remote sensing

MDPI

Article

Novel Credal Decision Tree-Based Ensemble Approaches for Predicting the Landslide Susceptibility

Alireza Arabameri [1], Ebrahim Karimi-Sangchini [2], Subodh Chandra Pal [3], Asish Saha [3], Indrajit Chowdhuri [3], Saro Lee [4,5,*] and Dieu Tien Bui [6]

[1] Department of Geomorphology, Tarbiat Modares University, Tehran 14117-13116, Iran; a.arabameri@modares.ac.ir
[2] Soil Conservation and Watershed Management Research Department, Lorestan Agricultural and Natural Resources Research and Education Center, AREEO, Khorramabad 6815144316, Iran; e.karimi64@gmail.com
[3] Department of Geography, The University of Burdwan, West Bengal 713104, India; scpal@geo.buruniv.ac.in (S.C.P.); asishsaha01@gmail.com (A.S.); indrajitchowdhuri@gmail.com (I.C.)
[4] Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro Yuseong-gu, Daejeon 34132, Korea
[5] Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 34113, Korea
[6] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; buitiendieu@duytan.edu.vn
* Correspondence: leesaro@kigam.re.kr

Received: 16 September 2020; Accepted: 8 October 2020; Published: 16 October 2020

Abstract: Landslides are natural and often quasi-normal threats that destroy natural resources and may lead to a persistent loss of human life. Therefore, the preparation of landslide susceptibility maps is necessary in order to mitigate harmful effects. The key objective of this research is to develop landslide susceptibility maps for the Taleghan basin of Alborz province, Iran, using hybrid Machine Learning (ML) algorithms, i.e., k-fold cross validation and ML techniques of credal decision tree (CDT), Alternative Decision Tree (ADTree), and their ensemble method (CDT-ADTree), which have been state-of-the-art soft computing techniques rarely used in the case of landslide susceptibility assessments. In this study, 22 key landslide causative factors (LCFs) were considered to explore their spatial relationship to landslides, based on local geomorphological and geo-environmental influences. The Random Forest (RF) algorithm was used for the identification of variables importance of different LCFs that are more prone to landslide susceptibility. A receiver operation characteristics (ROC) curve with area under the curve (AUC), accuracy, precision, and robustness index was used to evaluate and compare landslide susceptibility models. The output of the model performance shows that the CDT-ADTree model is the more robust model for the landslide susceptibility where the AUC, accuracy, and precision are 0.981, 0.837, and 0.867, respectively, than the standalone model of CDT and ADTree model. Therefore, it is concluded that the CDT-ADTree ensemble model can be applied as a new promising technique for spatial prediction of the landslide in further studies.

Keywords: landslide susceptibility; CDT-ADTree; k-fold cross validation; robustness index

1. Introduction

Several types of natural hazards and associated disasters such as earthquakes, volcanic eruptions, tsunamis, cloud bursts, soil erosion and so on occur on the Earth's surface and among these, landslides are both catastrophic and recurrent all around the globe [1,2]. A landslide can be defined as the movement of surface covers such as rock, soil, vegetation, and other organic materials downward of slope under the influence of gravity [3]. The major influential factors that are mainly responsible for the occurrences of landslides are rapidly rising population coupled with increasing need for natural resources, industrialization, destruction of forest cover, road construction and so on [4]. In addition to this, the global phenomenon of climate change significantly increases the incidence of landslides as it affects the pattern of rainfall. Every year, landslides damage forests, productive agricultural land, habitat area, network communication, as well as tourist spots. Therefore, landslides have caused enormous losses of life and property through the massive flooding in mountainous and foothill areas, tsunamis in coastal areas, and river pattern changes along with geomorphic and topographical changes [5,6]. Iran has faced various types of natural hazards and disasters, such as intensive soil erosion through gully formation, high-frequency flash floods, devastating land movements (landslides, debris flows) and so on. Therefore, due to their frequent occurrence, landslides and associated economic losses have become nationwide threats to Iran. The Iranian Ministry of National Natural Disaster Reduction Committee's report stated that monetary losses of almost 500 billion Rial were caused by landslides [7]. Basically, the presence of the unique natural characteristics such as physiographic, environmental, and climatic condition are highly prone to landslide activity in the northern part of mountainous regions in Iran [8]. In general, landslides are various types such as debris flow, rock avalanches, slumps and others. In the midst of several types of landslides, slumps are found nearly everywhere in this mountainous region. Debris flows and rock avalanches are nearly exceptional in this region but highly responsible for the loss of life. The present study area of the central Alborz Mountains in the province of Alborz, Iran, has produced a large number of landslides. Therefore, taking into account the devastating damage caused by landslides, it is mandatory to identify the location of a number of landslides that are likely to be mitigated with proper planning. In order to do this, several suitable geo-environmental factors were used to prepare the landslides susceptibility map (LSM) for the prevention and mitigation of hazardous areas.

Therefore, the accurate modeling and trustworthiness of LSMs prediction results largely depend on the accessibility of good quality data, the operational scale, and the appropriate methodology [9]. Thus, qualitative and quantitative methods were extensively applied for preparing LSMs [10]. In general, expert opinion, along with a geomorphological and heuristic approach, is needed for the qualitative analysis of LSMs. On the other hand, quantitative analysis is based on the relationship among several conditioning factors of landslides, which can be expressed by numerical values. Over the last few decades, GIS (geographic information system) has played an important role in the mapping of susceptibility as it is capable of handling vast amounts of spatial data. Therefore, GIS techniques along with several statistical methods have been applied for LSMs such as analytical hierarchical process (AHP) by Pal et al. [11], frequency ratio by Pal and Chowdhuri [12], logistic regression by Tsangaratos et al. [13], and evidential belief function by Pourghasemi and Kerle [14]. In recent times, several machine learning (ML) algorithms have been used to best predict LSMs by using suitable conditioning factors with appropriate ML models. The most notable ML algorithms that have been used frequently are random forest (RF) [15], decision trees (DT) [16], alternating decision tree (ADTree) [17], boosted regression trees (BRT) [15], support vector machine (SVM) [2], artificial neural network (ANN) [18], radial basis function (RBF) [19], grey wolf optimizer (GWO) [20], credal decision tree (CDT) [21] and many more for their better predictive performance with higher accuracy assessment. Recently, several ML algorithms have been merged together which is called an ensemble model, to get better prediction performance results rather than from a single ML algorithm. Extensive literature studies show that the individual ML algorithm is combined with another ML algorithm to provide accurate prediction of LSMs. In addition to this, we can also say that a single ML algorithm can be enhanced by applying ensemble models for perfect LSM prediction [22]. Several research studies have been carried out by different researchers all over the world on LSMs, some notable works using ML and ensemble algorithm are Pham et al. [22], Panahi et al. [23], Nhu et al. [24], and Dou et al. [25].

Thus, the main goal of our present research work is the development of the LSMs in the Taleghan Basin in the central Alborz Mountains in the province of Alborz, Iran. Therefore, in order to carry out our research objective here, we used twenty two appropriate landslide conditioning factors (LCFs) for this particular basin area. In addition, historical data of 188 landslide points, along with a total of 376 landslide points (188 for each landslide and non-landslide points) were randomly selected for further progress in our research work. K-fold cross validation has been applied into four fold groups i.e., Fold 1, 2, 3, and 4 of the randomly selected landslide points for training and validation purposes. In order to achieve our research objective in a suitable manner, we used two popular ML algorithms, the credal decision tree (CDT) [26] and the Alternating Decision Tree (ADTree) [27]. The advantages of using CDT and ADTree ML algorithms are basically for handling of large input model dataset with the proper assessment of prediction accuracy of the model's output. It is important to remember that the accuracy of an ensemble model's susceptibility prediction performance result is much better than a single ML algorithm. Therefore, in this study we applied a novel ensemble approach of CDT and ADTree. Basically, the performance of single CDT and ADTree algorithms has been improved through ensemble of these two ML models. This novel ensemble approach can take the place of a single ML algorithm by improving the predictive accuracy. Therefore, according to the intensive literature survey and to the best of our knowledge, there is currently no research work on the CDT-ADTree ensemble in landslide modeling. Thus, the ensemble approach is the novelty of this research work as the result of this proposed approach has enhanced the prediction accuracy. Finally, three landslide models were validated by statistical analysis of receiver operating characteristics—area under curve (ROC-AUC), overall accuracy (ACC), and precision (PRE) on the dataset. Therefore, based on the newly developed ensemble model of CDT-ADTree, the produced LSM can help the planners for management of environmental degradation and natural resources from fragile losses in this basin area.

2. Materials and Methods

2.1. Study Area

The Taleghan Basin is one of the main sub-basins of Sefidroud, found on the southern slopes of the central Alborz mountains in the province of Alborz, Iran. The present study area of the Teleghan basin is situated between 36°07′–36°36′ N latitudes and 50°59′–51°20′ E longitudes and occupies a region of 969 km^2 (Figure 1). The position of the Taleghan basin (between the two mountain ranges in the north and the south and surrounded by another mountain range in the east) provides a specific climatic characteristic in that area. The mean altitude is 2910 m and the mean slope is roughly 32.4%. According to the Alborz Meteorological Agency's predictions, this basin had a mean annual temperature of 7°–14° C and a mean annual precipitation of 532 mm over a 30-year period [28]. Based on the Emberger model [29,30], the area may be classified as a wet and sub-humid climate.

From a lithological point of view [31], the region can be divided into five general classes including quaternary rocks (young terraces), tertiary glacial sediments (Riss) with a marl sub-layer, uncondensed deposits of Miocene, igneous deposits and metamorphic rocks. The second level is particularly vulnerable to large motions, in particular landslides, owing to the presence of loose marl sub-layers and coarse-grained overhead sediments. The spatial chart of the sampling region and the definition of its lithological unit are shown in Table S1 in section 1 of Supplementary Materials.

