test the multi-collinearity. Tolerance and VIF values of <0.1 and above 10 indicate multi-collinearity problems. Arabameri et al. [1] and Roy and Saha [34] used Tolerance and VIF techniques to select suitable parameters for LSMs. In this study, different LCFs, namely elevation, slope, slope length, plan curvature, rainfall, CI, TPI, TWI, SPI, LULC, NDVI, lithology, soil type, distance from stream, distance from fault and distance to road were selected using the multi-collinearity test for the creation of LSMs using different models. The outcomes of the multi-collinearity analysis in this study (Table 2) show that the effective factors of landslides have no multi-collinearity problem because all the LCFs have lower than threshold values of tolerance and VIF. The maximum tolerance value was identified for TPI (0.955), and the VIF value for TWI (4.55) was identified from the conditioning factors. Therefore, all the proposed landslide conditioning factors were considered to be suitable to produce the LSMs using RF, ADTree and FLDA models in Gallicash catchment area.

Table 2. Multi-collinearity analysis among landslide conditioning factors.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	−1.017	0.222	–	−4.580	0.000	–	–
Rainfall	0.001	0.000	0.269	3.644	0.000	0.392	2.552
Soil type	0.303	0.111	0.150	2.744	0.006	0.711	1.407
LU/LC	0.186	0.040	0.249	4.710	0.000	0.766	1.306
Lithology	0.223	0.063	0.174	3.547	0.000	0.883	1.132
Slope	−0.005	0.004	−0.094	−1.162	0.246	0.324	3.085
Dis to road	-2.427×10^{-5}	0.000	−0.111	−2.177	0.030	0.824	1.214
TWI	−0.060	0.019	−0.402	−3.162	0.002	0.232	4.566
Dis to fault	-1.372×10^{-5}	0.000	−0.090	−1.594	0.112	0.667	1.498
PC	0.008	0.008	0.053	1.028	0.305	0.800	1.249
TPI	0.011	0.006	0.082	1.723	0.086	0.955	1.048
Elevation	0.000	0.000	−0.146	−2.512	0.012	0.628	1.592
NDVI	−0.453	0.268	−0.122	−1.692	0.092	0.410	2.437
CI	0.000	0.001	0.021	.351	0.726	0.593	1.687
Dis to stram	0.000	0.000	−0.141	−2.705	0.007	0.785	1.274
SPI	0.092	0.021	0.506	4.337	0.000	0.257	4.368
LS	−0.056	0.014	−0.345	−3.167	0.004	0.332	2.756

3.2. The Spatial Relationship Between Landslide Locations and Effective Factors by FR

The frequency ratio (FR) values were calculated based on the relationship between locations of landslides and the sixteen LCFs. The mathematical calculation of the FR value is presented in Table 3. An FR value over 1 indicates a high landslide susceptibility, whereas an FR value of 0 or close to 0 indicates that the sub-class of landslide conditioning factors does not affect landslide occurrence. The FR model was applied to the landslide conditioning factors using Equation (11).

Table 3. The spatial relationship between LCFs and landslides by FR model.

Factors	Class	Pixels in Domain		Pixels of Landslide		FR
		No	%	No	%	
Elevation (m)	<354	1,757,987	28.85	64	39.02	1.35
	354–768	1,528,329	25.08	48	29.27	1.17
	768–1216	1,255,594	20.60	40	24.39	1.18
	1216–1658	979,570	16.07	11	6.71	0.42
	>1658	573,032	9.40	1	0.61	0.06
Slope (^{0})	<5	1,533,843	25.17	33	20.12	0.80
	5–10	877,089	14.39	42	25.61	1.78
	10–15	1,031,179	16.92	27	16.46	0.97
	15–20	947,857	15.56	17	10.37	0.67
	20–30	1,198,335	19.67	37	22.56	1.15
	>30	505,070	8.29	8	4.88	0.59
PC	Concave	2,686,090	44.08	79	48.17	1.09
	Flat	528,638	8.68	6	3.66	0.42
	Convex	2,878,644	47.24	79	48.17	1.02
SPI	<9.08	893,698	14.68	11	6.71	0.46
	9.08–10.99	2,044,913	33.60	61	37.20	1.11
	10.99–12.9	2,162,864	35.54	58	35.37	1.00
	12.9–15.8	801,588	13.17	18	10.98	0.83
	>15.8	182,979	3.01	16	9.76	3.24
TPI	<−1.83	303,487	4.98	8	4.88	0.98
	−1.83–−0.18	1,031,274	16.92	23	14.02	0.83
	−0.18–1.37	1,212,301	19.89	26	15.85	0.80
	1.37–3.99	492,748	8.09	17	10.37	1.28
	>3.99	3,054,702	50.12	90	54.88	1.09
TWI	<4.89	2,838,332	46.57	73	44.51	0.96
	4.89–7.33	2,261,932	37.11	60	36.59	0.99
	7.33–11.08	798,883	13.11	14	8.54	0.65
	>11.08	195,364	3.21	17	10.37	3.23
LS (m)	<30.7	781,241	25.68	19	22.62	0.88
	30.7–65.08	720,877	23.69	20	23.81	1.00
	65.08–101.1	632,770	20.80	30	35.71	1.72
	101.1–140.7	527,320	17.33	10	11.90	0.69
	>140.7	380,217	12.50	5	5.95	0.48
CI (100/M)	<−32.54	586,802	9.63	22	13.41	1.39
	−32.54–−9.8	1,196,806	19.64	37	22.56	1.15
	−9.8–9.01	2,386,816	39.17	64	39.02	1.00
	9.01–31.76	1,387,504	22.77	31	18.90	0.83
	>31.76	536,304	8.80	10	6.10	0.69
Dis to road (m)	<500	762,121	12.51	46	28.05	2.24
	500–1000	709,095	11.64	30	18.29	1.57
	1000–1500	652,772	10.71	14	8.54	0.80
	1500–2000	591,152	9.70	7	4.27	0.44
	>2000	3,378,964	55.45	67	40.85	0.74
Dis to stream (m)	<100	1,058,295	17.38	70	42.68	2.46
	100–200	844,165	13.86	29	17.68	1.28
	200–300	834,301	13.70	10	6.10	0.45
	300–400	644,273	10.58	9	5.49	0.52
	>400	2,708,520	44.48	46	28.05	0.63
Rainfall (mm)	<425.9	929,156	15.25	2	1.22	0.08
	425.9–561.6	1,861,610	30.56	86	52.44	1.72
	561.6–703.6	1,490,446	24.47	18	10.98	0.45
	703.6–851.9	1,176,590	19.31	25	15.24	0.79
	>851.9	634,325	10.41	33	20.12	1.93