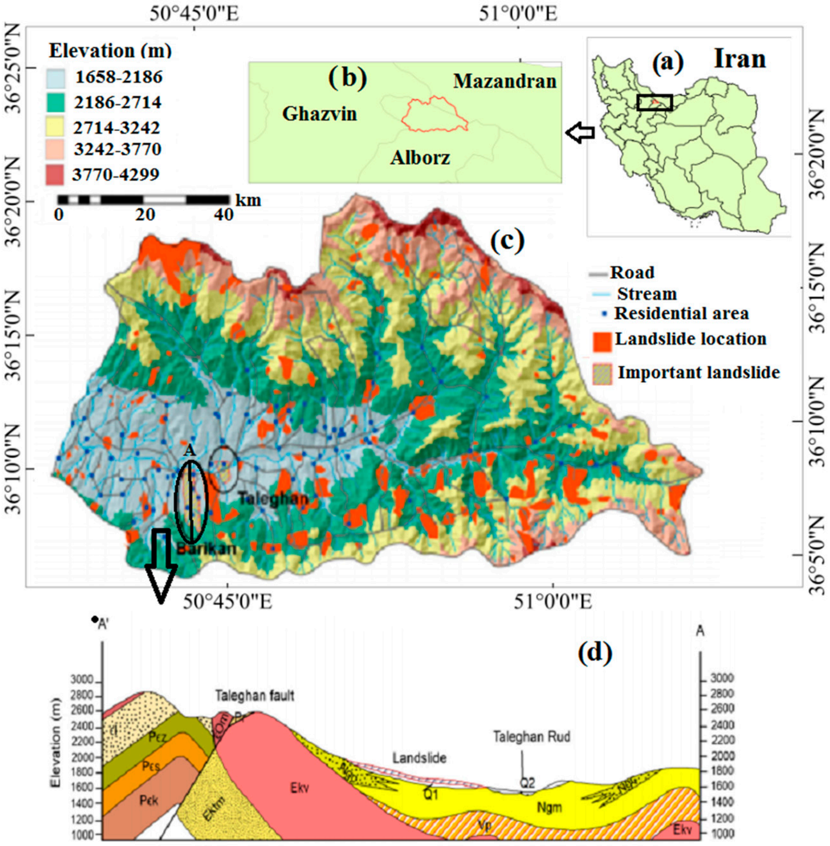

Figure 1. (**a**) Location of the study area in Iran, (**b**) location of the study area in Alborz and Mazandran provinces, (**c**) location of roads and Taleghan city in the study area, and (**d**) geological cross-section of the Barikan landslide.

According to the Soil Conservation and Watershed Research Institute [32], the most important landslide in this area is the Barikan landslide. The traveling mass of 455 ha was of a rotational form. The Barikan landslide was 1100 meters long, with an elevation of 4200 meters. Geologically, the residential areas are built on the Quaternary young soils, but the soils in the southern part of the village are primarily mudstone and siltstone. Several surveys have recorded landslides in numerous villages in the Barakin area [32,33]. The frequent occurrences of landslides in this area are highly responsible for the unique geographical location, surface, and subterranean water holding capacity and exceptional topographical characteristics. As a result, land collapse is an old phenomenon and there is often a reactivated landslide that has had a major impact on settlement areas around Barikan so that settlement houses are almost destroyed every time [32]. In addition, the movement of the river water in the Taleghan and Barikan regions is due to the hydraulic gradient (from the north to the south of the Barikan region), which causes the fracture of the land surface and its displacement in the clay deposits. As a result, land disruption occurs mainly in this region and fragmented settlements are found throughout this Barikan area.

2.2. Methodology

The present research study has been carried out by the following procedure to meet our research objective in an appropriate way (Figure 2).

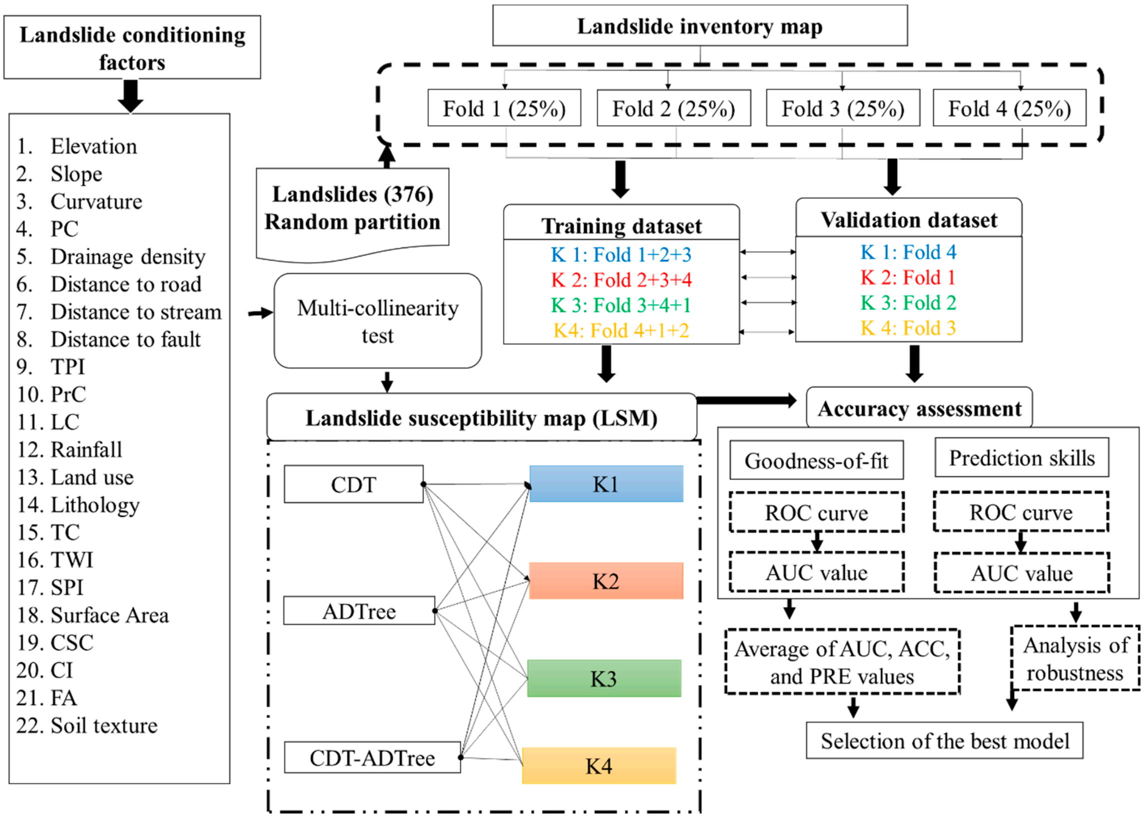

Figure 2. Flowchart of research in the study area.

First of all, a landslide inventory map was prepared based on the historical data of 188 landslide points for the last 21 years, along with a total of 376 randomly selected training and landslide validation points chosen to determine our research outcome. For further progress of this study, a total of 22 landslide conditioning factors (LCFs) namely convergence index (CI), cross-sectional curvature (CSC), curvature, distance to stream (DtS), drainage density (DD), elevation, distance to fault (DtF), distance to road (DtR), flow accumulation (FA), longitudinal curvature (LC), plan curvature (PC), profile curvature (PrC), rainfall, slope, stream power index (SPI), surface area (SA), Tangential curvature (TC), topography position index (TPI), topography wetness index (TWI), soil texture, land use/land cover (LU/LC), and lithology were selected for landslide susceptibility modeling based on the previous research study [27,34] and keeping in view the local topographical, climatological, and hydrological characteristics. After that, to know the relationship among these LCFs, we carried out the multi-collinearity test, which is a statistical procedure. The MC test was performed using two common techniques, i.e., inflation factor variance (VIF) and tolerance (TOL). Thereafter, it was also necessary to know which LCFs had more responsibility for landslide occurrences. To know the most responsible factor for the occurrences of landslides in this region, here we used the Random Forest (RF) algorithm for the identification of several variables importance Afterward, k-fold cross validation and ML modeling of CDT and ADTree along with their novel ensemble of CDT-ADTree was applied for modeling the LSMs. Finally, every model's result was validated through statistical analysis of ROC-AUC, ACC, and PRE methods for better accuracy assessment. The validation and accuracy assessment of the computation of every model's maps was also tested through the robustness method in this research study.

2.3. Landslide Inventory Map

In a landslide vulnerability analysis, the surface of the earth may be separated into two distinct zones, i.e., locations where landslides have already occurred and locations where landslides have not yet occurred but are likely to occur in the future [35]. Landslide modeling basically focused on several natural characteristics, which is a necessary requirement for the execution of construction and preparation of various steps. In this research, an inventory map of landslides was prepared based on the data collected from the Department of Natural Resources and Watershed Protection of Iran. Apart from this, Google Earth satellite images along with intensive field surveys were carried out to validate the inventory map. A total of 188 landslide events were reported in the study region. The total study area occupies an area of 110.87 km^2 and constitutes 11.44 per cent of the overall territory of the two provinces. The zone with the smallest and highest landslides affected area was about 0.005 km^2 and 8.8 km^2, respectively, within this study area. In comparison, the total landslide area of 82% of the landslides was 0.27 km^2. In total, six forms of landslides were reported in this area: translational (42%), slump (17%), rotational (12%), rock slide (10%), unknown (7%), and debris flow (3%).

Landslide inventories used for simulation and digitized as polygon forms from the Google Earth image are to be transferred to point repositories in such a way that simulations can be made [36]. Any methods, such as centroids [37], seed cells [38] and diagnostic fields [39], were widely used to convert landslide polygons into points shape file in the GIS environment. In this research study, we used the centroid method for converting polygon features to point features. For this analysis, first, the area size of the landslides was transformed into a raster layer with a 12.5 m pixel size, and then a sampling point per pixel was created for each landslide. This method, involving complicated measurement techniques, generates a significant number of pixels for wide-area landslides. In order to reduce ambiguity, a subset of points for each large polygon was randomly chosen to reflect internal improvements referring to the broad landslide region [40]. As a result, landslide polygons smaller than the average landslide region (0.27 km^2) were relocated to one sampling point by the method of centroid process and the number of polygon points greater than the average area was determined on the basis of the ratio of landslide area to the average landslide area. The ArcGIS software culminated in 376 points for the complete map of landslides.

It was also necessary to define non-land-sliding zones [39–41]. These areas were selected randomly [39]. In this study, the cross validation (CV) method was used for splitting the landslide inventory dataset into four k-fold groups. The general process of splitting k-folds is the dataset X of the landslide points, which was randomly divided into different k-folds, i.e., $X_1, X_2, \ldots\ldots X_n$. Each k-fold of this study had equal size. Thus, the k-fold model was attained n number of times and one for every time $t \in \{1, 2, \ldots n\}$. In the model t, it was trained with the dataset of X without the subset of X_t, and finally tested with X_t [41]. Thus, the entire dataset in this study was divided into four k-fold groups, i.e., Fold -1, Fold -2, Fold -3, and Fold -4 (Figure 3). Every time, 75% of the data were used for training and the remaining 25% were used for testing the landslide susceptibility map. Several types of factors are responsible for the numbers of folds applied to the study, i.e., length of inventory dataset, complexity, and the techniques used. Intensive studies on the k-fold CV model showed that various approaches have been applied for LSMs such as five-fold CV by Wiens et al. [42], four-fold CV by Václavík and Meentemeyer [43], Arabameri et al. [7], and three-fold CV by Boria et al. [44] and so on. In our study we used four-fold CV to eradicate the downbeat impacts of randomness on presentation of ML techniques to forecast LSMs. In this study area, a number of field photographs of landslide prone regions are shown in Figure 4 from which the pattern of landslides can be understood.

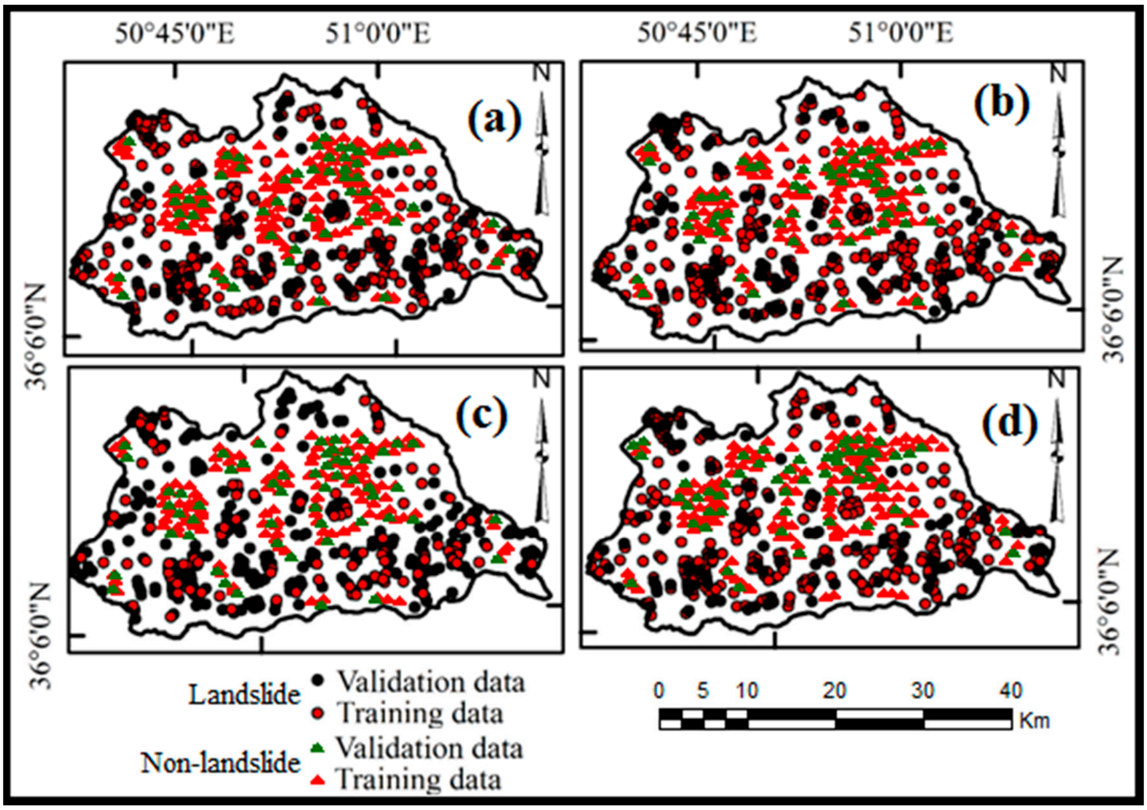

Figure 3. Landslide and non-landslide points of training and testing dataset using the four k-folds system (**a**) k1, (**b**) k2, (**c**) k3, and (**d**) k4.

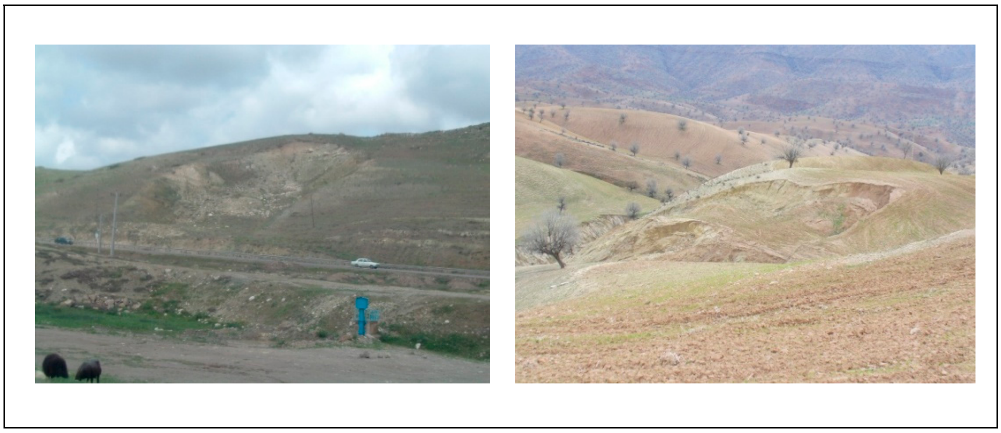

Figure 4. Field photographs show the pattern of landslide occurrences in some part of the present study area.

2.4. Landslide Conditioning Factors (LCFs)

The most important step in predicting landslide susceptibility maps is to choose a number of appropriate conditioning factors. Basically, conditioning factors of LSMs can be separated into several groups, i.e., topographical, climatological, hydrological, geological, and environmental factors. In general, there are no universal regulations for the selection procedure of several landslide conditioning factors [45]. Therefore, selection criterion of several LCFs largely depends on geographical features of an area, types of landslide occurrences, and the analysis procedure of the model [46]. A total of twenty two LCFs were selected to analyze the models based on the local geographical, geomorphological, and spatial characteristics in this research work. These twenty two LCFs are CI, CSC, curvature, distance to streams, drainage density, elevation, distance to faults, distance to roads, FA, LC, PC, PrC, rainfall, slope, stream power index (SPI), SA, TC, topographic positioning index (TPI), topographic wetness index (TWI), soil texture, LULC (land use/land cover), and lithology. All of these factors output maps (Figure S1a–v) have been provided in the Supplementary Materials of section 2. Selection of conditioning factors depends on the types of occurred landslide and geographical features. Therefore, in this study, all of these factors were selected based on the local topographical, climatological, and hydrological characteristics. Here, we tried to select the maximum number of LCFs for better analysis as the maximum number of factors gives accuracy on the model's analysis procedure. In addition to this, we have also tried to know which factors are more important for landslide occurrences than others. Thus, based on the aforementioned conditions we selected 22 factors based on the MC analysis. To prepare all of these LCFs, various data were collected from different sources, i.e., a 12.5 m resolution ALOSPALSAR DEM was downloaded from Alaska Satellite Facility (ASF) website for extraction of terrain and hydrological related factors. The lithology map was prepared from the geological map of this region, which was collected from the Geological Society of Iran (GSI) (http://www.gsi.ir/) at a scale of 1:100,000. The land use and road network maps were prepared from the topographic map collected from National Geographic Organization of Iran (www.ngo-org.ir) at a scale of 1:1:50,000. In addition to the topographic map, Google Earth and Landsat 8 satellite images were also used to verify land use and road maps. Some of the LCFs have been described in the following section.

Curvature generally indicates the rate of slope/gradient changes in a specific direction [47]. Curvature is classified into two types, i.e., plan and profile curvature. Both curvature types represent the topographic characteristics of an area and indicate different topographical perspectives through quantitative indices and analysis. Therefore, curvature and its types are important factors for modeling of LSMs. Several linear factors, i.e., distance to streams, roads, and faults are also considered most vital LCFs for mapping landslide susceptibility. In general, in the high mountainous region, the closer stream areas are more vulnerable to landslide occurrences. This is due to wetting capacity of the surface area from stream water and instability characteristics of the surrounding regions. Additionally, the stability of the slope is also responsible for stream erosion at the foot of the slope and thus this factor has an advantage on LSM. In the mountainous areas, steep slopes and road networks are very much prone to high runoff rate and this phenomenon triggers the landslide occurrences in a devastating way. Furthermore, during the construction of roads in the hilly areas land stability has been highly affected and as a consequence, land movements occur. The road map of the present study area has been collected from National Organization of Iran at a scale of 1:100.000 [7]. Distance to fault is also an important factor of landslide susceptibility mapping, which highly triggers the land movements. Faults are basically a gap in the rock surface and these features are unstable rock surface, which is basically instability in nature, and much prone to land movements [48]. All of these linear factors have been measured by using Euclidean distance buffering in the GIS platform. Drainage density is defined as total length of the all streams in a particular area. Slope instability of an area increases with higher drainage density and vice versa. Thus, the occurrence of landslides also depends on collective impacts of drainage density. The following equation has been used to calculate the drainage density [49]:

$$DD = \frac{\sum_{i=1}^{n} S_i}{a} \tag{1}$$

where, $\sum\limits_{i=1}^{n} S_i$ represents cumulative length of all streams, S in km, and 'a' is the cumulative area of watershed in km^2. Elevation is highly correlated with landslides and one of the major parameters used in LSM. It mainly impacts on vegetation cover, topographic characteristics, and human activity on a land surface. Depending on this impact, factors elevation also influences stability of the slope [50]. The pattern of rainfall-runoff is associated with rainfall. The intensity and duration of the rainfall influences the infiltration capacity of the pore spaces and increases the gravity of the rock materials due to the wetting capacity. It is a fact that heavy rainfall over a long period of time causes a high frequency of landslide occurrences. Slope is the most vital parameter for LSM due to the characteristics of its stability. Higher slope values are prone to landslide occurrences and vice versa. TPI is the measurement of slope position of a topographic feature. In general, TPI is used to differentiate between the central point elevation and the average elevation around the central point [51]. Positive and negative TPI values indicate the locations that are higher and lower than the average surrounding areas, respectively, whereas zero TPI value indicates that the area is flat or a constant slope. The TPI has been calculated using the following equation [52]:

$$TPI = M_0 - \frac{\sum_{n-1} M_n}{n} \tag{2}$$

where, M_0 is the elevation of the middle point, M_n is the elevation of the grid, and n is the total number of pixels in a neighborhood region of the digital elevation model (DEM) raster file.

In general, the erosional capacity of a stream is computed by SPI. In the case of LSM, it is essential to know the erosion capacity of the stream as it is highly responsible for the erosion of the riverside and the impact of the slope by subsidence or retreats. Therefore, SPI takes LSM into consideration. If the SPI value is high, then erosional capability of a slope surface is also high and vice versa. The value of SPI was determined by following equation [53]:

$$SPI = A_s * tan\beta \tag{3}$$

where, A_S indicates upslope contributing area and β indicates the slope angle. TWI analysis provides a proxy for soil moisture [54]. Therefore, TWI is essential for LSM as it largely depends on the moisture condition of land surface area where landslides are taking place. In this study, TWI was calculated by following equation [55]:

$$TWI = In\left(\frac{A_s}{tan\beta}\right) \tag{4}$$

Soil texture also determined the pattern of landslide and its frequency as it responds to rainfall, porosity, hardness, vegetation cover, etc. In the present study, four types of soil texture were recognized i.e., clay, clay loam, loam, and sandy loam. Like soil texture, LULC also play an important role in the occurrence of landslides. This is due to several land covers having diverse characteristics along with different slope stability. Therefore, LULC is considered one of the crucial factors for landslide assessment. Seven types of LULC were recognized in this study area and these are good range land, agri orchard, moderate rangeland, poor rangeland, urban, agriculture, and orchard. Lastly, lithological characteristics of an area directly impacted on landslide phenomenon. The characteristics of rock mass depend on several lithological units which immensely trigger landslide activities [56]. The lithological map of the present study area was collected from the Geological Society of Iran (GSI). Various lithological units and their respective descriptions have been given in Table S1 to Section 1 of Supplementary Materials.