Table 3. *Cont.*

Factors	Class	Pixels in Domain		Pixels of Landslide		FR
		No	%	No	%	
Lithology	A	453,438	7.42	1	0.61	0.08
	B	3,068,975	50.19	104	63.41	1.26
	C	74,810	1.22	0	0.00	0.00
	D	1,250,277	20.45	26	15.85	0.78
	E	543,327	8.89	13	7.93	0.89
	F	429,394	7.02	9	5.49	0.78
	G	236,729	3.87	6	3.66	0.95
	H	58,043	0.95	5	3.05	3.21
LU/LC	Afforest	1,582,944	25.89	20	12.20	0.47
	Agriculture	638,404	10.44	14	8.54	0.82
	Orchard	931	0.02	0	0.00	0.00
	Dry farming	2,883,017	47.15	74	45.12	0.96
	Water	42,648	0.70	1	0.61	0
	Agri-Orch	901,557	14.74	51	31.10	2.11
	Rock	5,361	0.09	1	0.61	6.95
	Urban	59,975	0.98	3	1.83	1.87
NDVI	<−0.201	3,517,295	57.72	98	59.76	1.04
	−0.201–0.36	929,798	15.26	41	25.00	1.64
	>0.36	1,647,011	27.03	25	15.24	0.56
Soil type	Rock Outcrops/Entisols	1,566,585	25.62	28	17.07	0.67
	Inceptisols	1,210,719	19.80	36	21.95	1.11
	Mollisols	1,797,624	29.40	44	26.83	0.91
	Alfisols	1,540,065	25.19	56	34.15	1.36
Dis to fault (m)	<500	670,236	11.00	15	9.15	0.83
	500–1000	637,781	10.47	14	8.54	0.82
	1000–1500	581,006	9.53	11	6.71	0.70
	1500–2000	503,170	8.26	11	6.71	0.81
	>2000	3,701,746	60.74	113	68.90	1.13

The topographic factors are of great importance for the preparation of LSM [98]. The elevation has a strong relationship with landslides occurrences. The sub-classes <354 m of elevation dataset has the highest FR values (FR = 1.35), indicating the zone of highest landslide susceptibility. Precipitation is the primary cause of landslide occurrences of a region. The long duration of heavy precipitation directly affects landslides due to the loosening of soils, increase of surface runoff velocity and accelerated rate of sediment transmission. The FR results showed that the sub-layer of >851 mm, which is the highest rainfall in the study area, has the maximum FR value of 1.93, followed by the sub-classes of <425 mm (0.08), 425–561 mm (1.72) and 561–703 mm (0.45). In the case of the slope, the steeply sloping area is the most susceptible to landslide occurrences in comparison to gently sloping areas. In this study, 5–10 degree and 20–30 degree slope angles are the classes with the maximum FR values of 1.78 and 1.15, followed by the <5 degrees (0.80), 10–15 degrees (0.97) and >30 degrees (0.59) sub-classes. For the slope length factor, the sub-class of 65.08–101.1 m has the highest FR value of 1.72, indicating a high landslide susceptibility, followed by the 30.7–65.08 m <30.7 m, 101.1–140.7 m and >140.7 m (0.48), respectively (Table 3). The topographic position index is one of the factors that determines the landslides occurrences. The sub-layer of 1.37–3.99 of TPI has the maximum FR values of 1.28. The highest FR values usually indicate zones of the highest landslides susceptibility, while lower FR values of the sub-layers of TPI indicate zones of lesser landslide susceptibility. With the help of the FR model, the concave (FR = 1.09) and convex (RF = 1.02) sub-layers of Plan curvature have been assigned more weights than the flat areas. The convex and the concave areas of mountains are more susceptible to landslides, while flat areas are not susceptible to landslides.

The convergence index (CI) is the topographic factor that plays a role in the landslides occurrence of a region. The values of CI vary from +100 to −100. The sub-layers of <−32.54 in the present study have a maximum FR value of 1.39, preceded by −32.54 to −9.8 (FR = 1.15) and −9.8 to 9.1 (FR = 1.0), respectively. The regions that have negative values of CI are more prone to landslides than the areas that have positive CI values. The correlation between landslide locations and road distance indicates that the subset <500 m has the maximum FR value (2.24), which indicates an extreme landslide risk. Landslides frequently occur close to roads in this region. The distance sub-classes of 1000−1500 m, 1500–2000 m and >2000 m indicate a weaker relationship to landslide occurrence with greater distance to the roads. In terms of the distance to streams, close proximity to a stream has a strong control on landslide occurrence. The FR results show that the sub-layer of <110 m, which is the least distance from a stream, has a direct impact on the occurrence of landslides due to the high velocity and discharge of the streams. Comparatively, a greater distance to streams does not directly affect the occurrence of landslides. The sub-layers of 200–300 m, 300–400 m, and >400 m distance have low FR values of 0.45, 0.52 and 0.63, respectively. The presence of a fault is one of the causes of landslides. Here, the sub-layer of >2000 m has the maximum FR value of 1.13, preceded by <500 m (0.83), 500−1000 m (0.82), 1000–1500 m (0.70) and 1500–2000 m (0.81), representing an increasing risk of landslides. The stream power index (SPI) directly affects the landslide susceptibility. The maximum stream power can transport the sediment and erode soil particles immediately, thus creating a favourable situation for landslides. The FR results show that the sub-class of >15.8 has the highest FR value of 3.24 and the highest landslide incidence, followed by the other sub-classes of <9.08 (0.46), 9.08–10.99 (1.11), 10.99–12.9 (1) and 12.9–15.8 (0.83), respectively. The geology, soil type, LU/LC and NDVI are the major landslide determining factors in this region. Geologically, the area consists of eight geological units. The outcomes of the FR model demonstrate that the geological sections H (Tre and TRJs lithological units) and B (Dp, and DCkh lithological units) have high FR values of 3.21 and 1.26, followed by A (0.08), C (0), D (0.78), E (0.89), F (0.78) and G (0.95), respectively. In the case of the soil type factor, the fragile, weak and softest soils have the highest risk of landslides of the region. The results of FR show that the inceptisols (FR = 1.11) and alfisols (FR = 1.36) have a stronger influence on landslide occurrence than the rock (0.67) and mollisols (0.91). LULC also plays a crucial role in the occurrence of landslides. Eight LULC classes were identified in the study area, namely afforested area, agriculture land, orchard, dry farming, water, agriculture-orchard, rock outcrops. The majority of landslides occurred in the mountain rocks area (FR = 6.95) followed by agriculture-orchard (FR = 2.11), urban areas (FR = 1.87), the forest (FR = 0.47), agriculture (FR = 0.82) and dry farming (FR = 0.96) areas. A high vegetation density can reduce the landslide risk and also postpone the soil transmission and erosion, while a high concentration of landslides can be found in the low vegetation area, which is more susceptible to the occurrence of landslides. In this study, the sub-class of −0.201 to 0.36 has the maximum FR of 1.64, suggesting susceptibility to landslides. Therefore, the FR model depicts the reciprocal relationship of the landslide locations and landslide conditioning factors, taken as the research purpose to manage the landslide mitigation. The pixel-based study using FR is more accurate and appropriate for landslide susceptibility mapping.