2.5. Multi-Collinearity (MC) Analysis

Multi-collinearity analysis refers to the linear relationship among two or more variables in a dataset and it indicates lack of orthogonality among the variables [57]. In a regression model, when two or more independent variables are correlated among each other then MC appears. In general, MC is a statistical technique in which near-linear dependence variables are correlated in a regression model. It is a fact that if there is absence of linear relationship among the variables then it is called orthogonal [58]. In mathematical terms, it can be defined as when vectors k recline in a subspace of dimension lower than k [57]. In the midst of predictor variables, a little bit of MC will highly create a big problem in the whole dataset and cause errors in the prediction output. Therefore, it is necessary to check MC analysis among the several LCFs for better prediction assessment. Several literature studies show that there are present various techniques to evaluate the MC, among them most popular are Pearson's correlation coefficients, variance inflation factors (VIF), and tolerances (TOL). [59,60]. In the present research study, two MC test methods of VIF and TOL were used to assessment MC among the different LCFs. The MC problems occur when the threshold value of VIF is >5 and TOL is <0.1. To calculate the VIF and TOL of MC, the following equation was used [61]:

$$TOL = 1 - R_j^2 \tag{5}$$

$$VIF = \frac{1}{TOL} \tag{6}$$

where, R_j^2 represent coefficient of multiple determination of j on the predictor variables.

2.6. Measuring the Importance of LCFs by Random Forest (RF)

The algorithm of RF is an ensemble-based machine learning classifier that was primarily introduced by Breiman [62]. The algorithm of RF is based on a non-parametric multivariate statistical method. The mechanism of RF is based on the construction of numerous decision trees at the initial or training phase through the combined action of bagging and random selection of the variables [62]. The method of bagging approach is that from the training dataset, several trees are substituted all the way through the subset in a RF algorithm. Therefore, it indicates that the same sample receives chances several times and others may possibly not receive chances at all [63]. The application of RF has wide perspective in classification, regression along with unsupervised learning. It is essential that two parameters produce the RF classifier trees and these parameters are *Ntree* i.e., number of decision trees and *Mtry* i.e., number of variables. Any kind of assumption does not necessitate establishing the relationship among explanatory and response variables in a RF algorithm. Thus, this model is very suitable for better prediction of new data cases. Several predictor variables' optimal condition is specified by following equation [64]:

$$log2(M+1) \tag{7}$$

where, M represents the input algorithm numbers and the mean square error (ε) of this model is represented by following equation:

$$\varepsilon = \left(v_{observed} - v_{response}\right)^2 \tag{8}$$

where, $v_{observed}$ indicates variable from observed data and $v_{response}$ indicates variable from the output result. Finally, the prediction of the model can be calculated by:

$$S = \frac{1}{K} \sum K^{th} v_{response} \tag{9}$$

where, S is the prediction result and K indicates individual trees in the RF model.

Importance of variables selection is a significant task and it is more vital in any kind of statistical analysis where a large number of LCFs are used. The algorithm of RF ML model can help and is used to identify several variables importance in a definite and accurate way. Here, we used the mean decrease accuracy (MDA) index for identifying the importance of variables by applying the RF model. Basically, the MDA index permutes out-of-bag (OOB) samples from the RF ML model and helps to determine the most important variables from the several conditioning variables in a dataset. The following equation has been used to compute the variables importance:

$$VI_j = \frac{1}{ntree} \sum_{t=1}^{ntree} EP_{tj} - E_{tj} \tag{10}$$

where, E_{tj} indicates the OOB error on tree t before permuting the values of X_j and EP_{tj} indicates the OOB error on tree t after permuting the values of X_j.

2.7. Methods for Landslide Susceptibility Assessment

2.7.1. Credal Decision Tree (CDT)

Credal set is a set of probability distribution or measurement in a given dataset. Basically, this credal set is proposed to express uncertainty or ambiguity about the probability of a dataset in the credal model. The credal set is also used to suggest the beliefs of Bayesian variables regarding the possible outcomes of the variables. The algorithm of CDT is principally based on split criterion processes and developed by Abellán and Moral [65]. In general, to solve the classification problems by using credal sets, CDT algorithm was proposed very first to do so [66]. Furthermore, the algorithm of CDT was created on uncertainty measures along with inaccurate probabilistic phenomena based on the credal set [65]. Keeping in view the avoidance production of difficult decision trees, a new idea was born during the construction of CDT algorithm. Thus, the new idea for CDT was developed to summarize the uncertainty measures basically for splitting decision trees and solved the classification problems [21]. Based on the two theories given by Dempster [67] and Shafer [68], along with the new idea which was developed during the development of CDT, these were extensively used to analyze the uncertainty measures of credal datasets. Abellán and Moral [65] measured the uncertainty of credal datasets by using the following equation:

$$Eu(x) = NG(x) + RG(x) \tag{11}$$

where, Eu represents the measure of uncertainty value, x represents the credal sets on the frame X, NG represents the function of non-specificity and RG represents the function of randomness in a credal dataset.

The basic arithmetical function for the CDT model may be described as follows, taking into consideration that variable Z is found among the values of $\{Z_1, \ldots, Z_k\}$. Thereafter, in a given dataset the distribution of probability i.e., $p\left(Z_j\right),\ j = 1, \ldots, k$ is explicit for every value of Z_j [66].

$$p\left(Z_j\right) \in \left[\frac{n_{zj}}{N+s}, \frac{n_{zj}+s}{N+s}\right], j = 1, \ldots, k \tag{12}$$

where, N represents the sample dataset size, n_{zj} represents frequency of event i.e., ($Z = Z_j$) and s represents the hyper-parameter and the value ranges from 1 to 2 [69].

On the basis of the above representation, a new kind of credal set on the variable Z, $K(Z)$ may be defined as:

$$k(z) = \left\{p|p\left(z_j\right) \in \left[\frac{n_{zj}}{N+s}, \frac{n_{zj}+s}{N+s}\right], j = 1, \ldots, k\right\} \tag{13}$$

2.7.2. Alternating Decision Trees (ADTree)

ADTree is a modernized version of the decision tree. It can generate a very precise classifier with the combination of boosting and decision trees algorithm [70]. The advantages of the ADTree algorithm are that it produces a lower number of nodes, is easily explainable, and creates a classification margin through confidence of measure. It also has the capability to identify and remove the gaps in the midst of boosting and decision trees algorithm. In addition to this with the help of AdaBoost algorithm, this model uses a smaller number of iterations in its function [70]. In general, this model consists of two layers i.e., one layer is decision or splitting nodes and the another one is option or prediction nodes [17]. Among these two layers, the situation is expressed and evaluated by decision nodes and prediction nodes consisting of numerical values in the model. The prediction nodes in ADTree algorithm present in leaves as well as roots if there is no connection among the additional decision nodes. The tree stem of ADTree initially searches the invariable prediction coefficient among the training dataset. Thereafter, by using boosting, the algorithm tree can grow based on reiteration of data and an added new rule. By following this, one decision node and two prediction nodes are produced [70]. Subsequently, a final prediction score is assigned with the contribution of weight and the summation of all these weights gives the prediction probability [71]. In addition to this, ADTree has the capability to make classes binary within the model for better analysis of the training and testing dataset. The algorithm of ADTree has been computed in different phases and presented as follows [24,72]:

Initial phase:

Let us consider the precondition and condition of the base chief rule of R_t and the early prediction value can be calculated by:

$$a = \frac{1}{2} In \frac{W_+(T)}{W_-(T)} \tag{14}$$

where, $W_+(T)$ and $W_-(T)$ represent the total positive and negative weights, respectively, and they also validate the training dataset.

Pre-adjustment phase:

Here, $Z_t(c_1, c_2)$ may be defined by precondition c_1 and condition c_2. In this phase, conjunction (AND) and negation (NOT) is denoted by $\wedge$ and $\neg$ respectively.

$$Z_t(c_1, c_2) = 2(\sqrt{W_+(c_1 \wedge c_2) W_-(c_1 \wedge c_2)} + \sqrt{W_+(c_1 \neg c_2) W_-(c_1 \neg c_2)} + W\,(\neg c_2) \tag{15}$$

Thereafter, by using formula 13, two prediction values 'a' as well as 'b' are assigned, and to minimize $Z_t(c_1, c_2)$ optimized of c_1 and c_2 are selected. Using R_t rules, run R_{t+1} and R_t so that c_1 (precondition) and c_2 (condition) should be equal.

$$a = \frac{1}{2} In \frac{W_+(c_1 \wedge c_2)}{W_-(c_1 \wedge c_2)},\ b = \frac{1}{2} In \frac{W_+(c_1 \wedge \neg c_2)}{W_-(c_1 \wedge \neg c_2)} \tag{16}$$

Afterward, update the weights for each repetition based on the following equation.

$$W_{i,t+1=} W_{i,t} e^{-r,t(x_i) y_t} \tag{17}$$

Output phase:

Finally, classification rule is obtained through summation of all weights and chief rule R_{t+1}:

$$class(x) = sign\left\{ \sum_{t=1}^{T} r_t(x) \right\} \tag{18}$$

2.7.3. Ensemble of CDT and ADTree

Ensemble models have advantages for their novelty and ability of comprehensive analysis of any kind of natural hazard susceptibility mapping [73]. In general, ensemble models are applied for high precision and predictive analysis for hazard-related susceptibility mapping. In another words, production of ML models are considerably improved by using ensemble models [22]. Thus, in this study we ensemble CDT and ADTree models to significantly reduce the limitation of these individual models. Literature studies show that ensemble method has been used in different fields to get maximum accuracy and predictive ability such as for gully erosion potentiality [74], flood susceptibility mapping [75] and so on. Therefore, in this study we also used a novel approach by ensemble of two single ML algorithms i.e., CDT and ADTree. The ensemble method of these two ML algorithms was performed in 'R' programming language. In general, statistical computing and graphics software are freely available and have been widely used by a number of researchers around the world. The ensemble of these two models is highly optimal regarding the prediction of landslide susceptibility assessment.

2.8. Validation and Accuracy Assessment

The predictive performance of different ML models in this study was carried out by accuracy (ACC) assessment, precision (PRE) analysis, and computing the values of area under curve (AUC) of receiver operating characteristics (ROC). In addition to the aforementioned method, it was also necessary to estimate model robustness for better predictive performance.