3.3. Landslide Susceptibility Models

In this landslide susceptibility assessment, the dependent factors were expressed as binary variables, i.e., landslides and non-landslides. The numerical and categorical values of landslide conditioning factors were extracted for creating the LSMs based on the models using the program R. The landslide susceptibility mapping based on the random forest (RF) was constructed using the relative weight of the mean decrees in accuracy and the mean decrees in accuracy of the Gini index of LCFs (Table 7). The RF model results show that the very low, low, moderate, high and very high landslide susceptibility zones of the LSM cover 25.75%, 28.61%, 19.46%, 14.82% and 11.36% areas of the watershed, respectively (Figure 5). Only 11.36% of the basin, mainly in the middle portion along the north-south line, has an even higher risk of landslides. Likewise, in the GIS setting, the LSM

generated using the FLDA model was categorized into five sub-layers (Figure 6). The results of the LSM based on the FLDA model show that 16.10% and 26.07% of the area have very high and high landslide susceptibility (LS), while 10.64%, 22.20%, and 24.99% of the area have very low, low, and moderate LS. The ADTree LS map was prepared by summarizing all prediction node values in the GIS environment (Figure 7). In this study, the ADTree model was classified into the five classes of very low, low, medium, high and very high according to the NBM of Jenk's. From total of study area, 31.27% has very low susceptibility to landslide occurance), 37.81% has low, 1.53% has moderate, 14.03 % has high and 15.37% has very high susceptibility to gully co-occurance.

Figure 5. Landslide susceptibility map based on the RF model.

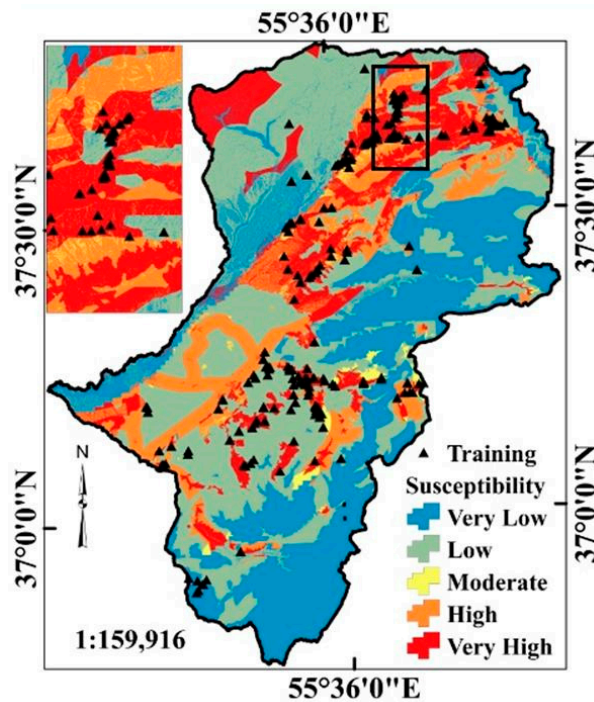**Figure 6.** Landslide susceptibility map based on the FLDA model.

Figure 7. Landslide susceptibility map based on the ADTree model.

3.4. Validation of Machine Learning Models

The success and predication curves (Figure 8) were generated using the training and testing data sets by AUROC curves. The AUROC values of the success rate curves (SRC) are 0.896 (89%) for ADTree, 0.927 (92%) for FLDA and 0.972 (97%) for RF models, which means that the LSMs are very good to excellent. The AUC values of the prediction rate curve (PRC) are 0.822 (82%) for ADTree, 0.795 (79%) for FLDA and 0.985 (98%) for RF models, which is also very good to excellent for the resulting LSMs (Figure 8). Very similar results have been found for the success and prediction curves. Of the selected machine learning ensemble methods, the RF model produces a more accurate LSM.
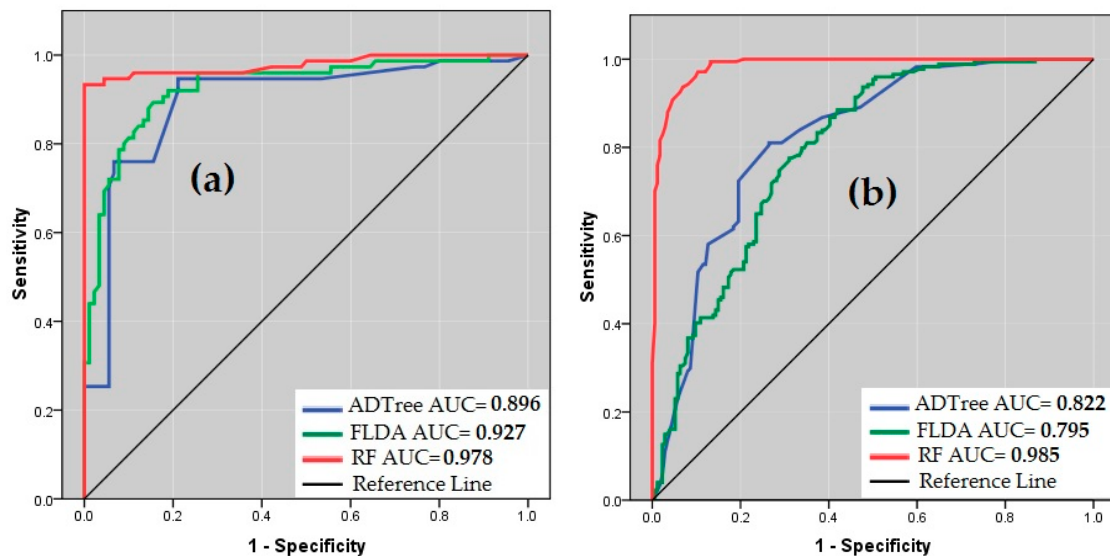

Figure 8. Validation of results: (**a**) Success rate curve (SRC), (**b**) Prediction rate curve (PRC).

The seed cell area index (SCAI) is a reliable validation technique. The validation results of the LSMs using the SCAI method are shown in Table 4. SCAI is the ratio of the percentage of each sub-layer

to the percentage of landslides that occur in the class [52]. If values of the SCAI, decreased from very low to very high classes of LS, the model is regarded as excellent [1,14–37]. The LSMs produced based on the ADTree, FLDA and RF models were categorized into five classes. These landslide susceptibility classes and the number of pixels of training and validation datasets were computed using the SCAI statistical method. The SCAI values of very high landslides susceptibility classes are 0.28, 0.34 and 0.14 for the FLDA, ADTree, and RF models, respectively (Table 4).

Table 4. Percentage of each susceptibility class along with SCAI values.