2.8.1. Accuracy (ACC) Assessment and Precision (PRE) Analysis

The ACC is a technique that gives the accurate prediction result as a method with the combination of hazard and non-hazard area in this study landslide and non-landslide area. The ACC measurement of a dataset is carried out by using four possible indices. These four indices are namely true positive (TP), true negative (TN), false positive (FP) and false negative (FN). In which, TP correctly indicates landslide pixels and FP indicates non-landslide pixels. On the other side, TN incorrectly indicates landslide pixels and TP indicates non-landslide pixels [5]. The ACC and PRE were calculated by using following equation [27]:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$PRE = \frac{TP}{(TP + FP)} \tag{20}$$

2.8.2. Receiver Operating Characteristics (ROC) Curve

The area under curve (AUC) of ROC is the success of a model to perfectly predict the occurrence and non-occurrence of a phenomenon; here it is landslides and non-landslides. In general, ROC is a two-dimensional diagram on the Y and X axis and represents sensitivity (true positive) and 1-specificity, respectively (false positive). The sensitivity and specificity have been correctly classified as susceptible and non-susceptible [76]. The predictability of a model was estimated through AUC analysis. Model's effectiveness was analyzed through training dataset and prediction ability was analyzed by validation dataset [46]. The advantage of ROC analysis is that the model performance is assessed at a range of thresholds. The value of AUC ranges from 0.5 to 1, in which higher values represent very good prediction and lower values represent a weaker prediction by the respective model. Furthermore, according to Yesilnacar [77], AUC-ROC can be classified into five categories i.e., 0.5–0.6 (poor), 0.6–0.7 (moderate), 0.7–0.8 (good), 0.8–0.9 (very good), and 0.9–1 (excellent). The value of AUC-ROC was calculated by using the following equation [74]:

$$A_{AUC} = \sum_{k=1}^{n} (X_{k+1} - X_k)\left(A_k + 1 - A_{k+1} - \frac{A_k}{2}\right) \tag{21}$$

where, A_{AUC} represents the area under curve, X_k indicates 1-specificity, A_k indicates sensitivity, and n is the number of thresholds at which sensitivity and specificity are classified in the ROC model.

2.8.3. Robustness Test

The capacity of a specific model to retain efficiency when small changes to the input take place is known as model robustness [78,79]. In general, robustness of a model may perhaps be deliberated through differentiating between the maximum and minimum evaluation criteria's accuracy result [80]. Therefore in this study, robustness analysis was calculated for ACC, PRE, and AUC from the four k-folds analysis to evaluate the models in the following way [81]:

$$R_{ACC} = ACC_{max} - ACC_{min} \tag{22}$$

$$R_{\mathrm{PRE}} = \mathrm{PRE}_{max} - \mathrm{PRE}_{min} \tag{23}$$

$$R_{ROC\_AUC} = ROC\_AUC_{max} - ROC\_AUC_{min} \tag{24}$$

where, R_{ACC}, R_{PRE} and R_{ROC_AUC} is robustness of accuracy, precision, and receiver operating characteristics—area under curve method, respectively. ACC_{max}, PRE_{max}, and ROC_AUC_{max} indicate maximum value of three evaluation criteria, similarly ACC_{min}, PRE_{min}, and ROC_AUC_{min} indicate minimum value of three evaluation criteria used in this study.

3. Results

3.1. Multi-Collinearity (MC) Analysis

The extent to which geo-environmental variables of landslides are interrelated to each other by linear relationship is called the MC effect of explanatory variables. MC exists individually based on the relationship with several conditioning variables and some dependent variables. The high linear relationship among the conditioning variables of landslide occurrence has resulted in the difficulty of estimating the model parameter. In addition to this, high MC also have inconsistent projections of dependent variables e.g., landslides, which are not suitable for future predictions of landslide occurrences. MC has been measured considering the tolerance (TOL) and variance inflation factor (VIF) methods and these techniques help to analyze each explanatory variable of landslide for better performance of the landslide susceptibility models. A TOL value below 0.2 and VIF above 10 indicate the MC problems among the variables [82]. When TOL and VIF of explanatory variables are above 0.2 and below 10, respectively, this indicates of no MC between the variables. The TOL and VIF of variables of landslide susceptibility in different k-fold CV landslide classification methods (K1 to K4) were calculated and shown in Table 1. The result shows that there are no issues with collinearity problems among the 21 landslide causative factors in differential k-fold CV methods for dividing of landslides.

Table 1. Multi-collinearity test among factors.

Factors	Collinearity Test (K1)		Collinearity Test (K2)		Collinearity Test (K3)		Collinearity Test (K4)	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Elevation	0.804	1.244	0.880	1.136	0.892	1.121	0.776	1.289
LC	0.859	1.164	0.916	1.092	0.908	1.101	0.852	1.174
Rainfall	0.681	1.468	0.728	1.374	0.717	1.395	0.676	1.479
DtF	0.489	2.045	0.524	1.908	0.524	1.908	0.504	1.984
SA	0.220	4.545	0.257	3.891	0.311	3.215	0.201	4.975
DD	0.351	2.849	0.398	2.513	0.388	2.577	0.358	2.793
CSC	0.235	4.255	0.265	3.774	0.274	3.650	0.234	4.274
DtS	0.499	2.004	0.520	1.923	0.551	1.815	0.492	2.033
Slope	0.634	1.577	0.694	1.441	0.682	1.466	0.643	1.555
DtR	0.594	1.684	0.692	1.445	0.635	1.575	0.614	1.629
FA	0.718	1.393	0.758	1.319	0.783	1.277	0.710	1.408
LULC	0.213	4.695	0.298	3.356	0.329	3.040	0.340	2.941
PC	0.266	3.759	0.316	3.165	0.316	3.165	0.281	3.559
PrC	0.270	3.704	0.263	3.802	0.339	2.950	0.309	3.236
CI	0.262	3.817	0.282	3.546	0.287	3.484	0.258	3.876
lithology	0.211	4.739	0.247	4.049	0.350	2.857	0.308	3.247
SPI	0.372	2.688	0.412	2.427	0.309	3.236	0.281	3.559
TWI	0.286	3.497	0.325	3.077	0.227	4.405	0.197	5.076
TC	0.206	4.854	0.264	3.788	0.244	4.098	0.214	4.673
TPI	0.333	3.003	0.374	2.674	0.365	2.740	0.347	2.882
Soil Texture	0.343	2.915	0.282	3.546	0.279	3.584	0.351	2.849

3.2. Application of the Models to Landslide Susceptibility Mapping

Figure 5a–l show landslide susceptibility maps produced by CDT, ADTree, and CDT-ADTree model for four different k-fold CV landslides classification methods (K1 to K4 CV). The mentioned maps were classified into five categories of very low, low, moderate, high, and very high using the quintile raster classification methods in the ArcGIS platform. The optimism CDT, ADTree, and an ensemble of CDT-ADTree are associated with better prediction capability than any other resample method. The percentage of area in five different susceptibility zones in the CDT, ADTree, and an ensemble of CDT-ADTree model are almost same in all k-fold CV classification methods. Different landslide susceptibility scenarios in the three aforementioned models were shown in Figure 6. In the CDT-ADTree ensemble model, it was found that most of the area is associated with moderate susceptibility zone (36.09%), whereas the remaining part of the study area was related to very low (33.13%), low (17.47%), very high (7.08%), and high (6.23%) susceptibility zones in the K1-fold CV landslide distribution method. However, for the other two landslide susceptibility models i.e., ADTree and CDT, the landslide susceptibility zone is drastically different from the CDT-ADTree ensemble model. In ADTree K1 method, most of the area associated with high (26.16%) susceptibility zone and 13.92 per cent area belongs to very low to high susceptibility zones. In the case of the CDT K1 method, most of the area was associated with moderate (27.43%) and low (25.75%) landslide susceptibility zone, where the rest of the basin is distributed in a high (23.55%), very high (12.0%), and very low (11.28%) landslide susceptibility zones. In the K2, K3, and K4-fold CV landslide distribution method, the above models presented more or less the same scenario as mentioned for K1-fold CV of susceptibility area and their spatial distribution (Figure 6).