Model	Class	Pixels in Domain		Pixels of Landslide		SCAI
		No	%	No	%	
ADTree	Very Low	1,865,100	31.27	7	3.07	10.18
	Low	2,255,466	37.81	43	18.86	2.00
	Moderate	91,472	1.53	5	2.19	0.70
	High	836,774	14.03	49	21.49	0.65
	Very High	916,639	15.37	124	54.39	0.28
FLDA	Very Low	648,691	10.64	2	0.88	12.14
	Low	1,352,591	22.20	7	3.07	7.23
	Moderate	1,522,829	24.99	30	13.16	1.90
	High	1,588,744	26.07	82	35.96	0.72
	Very High	981,082	16.10	107	46.93	0.34
RF	Very Low	1,569,307	25.75	3	1.32	19.57
	Low	1,743,560	28.61	3	1.32	21.75
	Moderate	1,185,580	19.46	14	6.14	3.17
	High	902,913	14.82	27	11.84	1.25
	Very High	692,408	11.36	181	79.39	0.14

The efficiency values of the FLDA, ADTree, and RF models using the testing dataset are 0.69, 0.77 and 0.69, respectively. The efficiency values of the FLDA, ADTree, and RF models using a training dataset are 0.74, 0.77 and 0.69, respectively. The TSS values of the FLDA, ADTree, and RF models using the testing dataset are 0.37, 0.55 and 0.39, respectively (Table 5). The TSS values of the FLDA, ADTree, and RF models using the training dataset are 0.47, 0.53 and 0.38, respectively. The AUROC, SCAI, efficiency and TSS methods are a good representation of reality. The outcome of the validation also indicates a strong relationship between the current landslide locations (validation data set) and the model's expected landslide locations. Above all, as per the validation methods used here, the RF machine learning technique performs better than the other models for creating LSMs in the Gallicash watershed in Iran.

Table 5. Values of efficiency (E) and true skill statistic (TSS).

Models	Validation Dataset (VD)			Training Dataset (TD)		
	FLDA	RF	ADTree	FLDA	RF	ADTree
TN	44	58	54	122	128	117
FP	31	17	21	52	46	57
FN	16	17	25	40	35	51
TP	59	58	50	134	139	123
TPR	0.79	0.77	0.67	0.77	0.80	0.71
FPR	0.41	0.23	0.28	0.30	0.26	0.33
efficiency	0.69	0.77	0.69	0.74	0.77	0.69
TSS	0.37	0.55	0.39	0.47	0.53	0.38

4. Discussion

4.1. Model Performance and Comparison

The northern parts of Iran are frequently affected by landslides, which is hindering the economic growth of this region. Landslide susceptibility mapping may contribute to an effective solution for this region to reduce the financial losses and human deaths caused by the landslides. The landslide susceptibility mapping is an invaluable tool for deploying protection measures and managing sudden landslide occurrences in this region. To prepare an LSM, selecting suitable models and geo-environment factors is an important task to properly forecast areas that may eventually be affected by a landslide. In the forested area, it is difficult to investigate the topographic conditions and to map the landslide susceptibility. Due to the obstruction of a large canopy cover, ASTER and SRTM low-resolution images cannot provide sufficient information on topographical characteristics in forest areas [114]. The scarcity of the morphological information in the densely forested areas is creating difficulties in the environmental management in the study area. These problems can be solved using radar, which is an effective and reliable remote sensing technique that provides accurate information regarding natural resources. synthetic aperture radar (SAR) techniques leads to lightweight cost-effective imaging sensors of high resolution [115]. The ALOS PALSAR, ERS 1, 2, JERS Radar sat and Envisat have been used in this field. The ALOS PALSAR DEM is more suitable for collecting information from densely forested areas and also produces a more accurate DEM than ASTER and SRTM [108,114,116,117]. In this study, the PASLAR DEM with a resolution of 12.5 m × 12.5 m was used to extract the topographical and hydrological characteristics in the Gallicash watershed.

Recently, machine learning techniques have become more popular in the scientific community for environmental modeling as these methods are helpful in effectivley explaining the complex relationships between environmental predictors and responses. So far, many machine learning ensemble approaches, including RF, FLDA, BRT, AdaBoost, multiboost, bagging, CART, NBT, ADTree have been used for the evaluation of LSMs [1,29–31]. different approaches and methods were developed and implemented around the world for the spatial assessment of environmental hazards.In this study, the RF, ADTree, and FLDA models are used for the assessment of the LSMs in the Gallicash watershed. The LSMs of the present study were prepared using the R program, remote sensing, and GIS. It is impossible to avoid the problem of overfitting because the sampling ratio of 70:30 was established for the training and test datasets without conducting a sampling ratio accuracy test, and despite a multi-collinearity check having been carried out, noise still exists in the landslide conditioning factors data. A key advantage of the machine learning algorithm used in this research is that it automates the task of searching multiple databases in order to collect crucial information. Such algorithms can compensate for particular theories that can be used together with automating the analysis of large amounts of data to support planning decisions.

The LSMs prepared using the RF, ADTree, and FLDA models were categorized into five separate landslide susceptibility classes, namely very low, low, moderate, high and very high, according to the four classification methods, namely quantile, natural breaks of the Jenk's classification, equal interval, and geometric interval. By comparing the results of each classification method and the distribution of landslides of the training and validation dataset on the high and very high classes of landslide susceptibility, it was noticed that the most reliable distribution was given by the natural break classification method. The main advantage of the natural break classification method is that there is no class bias, and the intra-class deviation is minimum and inter-class deviation is maximum. This is in line with the results of Arabameri et al. [1] since the natural break method is a reliable classifier in the mapping of landslide susceptibility. In the Gallicash watershed, the very high LS classes of the RF, ADTree, and FLDA models cover an area of about 11.36%, 15.37% and 16.10% (Table 4). To determine the goodness of fit and prediction accuracy of LSMs, the AUROC curve, SCAI, efficiency, and TSS was used in this study area. The AUROC curve produced the successive curve rate (SCR) and predication curve rate (PCR) using the training and validation datasets. The succession and the prediction values

of the AUROC curve are shown in Figure 8. The AUROC, SCAI, efficiency and TSS methods have shown that all the models are suitable and efficient enough for producing the LSMs. Our results of these models are rational and justified by the works of Arabameri et al. [1], Dieu Tien Bui et al. [31], Lee et al. [40] and Chen et al. [14]. Above all, the study shows that the RF model is more suitable for the landslide susceptibility mapping than the ADTree and FLDA models. The severe landslide risk zones in this region have been found in the middle part along the north-south axis of the study area, likely due to the mountainous, elevated, slopping, high convergence index and topographical position of the landforms in this area.

4.2. Variable Contribution Analysis

In this work, a vital machine learning ensemble approach, namely the RF model, was used for landslide susceptibility mapping in the Gallcash river watershed. In the program R, the presence or absence of landslides, expressed as 0 and 1, were considered as the target variables, and the LCFs were considered as the independent variables. The point's values of the studies have been extracted LCFs regarding the target variables. The results of the RF model show that the out-of-bag error is 4.7 % (Table 6). The RF algorithm calculates the importance of FCFs for the construction of LSM (Table 7). The LCFs such as TPI (21.09), distance to stream (7.72), elevation (5.8), rainfall (8.17), plan curvature (2.99), lithology (2.78), SPI (2.78), CI (2.93), TWI (3.01), distance to road (4.62), NDVI (3.31), soil (1.42), distance to fault (3.88), slope (3.35) and LULC (1.28) were assigned a weight by the RF model (Table 7). The TPI, distance to stream, elevation, and rainfall were shown to have a great effect on landslides, followed by plan curvature, lithology, SPI, CI, TWI, distance from road, NDVI, soil, distance from fault, slope and LULC.

Table 6. Confusion matrix from the RF model (0 = no landslide, 1 = landslide).

	0	1	Class Error
0	113	6	0.0504
1	20	104	0.1612

Table 7. The relative influence of landslide conditioning factors in the RF model.