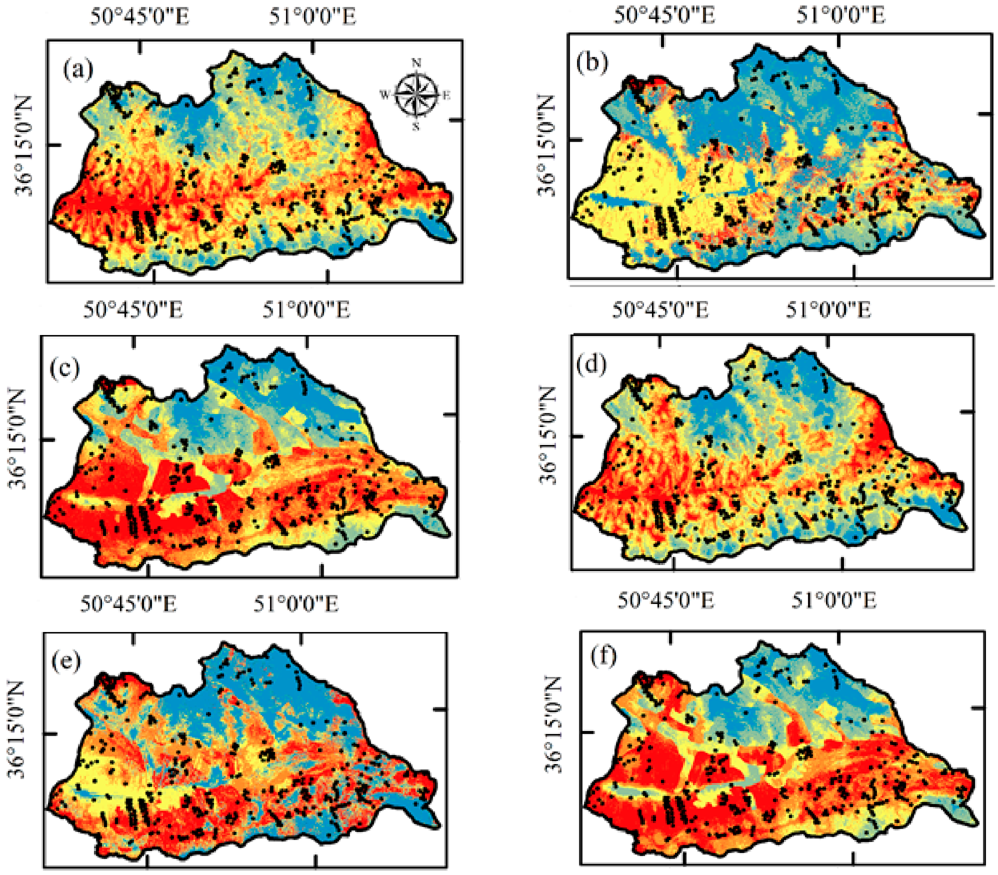

Figure 5. *Cont.*

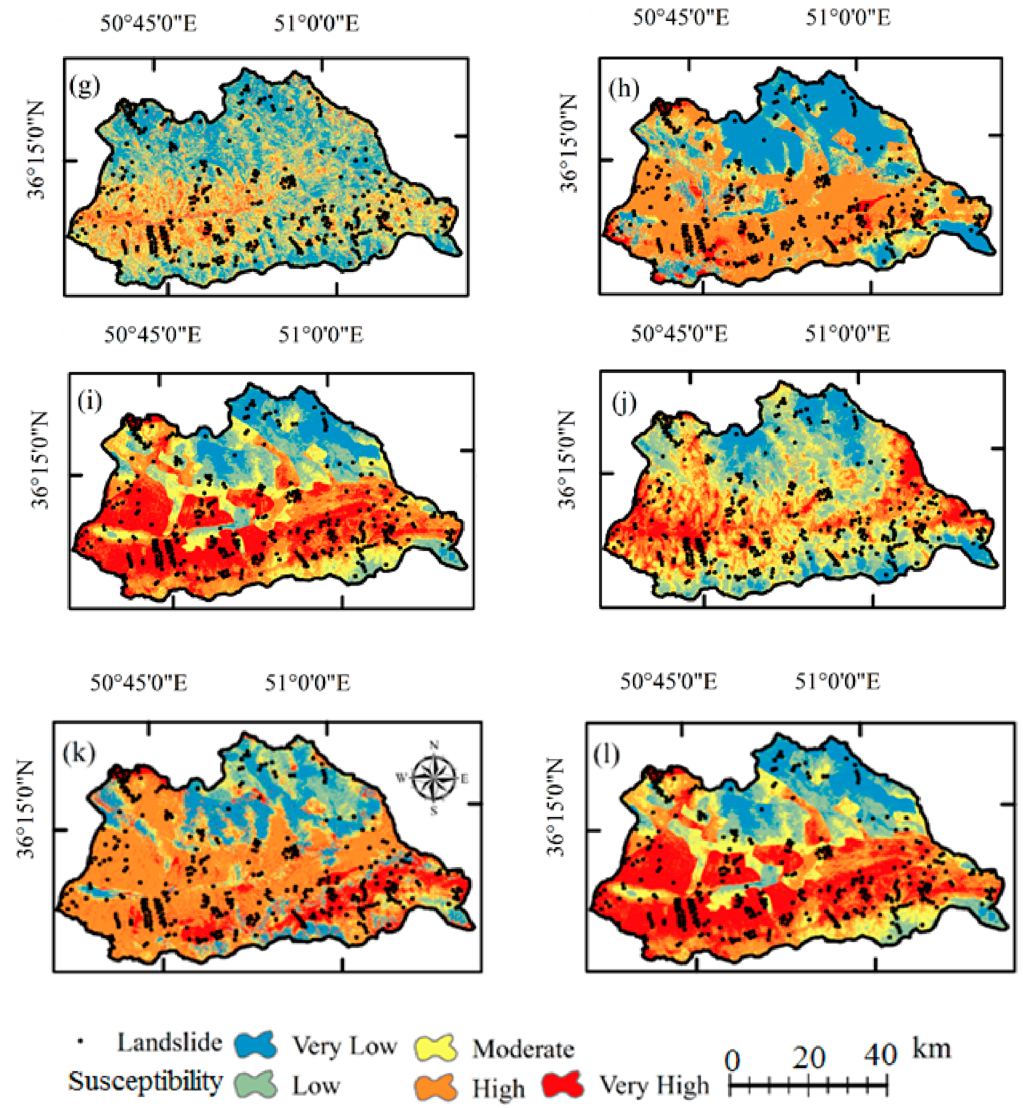

Figure 5. Landslide susceptibility mapping produced by using K-fold-1 (**a**) CDT, (**b**) CDT-ADTree, and (**c**) ADTree; using K-fold-2 (**d**) CDT, (**e**) CDT-ADTree, and (**f**) ADTree; using K-fold-3 (**g**) CDT, (**h**) CDT-ADTree, and (**i**) ADTree; and using K-fold-4 (**j**) CDT, (**k**) CDT-ADTree, and (**l**) ADTree.

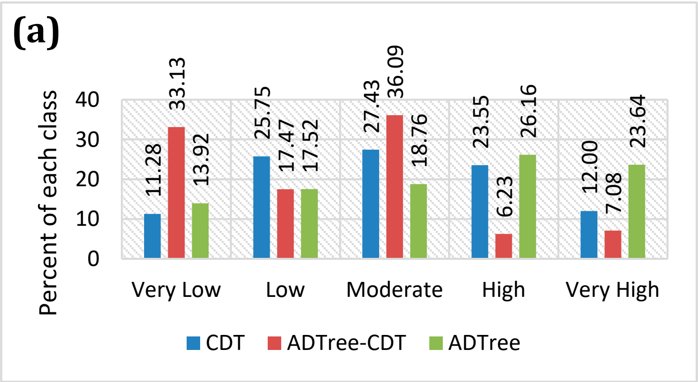

Figure 6. *Cont.*

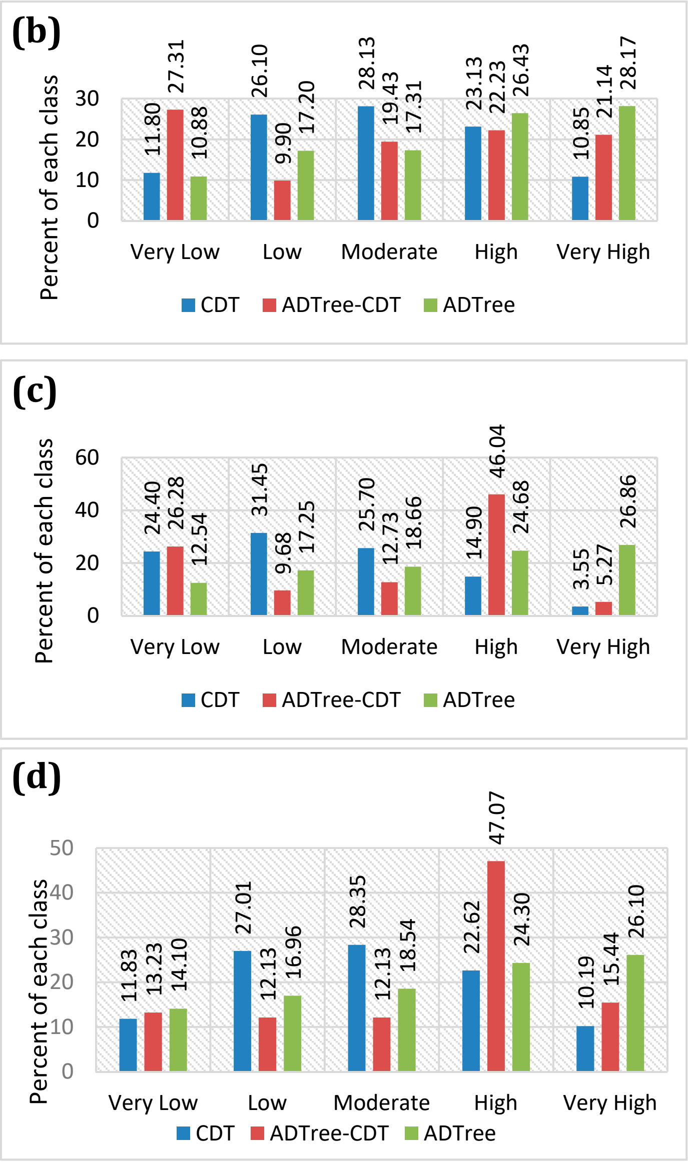

Figure 6. Percentage of each susceptibility class using (**a**) K-fold-1 for CDT, ADTree-CDT and ADTree, (**b**) K-fold-2 for CDT, ADTree-CDT and ADTree, (**c**) K-fold-3 for CDT, ADTree-CDT and ADTree, and (**d**) K-fold-4 for CDT, ADTree-CDT and ADTree.

3.3. Evaluation of the Landslide Susceptibility Models

The evaluation results of landslide susceptibility using the training dataset strengthened the landslide susceptibility model, whereas using the validating dataset, the evaluation outcome reflected suitable prediction analysis of the models. In this study, Table 2 showed the performance validation result of ADTree, CDT, ADTree-CDT machine learning ensemble models in K1 to K4-fold CV landslide classification methods. The several validation indexes, like AUC of ROC curve, accuracy, and precision were applied for calculating and evaluating landslide susceptibility model performance with considering training and testing dataset of landslide. The evaluation results of the landslide susceptibility mapping (Table 2 and Figure 8) indicated that the performance of all models was good, but the ensemble of the CDT-ADTree package was much better than the other two single ML models. The results of the ADTree-CDT ensemble method presented the K3-fold CV with the highest accuracy (0.803), the K2-fold CV has the lowest accuracy (0.7393) and the ADTree-CDT average accuracy is 0.806. The mean AUC, accuracy and precision were extracted from the mean K1, K2, K3 and K4-fold CV methods used to generalize the evaluation criteria for each model. Precision criteria of the evaluation showed that CDT-ADTree had the highest average accuracy in K2 (0.691), followed by ADTree in K4 (0.66), and CDT in K1 (0.562). The AUC of the mentioned three model in K1 to K4-fold CV classification method were graphically presented in Figure 7 by the receiver operating characteristic curve (ROC). The ADTree-CDT model has the highest AUC and success compared to the ADTree and CDT models, which demonstrate the excellent predictive capability of the models. Figure 7 showed the AUC value of ROC for the ADTree-CDT in K1 to K4-fold CV method was 0.828, 0.795, 0.803, and 0.801, respectively. The AUC value of ROC for the ADTree in K1 to K4-fold CV method was 0.746, 0.756, 0.728 and 0.721, respectively. The same as the AUC value for CDT in K1 to K4-fold CV method was 0.647, 0.611, 0.582 and 0.639, respectively (Figure 8). Therefore, based on the evaluation and validation of these three models it was revealed that the ensemble of CDT-ADTree is much better than the others. The validation results by using the aforementioned statistical index shows that all of the three model's results established a strong agreement between the existing landslide and the future predictive maps of landslide susceptibility models.

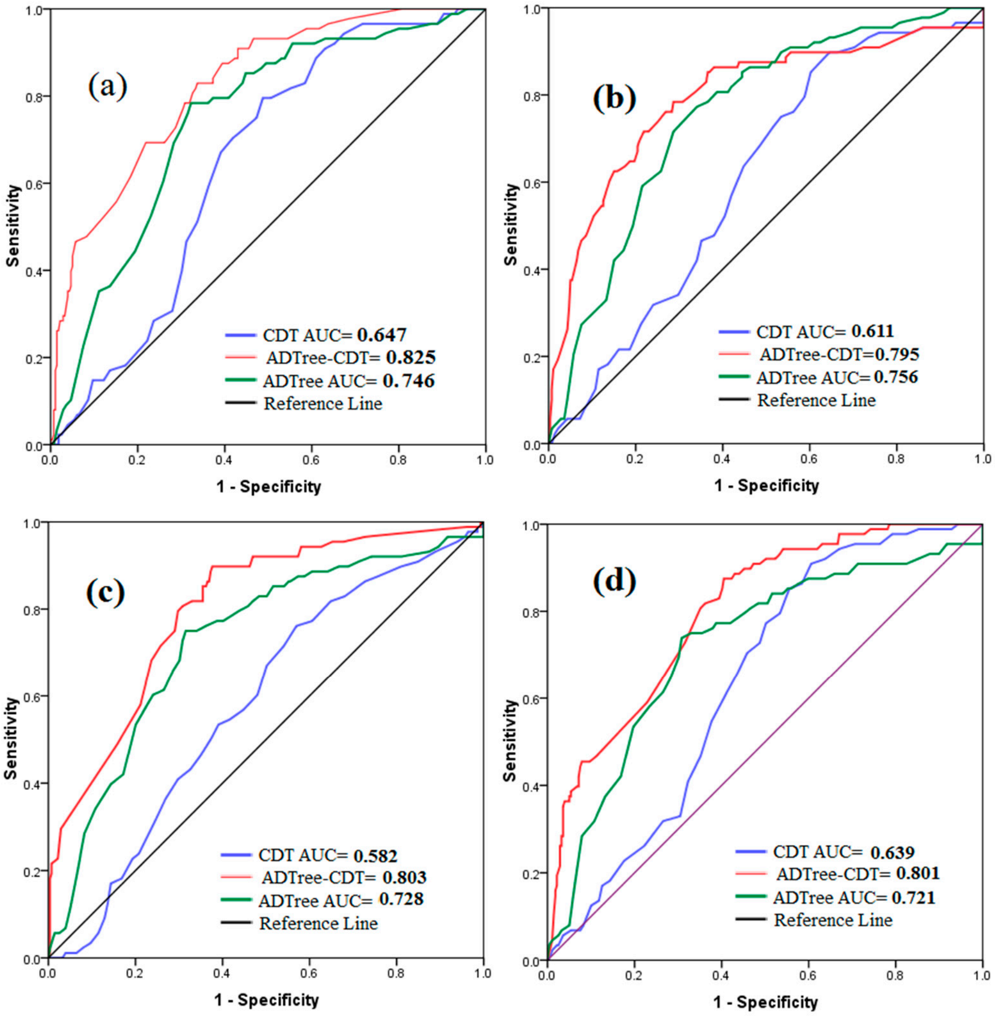

Figure 7. Area under the receiver operation characteristics (ROC) curve (**a**) K1, (**b**) K2, (**c**) K3, (**d**) K4.

Table 2. The accuracy assessment and validation result of three Machine Learning (ML) models by using area under the curve (AUC), accuracy, and precision index.

Criteria	Models	K1	K2	K3	K4	Average
AUC	ADTree	0.746	0.756	0.728	0.721	0.737
	CDT	0.647	0.611	0.582	0.639	0.619
	ADTree-CDT	0.828	0.795	0.803	0.801	0.806
Accuracy	ADTree	0.669	0.626	0.627	0.63	0.638
	CDT	0.502	0.483	0.495	0.516	0.499
	ADTree-CDT	0.652	0.661	0.664	0.659	0.659
Precision	ADTree	0.659	0.636	0.657	0.66	0.653
	CDT	0.562	0.523	0.525	0.526	0.534
	ADTree-CDT	0.687	0.691	0.684	0.689	0.688

3.4. Models Evaluation Through Robustness Testing

The robustness of the prognostic model is expressed as the stability of the performance of the models when the alteration takes place in the training and validation samples [73]. This measure is calculated by differentiating the maximum and minimum model evaluation criteria, such as AUC, accuracy, and true skill statistic (TSS). In this study, the robustness of the models was measured by the average value of the model evaluators using landslide sample datasets. The result of robustness based on AUC, accuracy, and precision criteria are presented in Figure 8. The findings of Figure 8 depict that all three models sounded stable and robust in the validating phase because the difference of maximum and minimum evaluator is very minute. As can be seen from Figure 8a, ADTree-CDT has the minimum AUC robustness index, which means the highest stable compared to CDT and ADTree model. From Figure 8b, it is observed that the accuracy and robustness of the ADTree-CDT is better than the CDT and ADTree model. Figure 8c indicates the ADTree-CDT model is a stable model with the minimum precision–robustness index, followed by the CDT and ADTree model. Therefore, all three models were robust based on AUC, accuracy, and precision robustness index and the ADTree-CDT model has the best robustness indices except for the AUC index.

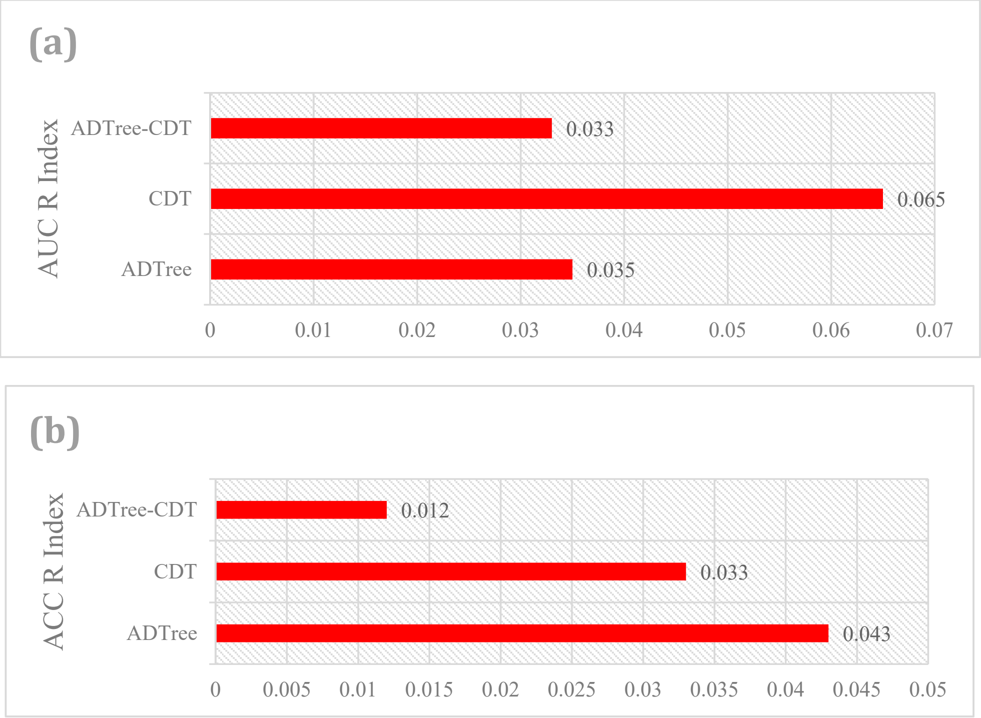

Figure 8. *Cont.*

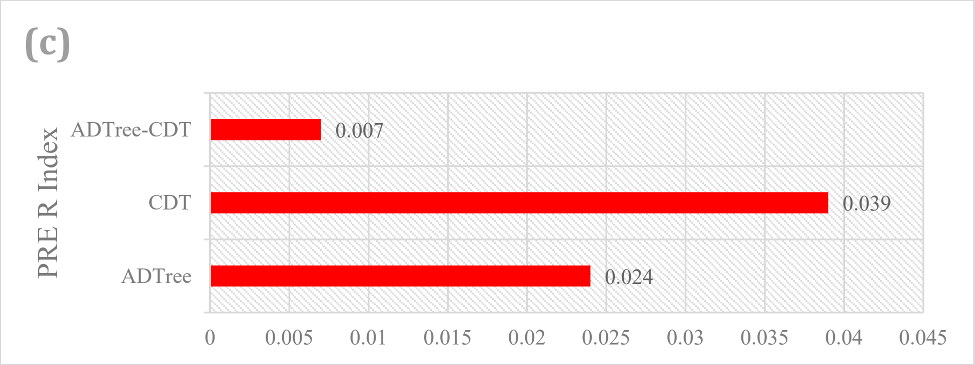

Figure 8. Robustness of models using (**a**) area under the curve, (**b**) accuracy (ACC), (**c**) precision (PRE).

3.5. Contribution of the Factors in the Modelling Process

The contributions of landslide susceptibility conditioning factors were assessed using the MDA index in RF algorithm and the lithology was found to be the most important factor of the landslide susceptibility model in the various k-fold CV classification methods (K1 to K4) (Table 3). After the lithology, elevation is also a significant variable followed by slope factor of different k-fold CV classification methods. According to the K1-fold CV method, lithology, elevation, slope, distance to fault, TWI, soil texture, rainfall, and surface area are the most important factors whereas the rest of the other variables, such as land use and land cover, cross-sectional curvature, flow accumulation, distance to stream, topographic position index etc., are the less important variables, which were identified by the modeling procedure. The K2, K3, and K4-fold CV classification suggested the approximately same scenario of variable importance. After the importance values analysis, we found that there is a strong relationship between the factors of lithology, elevation, slope, distance to fault, distance to river, rainfall, and TWI, and the occurrences of landslide in this study area. On the other hand, land use and land cover, cross-sectional curvature, flow accumulation, and the topographic position index have less importance on landslide occurrence in the study area.

Table 3. Several variables importance in landslide occurrences using random forest (RF) algorithm in four k-fold systems of the present study area.

Factors	K1	K2	K3	K4
Lithology	36.80	34.38	33.96	33.76
Elevation	25.76	25.48	28.73	27.61
Slope	6.11	5.33	5.86	6.50
DtR	2.75	3.24	2.65	3.91
TWI	4.03	5.18	4.21	4.44
Rainfall	3.33	4.06	4.28	3.61
DtF	4.26	6.51	4.76	4.98
Soil texture	3.37	3.69	3.71	3.06
SA	2.64	2.44	2.85	2.61
PC	1.51	1.50	1.48	1.33
Curvature	1.69	1.68	1.54	1.81
DtS	1.60	1.55	1.66	1.31

Table 3. *Cont.*

Factors	K1	K2	K3	K4
TPI	1.53	1.76	1.46	1.57
DD	1.65	1.26	1.43	1.67
PrC	1.63	1.25	1.13	1.12
CSC	1.40	1.29	1.24	1.29
CI	2.16	2.31	1.82	1.88
LU/LC	1.43	1.42	1.18	1.72
LC	2.58	2.16	2.41	2.05
SPI	1.46	1.88	1.46	1.62
FA	0.95	0.72	0.77	0.92
TC	1.56	1.27	1.75	1.49

4. Discussion

Landslides have caused huge economic losses and have endangered the shelter of human settlements, particularly in the hilly and mountainous regions throughout the world [56]. The landslide susceptibility technique has been used for last 30 years to tackle the spatial prediction analysis of landslide occurrences. Assortments of approaches have been applied for spatial prediction of landslides or landslide susceptibility over the world and the objectives of all these methods are similar. Initially, the landslide susceptibility model was evolved by using several statistical methods like analytical hierarchical process (AHP), multi-criteria decision analysis (MCDA), and very recently, natural hazards researchers have focused on the application of machine learning and hybrid models to do the same. Ensemble hybrid machine learning methods have also increased in popularity in the field of geospatial modeling. Thus, machine learning and ensemble approaches have been widely used for the assessment of landslide susceptibility due to some of their great advantages in modeling as well as accuracy in output results. The objective of this study is to establish a new hybrid machine learning model for landslide susceptibility mapping at the Taleghan basin. Throughout this research work, the empirical study was investigated by the application of three decision tree classifiers, such as ADTree, CDT, and the ensemble of CDT-ADTree and CV methods (K1 to K4-fold classification) for preparing the landslide susceptibility maps.