Factor	Weight
TPI	21.09
Distance to stream	7.72
Elevation	5.8
Rainfall	8.17
Plan curvature	2.99
Lithology	2.78
SPI	4.42
CI	2.93
TWI	3.01
Distance to road	4.62
NDVI	3.31
Soil	1.42
Distance to fault	3.88
Slope	3.35
LULC	1.28
LS	1.01

The relative importance of each of the LCFs was also calculated using the BRT machine learning technique. The results of the BRT model indicate that the TPI (0.5552) has the highest contribution to landslide susceptibility, followed by distance to stream (0.1148), elevation (0.0573), rainfall (0.0497), plan curvature (0.0419), lithology (0.0306), SPI (0.0270), CI (0.0225), TWI (0.0224), distance to road

(0.0219), NDVI (0.0218), soil type (0.0165), distance to fault (0.0086), slope (0.0075), LULC (0.0017) and LS (0.0003). All the LCFs used in susceptibility modeling play a major role in making the area prone to landslides (Figure 9).
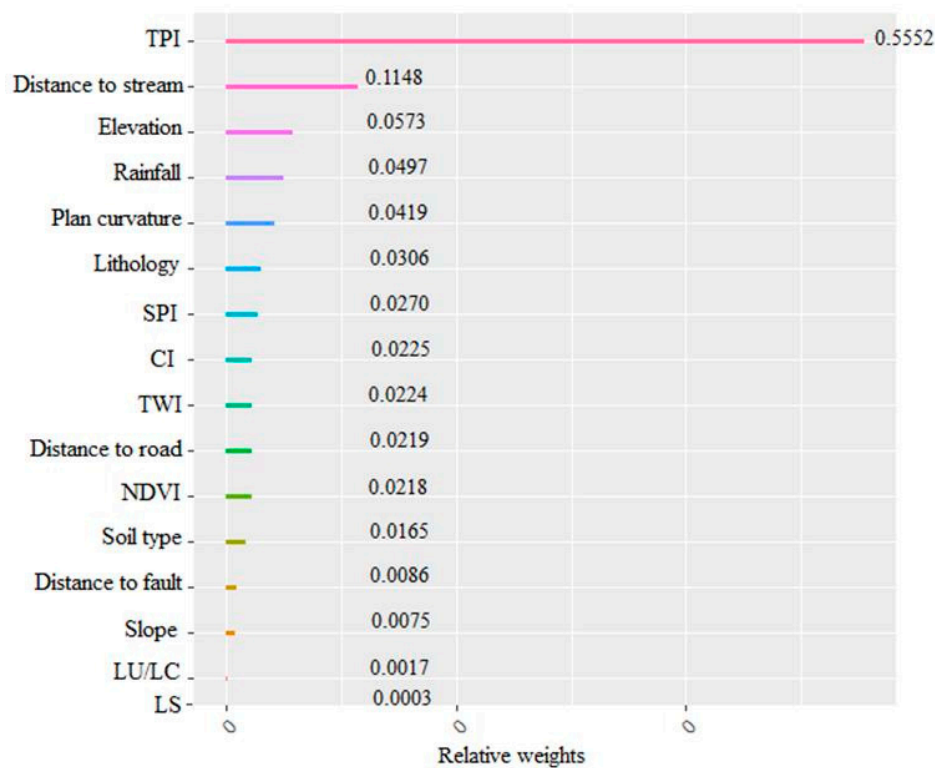

Figure 9. The relative importance of landslide conditioning factors using boosted regression tree model.

5. Conclusions

The presented study analyzes landslides and how they are impacted by tectonic activities, adverse topographical and hydrological conditions, land-use change, road construction and unnecessary activities of people in the Gallicash catchment, northern Iran. A total of 249 landslides were identified and analyzed to evaluate the LS conditions in the study area. LSMs based on an environmental hazards assessment is an important aspect of sustainable environmental management and planning. In this study, sixteen landslide conditioning factors and three reliable machine learning ensemble models, namely RF, ADTree and FLDA, were chosen to demonstrate the areas at risk of landslides. Landslide inventory datasets were considered to be the dependent variables and LCFs were presumed to be independent variables. The dependent variables are expressed as the binary values of 0 and 1, where 1 and 0 indicate the presence and absence of landslides. The high-quality data, such as the ALOS PALSAR DEM and Sentinel-2 image with a resolution of 12.5 m × 12.5 m were used for extracting the topographic, hydrological and geomorphological datasets to prepare LSMs. Of the different lithological segments in this study area, the H (Tre and TRJs lithological units) and B (Dp, and DCkh lithological units) geological sections have a strong influence on the occurrence of landslides. Similarly, alfisoil and inceptisoil soils contribute significantly to landslides occurring in this region. The machine learning software R Studio was used to run machine learning algorithms such as RF, ADTree, FLDA, etc. RS and GIS are the basic tools for performing the landslide susceptibility mapping. The integrations of R, RS and GIS have created a strong foundation for the construction of the LSMs using machine learning algorithms. The results of RF, ADTree and FLDA models show that 11.36% (RF), 15.37% (ADTree), and 16.10% (FLDA) of the area of the whole catchment belong to the very high landslide susceptibility class. Several common techniques, such as the ROC curve, SCAI, Efficiency and TSS, were used to verify the performance of the three chosen models. All models have strong LS simulation abilities. Finally, the

RF model shows slightly better accuracy (AUC of SRC = 0.978, AUC of PRC = 0.985, the efficiency of VD = 0.77, efficiency of TD = 0.77, TSS of VD = 0.55 & TSS of TD = 0.53) than the other models in modeling the landslide susceptibility in this region as per the results of the validation techniques. The most critical zones of landslide susceptibility have been identified in the central part along the north-south axis in the Gallicash catchment. This could identify the relationships between the spatial distribution of landslide locations and landslide conditioning factors. The LSMs prepared based on the RF, ADTree, and FLDA may be utilized by decision-makers, planners, academic researchers, engineers, and government agencies to mitigate the loss of human life and property. The results of this study will also support sustainable development measures and management of the environment in the Gallicash catchment and landslides prone areas in general.

References

1. Arabameri, A.; Pradhan, B.; Rezaei, K.; Sohrabi, M.; Kalantari, Z. GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. *J. Mt. Sci.* **2019**, *16*, 595–618. [CrossRef]

2. IAEG Commission on Landslides. Suggested nomenclature for landslides. *Bull. Int. Assoc. Eng. Geol.* **1990**, *41*, 3–16. [CrossRef]

3. Lin, L.; Lin, Q.; Wang, Y. Landslide susceptibility mapping on a global scale using the method of logistic regression. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1411–1424. [CrossRef]

4. Nadim, F.; Kjekstad, O.; Peduzzi, P. Global landslide and avalanche hotspots. *Landslides* **2006**, *3*, 159–173. [CrossRef]

5. Haftlang, K.K.; Lang, K.K.H. *The Book of Iran: A Survey of the Geography of Iran*; Alhoda: Tehran, UK, 2003; p. 17. ISBN 978-964-94491-3-5.

6. Aghda, S.F.; Bagheri, V.; Razifard, M. Landslide Susceptibility Mapping Using Fuzzy Logic System and Its Influences on Mainlines in Lashgarak Region, Tehran, Iran. *Geotech. Geol. Eng.* **2018**, *36*, 915–937. [CrossRef]

7. National Geosciences Database. 2017. Available online: www.ngdir.ir (accessed on 21 August 2018).

8. Piacentini, D.; Devoto, S.; Mantovani, M.; Pasuto, A.; Prampolini, M.; Soldati, M. Landslide susceptibility modeling assisted by Persistent Scatterers Interferometry (PSI): An example from the northwestern coast of Malta. *Nat. Hazards* **2015**, *78*, 681–697. [CrossRef]

9. Pradhan, A.M.S.; Kim, Y.T. Relative effect method of landslide susceptibility zonation in weathered granite soil: A case study in Deokjeok-ri Creek, South Korea. *Nat. Hazards* **2014**, *72*, 1189–1217. [CrossRef]

10. Hong, H.; Tsangaratos, P.; Ilia, I. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Sci. Total Environ.* **2018**, *625*, 575–588. [CrossRef]

11. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Ahmad, B.B. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena* **2018**, *163*, 399–413. [CrossRef]

12. Ahlmer, A.K.; Cavalli, M.; Hansson, K.; Koutsouris, A.J.; Crema, S.; Kalantari, Z. Soil moisture remote-sensing applications for identification of flood-prone areas along transport infrastructure. *Environ. Earth Sci.* **2018**, *77*, 533. [CrossRef]

13. Nsengiyumva, J.; Luo, G.; Nahayo, L.; Huang, X.; Cai, P. Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation Model in Rwanda. *Int. J. Environ. Res. Public Health* **2018**, *15*, 243. [CrossRef]

14. Chen, W.; Xie, X.; Peng, J.; Shahabi, H.; Hong, H.; Bui, D.T.; Duan, Z.; Li, S.; Zhu, A.X. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena* **2018**, 1–17. [CrossRef]

15. 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]

16. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. [CrossRef]

17. Arabameri, A.R.; Pourghasemi, H.R.; Yamani, M. Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environ. Earth Sci.* **2017**, *76*, 832. [CrossRef]

18. Pradhan, B.; Jebur, M.N.; Shafri, H.Z.M.; Tehrany, M.S. Data fusion technique using wavelet transform and taguchi methods for automatic landslide detection from airborne laser scanning data and QuickBird satellite imagery. *Trans. Geosci. Remote Sens.* **2016**, *54*, 1–13. [CrossRef]

19. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [CrossRef]

20. Arabameri, A.; Pradhan, B.; Rezaei, K.; Lee, C.W. Assessment of Landslide Susceptibility Using Statistical-and Artificial Intelligence-Based FR–RF Integrated Model and Multiresolution DEMs. *Remote Sens.* **2019**, *11*, 999. [CrossRef]

21. Arabameri, A.; Pradhan, B.; Pourghasemi, H.; Rezaei, K.; Kerle, N. Spatial Modelling of Gully Erosion Using GIS and R Programing: A Comparison among Three Data Mining Algorithms. *Appl. Sci.* **2018**, *8*, 1369. [CrossRef]

22. Arabameri, A.; Rezaei, K.; Pourghasemi, H.R.; Lee, S.; Yamani, M. GIS based gully erosion susceptibility mapping: A comparison among three data-driven models and AHP knowledge-based technique. *Environ. Earth Sci.* **2018**, *77*, 628. [CrossRef]

23. Arabameri, A.; Pradhan, B.; Rezaei, K.; Yamani, M.; Pourghasemi, H.R.; Lombardo, L. Spatial modelling of gully erosion using Evidential Belief Function, Logistic Regression and a new ensemble EBF–LR algorithm. *Land Degrad. Dev.* **2018**, *29*, 4035–4049. [CrossRef]

24. Arabameri, A.; Pourghasemi, H.R. Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, 1st ed.; Pourghasemi, H.R., Gokceoglu, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2019.

25. Arabameri, A.; Pradhan, B.; Rezaei, K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J. Environ. Manag.* **2019**, *232*, 928–942. [CrossRef]

26. Arabameri, A.; Rezaei, K.; Cerdà, A.; Conoscenti, C.; Kalantari, Z. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Sci. Total Environ.* **2019**, *660*, 443–458. [CrossRef]

27. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Total Environ.* **2019**, *658*, 160–177. [CrossRef]

28. Arabameri, A.; Pradhan, B.; Rezaei, K. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci. J.* **2019**, *1*, 1–18. [CrossRef]

29. 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]

30. Hong, H.; Liu, J.; Zhu, A.X.; Shahabi, H.; Pham, B.T.; Chen, W.; Pradhan, B.; Bui, D.T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ. Earth Sci.* **2017**, *76*, 689. [CrossRef]

31. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Pradhan, B.; Chen, W.; Khosravi, K.; Panahi, M.; Bin Ahmad, B.; Saro, L. Land subsidence susceptibility mapping in south korea using machine learning algorithms. *Sensors* **2018**, *18*, 2464. [CrossRef]

32. 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]

33. Roy, J.; Saha, S. Assessment of land suitability for the paddy cultivation using analytical hierarchical process (AHP): A study on Hinglo river basin, Eastern India. *Model. Earth Syst. Environ.* **2018**, *4*, 601–618. [CrossRef]

34. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenviron. Disasters* **2019**, *6*, 11. [CrossRef]

35. Roy, J.; Saha, S. GIS-based Gully Erosion Susceptibility Evaluation Using Frequency Ratio, Cosine Amplitude and Logistic Regression Ensembled with fuzzy logic in Hinglo River Basin, India. *Remote Sens. Appl. Soc. Environ.* **2019**, *15*, 100247. [CrossRef]

36. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A Novel Ensemble Approach for Landslide Susceptibility Mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. *Remote Sens.* **2019**, *11*, 2866. [CrossRef]

37. Saha, S. Groundwater potential mapping using analytical hierarchical process: A study on Md. Bazar Block of Birbhum District, West Bengal. *Spat. Inf. Res.* **2017**, *25*, 615–626. [CrossRef]

38. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138. [CrossRef]

39. Paul, G.C.; Saha, S.; Hembram, T.K. Application of the GIS-Based Probabilistic Models for Mapping the Flood Susceptibility in Bansloi Sub-basin of Ganga-Bhagirathi River and Their Comparison. *Remote Sens. Earth Syst. Sci.* **2019**. [CrossRef]

40. Lee, M.J.; Choi, J.W.; Oh, H.J.; Won, J.S.; Park, I.; Lee, S. Ensemble based landslide susceptibility maps in Jinbu area. Korea. *Environ. Earth. Sci.* **2012**, *67*, 23–37. [CrossRef]

41. Arabameri, A. Application of the Analytic Hierarchy Process (AHP) for locating fire stations: Case study Maku City. *Merit Res. J. Art Soc. Sci. Humanit.* **2014**, *2*, 1–10.

42. Arabameri, A.; Ramesht, M.H. Site Selection of Landfill with emphasis on Hydrogeomorphological–environmental parameters Shahrood-Bastam watershed. *Sci. J. Manag. Syst.* **2017**, *16*, 55–80.

43. Arabameri, A. Zoning Mashhad Watershed for artificial recharge of underground aquifers using topsis model and GIS technique. *Glob. J. Hum. Soc. Sci. B Geogr. Geo Sci. Environ. Disaster Manag.* **2014**, *14*, 45–53.

44. Arabameri, A.; Roy, J.; Saha, S.; Blaschke, T.; Ghorbanzadeh, O.; Tien Bui, D. Application of Probabilistic and Machine Learning Models for Groundwater Potentiality Mapping in Damghan Sedimentary Plain, Iran. *Remote Sens.* **2019**, *11*, 3015. [CrossRef]

45. Tien Bui, D.; Le, K.T.; Nguyen, V.; Le, H.; Revhaug, I. Tropical forest fire susceptibility mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, using GIS-based Kernel logistic regression. *Remote Sens.* **2016**, *8*, 347. [CrossRef]

46. Saadat, H.; Bonnell, R.; Sharifi, F.; Mehuys, G.; Namdar, M.; Ale-Ebrahim, S. Landform classification from a digital elevation model and satellite imagery. *Geomorphology* **2008**, *100*, 453–464. [CrossRef]

47. Thai Pham, B.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Talebpour Asl, D.; Bin Ahmad, B.; Kim Quoc, N.; Lee, S. Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms. *Sustainability* **2019**, *11*, 4386. [CrossRef]

48. Pham, B.T.; Prakash, I. Evaluation and comparison of LogitBoost Ensemble, Fisher's Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto Int.* **2017**, *3*, 316–333. [CrossRef]

49. Freund, Y.; Mason, L. *The Alternating Decision Tree Learning Algorithm*; ICML: New York, NY, USA, 1999; pp. 124–133.

50. IRIMO. Summary Reports of Iran's Extreme Climatic Events. Ministry of Roads and Urban Development, Iran Meteorological Organization. Available online: www.cri.ac.ir (accessed on 28 August 2018).

51. Azari, M.; Saghafian, B.; Moradi, H.R.; Faramarzi, M. Effectiveness of Soil and Water Conservation Practices Under Climate Change in the Gorganroud Basin, Iran. *Clean Soil Air Water* **2017**, *45*, 1700288. [CrossRef]

52. Shahpasandzadeh, M. *Seismology and Seismotectonics of Golestan Province, Northeast Iran*; International Institute Seismology and Earthquake Engineering, Seismology Research Institute of the Seismic Group: Tehran, Iran, 2004; p. 8. (In Persian)

53. Lar Consulting Engineering. *The Study on Flood and Debris Flow in the Golestan Province, Regional Water Board in Golestan*; Ministry of Energy: Tehran, Iran, 2007.

54. Jenks, G.F. The Data Model Concept in Statistical Mapping. *Int. Yearb. Cartogr.* **1967**, *7*, 186–190.

55. McMaster, R. In Memoriam: George F. Jenks (1916–1996). *Cartogr. Geogr. Inf. Sci.* **1997**, *24*, 56–59. [CrossRef]

56. Yilmaz, C.; Topal, T.; Suzen, M.L. GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). *Environ. Earth Sci.* **2012**, *65*, 2161–2178. [CrossRef]

57. Van Westen, C.J.; van Asch, T.W.J.; Soeters, R. Landslide hazard and risk zonation—Why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [CrossRef]

58. Youssef, A.M.; Maerz, N.H.; Hassan, A.M. Remote sensing applications to geological problems in Egypt: Case study, slope instability investigation, Sharm El-Sheikh/Ras- Nasrani Area, Southern Sinai. *Landslides* **2009**, *6*, 353–360. [CrossRef]

59. Iranian Landslide Working Party (ILWP). *Iranian Landslides List*; Forest, Rangeland and Watershed Association: Tehran, Iran, 2007; p. 60.

60. Forestry, Rangeland and Watershed Organization (FRWO). List of Landslides in the Iran; Study Group on Landslides, Office of Engineering and Design Evaluation: 2013. Available online: http://www.frw.org.ir/02/Fa/default.aspx (accessed on 2 February 2020).

61. Arabameri, A.; Pradhan, B.; Rezaei, K.; Lee, S.; Sohrabi, M. An ensemble model for landslide susceptibility mapping in a forested area. *Geocarto Int.* **2019**. [CrossRef]

62. Chung, C.-J.F.; Fabbri, A.G. Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards* **2003**, *30*, 451–472. [CrossRef]

63. Marchesini, I.; Ardizzone, F.; Alvioli, M.; Rossi, M.; Guzzetti, F. Non-susceptible landslide areas in Italy and in the Mediterranean region. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 2215–2231. [CrossRef]

64. Frattini, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 66–72. [CrossRef]

65. Li, Z.; Zhu, Q.; Gold, C. *Digital Terrain Modeling: Principles and Methodology*; CRC Press: Boca Raton, FL, USA, 2005.

66. Wentworth, C.K. A simplified method of determining the average slope of land surfaces. *Am. J. Sci.* **1930**, *117*, 184–194. [CrossRef]

67. Zevenbergen, L.W.; Thorne, C.R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* **1987**, *12*, 47–56. [CrossRef]

68. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [CrossRef]

69. Gallant, J.C.; Wilson, J.P. Primary topographic attributes. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; Wiley: New York, NY, USA, 2000; pp. 51–85.

70. Wischmeier, W.H.; Smith, D.D. *Predicting Rainfall Erosion Losses—A Guide to Conservation Planning*; Agriculture Handbook No. 537; US Department of Agriculture Science and Education Administration: Washington, DC, USA, 1978; p. 163.

71. Kiss, R. Determination of drainage network in digital elevation model. *Util. Limit. J. Hung. Geomath.* **2004**, *2*, 16–29.

72. Ay, N.; Amari, S.-I. A Novel Approach to Canonical Divergences within Information Geometry. *Entropy* **2015**, *17*, 8111–8129. [CrossRef]

73. Anderson, C.G.; Maxwell, D.C. *Starting a Digitization Center*; Elsevier: Amsterdam, The Netherlands, 2004; ISBN 978-1843340737.

74. Bayraktar, H.; Turalioglu, S. A Kriging-based approach for locating a sampling site—In the assessment of air quality. *Stoch. Environ. Res. Risk Assess.* **2005**, *19*, 301–305. [CrossRef]

75. Myung, I.J. Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.* **2003**, *47*, 90–100. [CrossRef]

76. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [CrossRef]

77. Pradhan, B.; Seeni, M.I.; Nampak, H. Integration of LiDAR and QuickBird data for automatic landslide detection using object-based analysis and random forests. In *Laser Scanning Applications in Landslide Assessment*; Pradhan, B., Ed.; Springer: Cham, Switzerland, 2017. [CrossRef]

78. Cama, M.; Conoscenti, C.; Lombardo, L.; Rotigliano, E. Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: A test in the Giampilieri catchment (Sicily, Italy). *Environ. Earth Sci.* **2016**, *75*, 238. [CrossRef]

79. Du, G.; Zhang, Y.; Iqbal, J. 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. [CrossRef]

80. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

81. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman & Hall: New York, NY, USA, 1984.

82. Hansen, L.; Salamon, P. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 993–1001. [CrossRef]

83. Breiman, L.; Cutler, A. Available online: http://www.stat.berkeley.edu/users/Breiman/RandomForests/ccpapers.html (accessed on 28 August 2018).

84. Micheletti, N.; Foresti, L.; Robert, S.; Leuenberger, M.; Pedrazzini, A.; Jaboyedoff, M.; Kanevski, M. Machine learning feature selection methods for landslide susceptibility mapping. *Math. Geosci.* **2014**, *46*, 33–57. [CrossRef]

85. Calle, M.L.; Urrea, V. Letter to the Editor: Stability of random forest importance measures. *Brief. Bioinform.* **2010**, *12*, 86–89. [CrossRef]

86. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **2000**, *40*, 139–157. [CrossRef]

87. Jolicoeur, P. Fisher_s linear discriminant function. In *Introduction to Biometry*; Springer: Berlin, Germany, 1999; pp. 303–308.

88. Gilbert, E.S. The effect of unequal variance-covariance matrices on Fisher_s linear discriminant function. *Biometrics* **1969**, *25*, 505–515. [CrossRef]

89. Yin, H.; Fu, P.; Meng, S. Sampled FLDA for face recognition with single training image per person. *Neuro Comput.* **2006**, *69*, 2443–2445. [CrossRef]

90. Robinzonov, N. Advances in Boosting of Temporal and Spatial Models. Ludwig-Maximilians-Universitat München. 2013. Available online: http://edoc.ub.uni-muenchen.de/15338/ (accessed on 28 August 2018).

91. Aertsen, W.; Kint, V.; Van Orshoven, J.; Muys, B. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environ. Model. Softw.* **2011**, *26*, 929–937. [CrossRef]

92. James, G.; Witten, D.; Hastie, T. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013; pp. 856–875.

93. Breiman, L. Arcing Classifiers. *Ann. Stat.* **1998**, *26*, 801–849. [CrossRef]

94. Therneau, T.M.; Atkinson, B.; Ripley, B. RPART: Recursive Partitioning and Regression Trees. *R Package Version* **2014**, *4*, 1–8.

95. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

96. Williams, G.J. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*; Springer: New York, NY, USA, 2011; p. 374. [CrossRef]

97. Wang, Q.; Li, W. A GIS-based comparative evaluation of analytical hierarchy process and frequency ratio models for landslide susceptibility mapping. *Phys. Geogr.* **2017**, *38*, 318–337. [CrossRef]

98. Rahmati, O.; Haghizadeh, A.; Pourghasemi, H.R.; Noormohamadi, F. Gully erosion susceptibility mapping: The role of GIS based bivariate statistical models and their comparison. *Nat. Hazards* **2016**, *82*, 1231–1258. [CrossRef]

99. Oh, H.; Lee, S.; Hong, S.M. Landslide susceptibility assessment using frequency ratio technique with iterative random sampling. *J. Sens.* **2017**, 1–21. [CrossRef]

100. Pradhan, B. An Assessment of the use of an advanced neural network model with Five different training strategies for the preparation of landslide susceptibility maps. *J. Data Sci.* **2011**, *9*, 65–81.

101. Arabameri, A.; Cerda, A.; Rodrigo-Comino, J.; Pradhan, B.; Sohrabi, M.; Blaschke, T.; Tien Bui, D. Proposing a Novel Predictive Technique for Gully Erosion Susceptibility Mapping in Arid and Semi-arid Regions (Iran). *Remote Sens.* **2019**, *11*, 2577. [CrossRef]

102. Arabameri, A.; Chen, W.; Lombardo, L.; Blaschke, T.; Tien Bui, D. Hybrid Computational Intelligence Models for Improvement Gully Erosion Assessment. *Remote Sens.* **2020**, *12*, 140. [CrossRef]

103. Arabameri, A.; Chen, W.; Loche, M.; Zhao, X.; Li, Y.; Lombardo, L.; Cerda, A.; Pradhan, B.; Bui, D.T. Comparison of machine learning models for gully erosion susceptibility mapping. *Geosci. Front.* **2019**, in press. [CrossRef]

104. Arabameri, A.; Cerda, A.; Tiefenbacher, J.P. Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. *Water* **2019**, *11*, 1129. [CrossRef]

105. Arabameri, A.; Chen, W.; Blaschke, T.; Tiefenbacher, J.P.; Pradhan, B.; Tien Bui, D. Gully Head-Cut Distribution Modeling Using Machine Learning Methods—A Case Study of N.W. Iran. *Water* **2020**, *12*, 16. [CrossRef]

106. Arabameri, A.; Blaschke, T.; Pradhan, B.; Pourghasemi, H.R.; Tiefenbacher, J.P.; Bui, D.T. Evaluation of Recent Advanced Soft Computing Techniques for Gully Erosion Susceptibility Mapping: A Comparative Study. *Sensors* **2020**, *20*, 335. [CrossRef]

107. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266. [CrossRef]

108. Oh, H.J.; Kim, Y.S.; Choi, J.K.; Park, E.; Lee, S. GIS mapping of regional probabilistic groundwater potential in the area of Pohang City. Korea. *J. Hydrol.* **2011**, *399*, 158–172. [CrossRef]

109. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and Naïve Bayes models. *Math. Probl. Eng.* **2012**, *2012*, 974638.

110. Oh, H.J.; Lee, S. Assessment of ground subsidence using GIS and the weights-of evidence model. *Eng. Geol.* **2010**, *115*, 36–48. [CrossRef]

111. Corsini, A.; Cervi, F.; Ronchetti, F. Weight of evidence and artificial neural networks for potential groundwater spring mapping: An application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology* **2009**, *111*, 79–87. [CrossRef]

112. Süzen, M.L.; Doyuran, V. A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environ. Geol.* **2004**, *45*, 665–679. [CrossRef]

113. Allouche, O.; Tsoar, A.; Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **2006**, *43*, 1223–1232. [CrossRef]

114. Pradhan, B.; Hagemann, U.; Tehrany, M. An easy to use ArcMap based texture analysis program for extraction of flooded areas from TerraSAR-X satellite image. *Comput. Geosci.* **2013**, *63*, 34–43. [CrossRef]

115. García-Davalillo, J.C.; Herrera, G.; Notti, D.; Strozzi, T.; Álvarez-Fernández, I. DInSAR analysis of ALOS PALSAR images for the assessment of very slow landslides: The Tena Valley case study. *Landslides* **2014**, *11*, 225–246. [CrossRef]

116. Honda, K.; Nakanishi, T.; Haraguchi, M.; Mushiake, N.; Iwasaki, T.; Satoh, H.; Kobori, T.; Yamaguchi, Y. *Application of Exterior Deformation Monitoring of Dams by DInSAR Analysis Using ALOS PALSAR*; The IEEE International Geoscience and Remote Sensing Symposium (IGARSS): Munich, Germany, 2012.

117. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]