In this modeling process and analysis, we processed multi-collinearity tests among the conditioning factors and selected 22 factors. The selection of landslide conditioning factors was based on the geo-environmental characteristics of the study area and on similar landslide susceptibility studies. Literature studies have shown that CDT and ADTree ML models were used separately in the field of landslide evaluation studies, in two different areas with the combination of other ML algorithms [21,27]. From this research study, we chose maximum conditioning factors along with some other factors that are favorable for this research area. Therefore, in this study we used these two ML algorithms as a standalone along with their ensemble to get better prediction results as the ensemble of ML models always give better results than the single ML models. Since it is important to quantify the multi-collinearity of the selected variables for landslide susceptibility modeling, the importance of the predictors was determined [7]. According to the importance of variable analysis by using random forest algorithm of the k-fold CV methods, lithology is the most important landslide factor in this study area, followed by elevation in all four CV methods. In the case of CDT model-based study on landslides [21], the variables importance was calculated by using correlation attribute evaluation (CAE) methods and it was found that NDVI is the most important factor, followed by distance to road and land use. In another case of ADTree-based landslide studies, Chi-square attribute evaluation (CSAE) techniques were used for variables importance and it was noticed that land cover is the most important followed by geomorphology and valley depth [27]. Therefore, our research study shows that the combination of these two ML models gives better results than the other two mentioned above. Furthermore, many studies have shown that slope instability is mainly due to landslides in the mountainous region due to steep terrain and the sun-facing slope of the hills [83,84], and in this study slope is the third most important factor for occurrences of landslides.

Over the last two decades, several statistical and empirical models have been used to predict spatial distribution of landslide susceptibility mapping using remote sensing data and GIS technology [12,85]. These statistical techniques have also some limitations in mapping landslide susceptibility, whilst the approach of machine learning ensemble has been successfully popularized due to the optimal prediction accuracy of this landslide susceptibility model in order to analyze the multifunctional link between response (landslide) and predictor (conditioning factors) variables [86]. The ADTree model has been considered for the spatial prediction of landslides around the world, and for the Taleghan River Basin we also used this model for the ability to classify binary classes which was highly incorporated into landslide susceptibility studies [27]. It is well known to all that the ADTree is an interpretable and robust algorithm against in binary classification error in comparison to the individual/base decision tree stump classifiers [70]. On the other hand, CDT is a newly invented tree classifier approach to landslide susceptibility studies that addresses classification problems with the credal set algorithm [21,65]. In this study, we have applied the ensemble of CDT and ADTree tree classifier approaches (CDT-ADTree) together to establish a novel landslide susceptibility method.

Furthermore, several researchers used the AUC of the ROC curve, accuracy, and precision to compare the goodness-of-fit and performance of the model [87,88]. Here, each modeling approach has been assessed on the basis of the K1 to K4-fold CV of landslide subsets using the above mentioned reliability measures. Additionally, the result shows that all three models performed well in different k-fold classifications. The CDT-ADTree model is a robust landslide susceptibility model where the average AUC and accuracy are 0.806, 0.659 and 0.688, respectively, compared to the stand-alone CDT and ADTree models. From the other studies, it can be observed that the stand-alone CDT and ADTree ML methods do not have as good accuracy for landslide susceptibility assessment [21,27] as the combination of CDT-ADTree. Previous landslide susceptibility studies [89] in the Taleghan basin showed that seven different statistical, machine learning, and ensemble methods had an AUC of 0.64 to 0.79, whereas the best AUC (0.79) was the maximum entropy model.

5. Conclusions

The main objective of the present research work was to analyze the landslide susceptibility assessment through standalone and ensemble ML approaches, and the present objective was fulfilled in a precise way for the Taleghan basin, Iran. In this study, two single ML algorithms, namely CDT and ADTree, along with ensemble of CDT-ADTree classifier model with four k-folds CV, were successfully used for assessing the landslide susceptibility. The performance of the ensemble model was evaluated and compared with the single ML algorithms. Therefore, accuracy assessment and performance of landslide susceptibility models were compared to each other by using ROC curve with AUC, accuracy, precision, and robustness statistical index. The findings of the model evaluation from the training and validating datasets have shown that the landslide susceptibility model prepared from the ML and ensemble model performed well in prediction analysis and ensemble model represented the best fitted model. The output of the model performance demonstrates that the CDT-ADTree ensemble classifier has the most outstanding performance over the K1 to K4-fold CV than the single ML models. Therefore, it is concluded that the novelty of CDT-ADTree ensemble model can be applied as a new promising technique for spatial prediction of the landslide in further studies. The importance of variables is the secondary finding of the landslide susceptibility analysis. Each k-fold CV method indicates that the aforementioned 22 determining factors were more or less responsible for the spatial landslide modeling, among them lithology is the most important. It was determined, in the findings of the importance of the variables, that lithology and elevation was most important, followed by slope.

On the other hand, variables importance results show that land use and land cover, cross-sectional curvature, flow accumulation, distance to stream, and TPI factors were the least important factors in the modeling procedure for the present study area. Furthermore, the prediction results of this study will help decision makers, planners, and local people to utilize land use in a proper and beneficial way along with most importantly, mitigate the landslide risk in a sustainable manner and minimize the lives and economic losses. Therefore, here we suggest applying the CDT-ADTree ensemble approach to landslide susceptibility mapping in other landslide prone areas to check their occurrences and control devastating damages.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-4292/12/20/3389/s1, Figure S1: title: Landslide points and several landslide conditioning factors (**a**) convergence index (CI), (**b**) cross-sectional curvature (CSC), (**c**) curvature, (**d**) distance to stream (DtS), (**e**) drainage density (DD), (**f**) elevation, (**g**) distance to fault (DtF), (**h**), distance to road (DtR), (**i**) flow accumulation (FA), (**j**) longitudinal curvature (LC), (**k**) plan curvature (PC), (**l**) profile curvature (PrC), (**m**) rainfall, (**n**) slope, (**o**) stream power index (SPI), (**p**) surface area (SA), (**q**) Tangential curvature (TC), (**r**) topography position index (TPI), (**s**) topography wetness index (TWI), (**t**) soil texture, (**u**) land use/land cover (LU/LC). (**v**) Lithology. Table S1: title: Abbreviation and their respective description of several lithological units present in the study area.

Author Contributions: Conceptualization, A.A.; Methodology, A.A.; formal analysis, A.A; investigation, A.A.; resources, A.A.; supervision, A.A.; writing—original draft preparation, A.A., S.C.P., I.C., A.S.; writing—review and editing, A.A., S.C.P., I.C., A.S., S.L., E.K.-S., D.T.B.; Visualization, A.A.; Supervision, A.A., S.L., D.T.B.; Funding Acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Basic Research Project of the Korea Institute of Geoscience and M ineral Resources (KIGAM) and Project of Environmental Business Big Data Platform and Center Construction funded by the Ministry of Science and ICT.

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

References

1. Oh, H.-J.; Lee, S. Shallow landslide susceptibility modeling using the data mining models artificial neural network and boosted tree. *Appl. Sci.* **2017**, *7*, 1000. [CrossRef]
2. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Xu, C.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [CrossRef]
3. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; Natural Hazards Serial; Unesco: Paris, France, 1984.
4. Arabameri, A.; Pourghasemi, H.R.; Yamani, M. Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environ. Earth Sci.* **2017**, *76*, 832. [CrossRef]
5. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tran, T.-T.-T.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [CrossRef]
6. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [CrossRef]
7. Arabameri, A.; Saha, S.; Roy, J.; Chen, W.; Blaschke, T.; Tien Bui, D. Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran. *Remote Sens.* **2020**, *12*, 475. [CrossRef]
8. 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.
9. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena* **2008**, *72*, 1–12. [CrossRef]
10. Faiz, M.A.; Liu, D.; Fu, Q.; Sun, Q.; Li, M.; Baig, F.; Li, T.; Cui, S. How accurate are the performances of gridded precipitation data products over Northeast China? *Atmos. Res.* **2018**, *211*, 12–20. [CrossRef]
11. Pal, S.C.; Das, B.; Malik, S. Potential Landslide Vulnerability Zonation Using Integrated Analytic Hierarchy Process and GIS Technique of Upper Rangit Catchment Area, West Sikkim, India. *J. Indian Soc. Remote Sens.* **2019**, *47*, 1643–1655. [CrossRef]

12. Pal, S.C.; Chowdhuri, I. GIS-based spatial prediction of landslide susceptibility using frequency ratio model of Lachung River basin, North Sikkim, India. *SN Appl. Sci.* **2019**, *1*, 416. [CrossRef]
13. Tsangaratos, P.; Ilia, I.; Hong, H.; Chen, W.; Xu, C. Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. *Landslides* **2017**, *14*, 1091–1111. [CrossRef]
14. Pourghasemi, H.R.; Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ. Earth Sci.* **2016**, *75*, 185. [CrossRef]
15. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [CrossRef]
16. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [CrossRef]
17. Wu, Y.; Ke, Y.; Chen, Z.; Liang, S.; Zhao, H.; Hong, H. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena* **2020**, *187*, 104396. [CrossRef]
18. Pascale, S.; Parisi, S.; Mancini, A.; Schiattarella, M.; Conforti, M.; Sole, A.; Murgante, B.; Sdao, F. Landslide susceptibility mapping using artificial neural network in the urban area of Senise and San Costantino Albanese (Basilicata, Southern Italy). In Proceedings of the 13th International Conference on Computational Science and Its Applications, Ho Chi Minh City, Vietnam, 24–27 June 2013; pp. 473–488.
19. Pham, B.T.; Shirzadi, A.; Bui, D.T.; Prakash, I.; Dholakia, M.B. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India. *Int. J. Sediment Res.* **2018**, *33*, 157–170. [CrossRef]
20. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* **2019**, *175*, 430–445. [CrossRef]
21. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel Entropy and Rotation Forest-Based Credal Decision Tree Classifier for Landslide Susceptibility Modeling. *Entropy* **2019**, *21*, 106. [CrossRef]
22. Tien Bui, D.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K.; et al. Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning Algorithm. *Remote Sens.* **2019**, *11*, 931. [CrossRef]
23. Panahi, M.; Gayen, A.; Pourghasemi, H.R.; Rezaie, F.; Lee, S. Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Sci. Total Environ.* **2020**, *741*, 139937. [CrossRef]
24. Nhu, V.-H.; Zandi, D.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Al-Ansari, N.; Singh, S.K.; Dou, J.; Nguyen, H. Comparison of Support Vector Machine, Bayesian Logistic Regression, and Alternating Decision Tree Algorithms for Shallow Landslide Susceptibility Mapping along a Mountainous Road in the West of Iran. *Appl. Sci.* **2020**, *10*, 5047. [CrossRef]
25. Dou, J.; Yunus, A.P.; Tien Bui, D.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [CrossRef] [PubMed]
26. Nguyen, P.T.; Ha, D.H.; Nguyen, H.D.; Van Phong, T.; Trinh, P.T.; Al-Ansari, N.; Le, H.V.; Pham, B.T.; Ho, L.S.; Prakash, I. Improvement of Credal Decision Trees Using Ensemble Frameworks for Groundwater Potential Modeling. *Sustainability* **2020**, *12*, 2622. [CrossRef]
27. 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]
28. Available online: https://www.cri.ac.ir/index.php/fa/ (accessed on 31 July 2020).
29. Emberger, L. *La Végétation de la Région Méditerranéenne: Essai d'une Classification des Groupements Végétaux*; Librairie Générale de l'Enseignement: Paris, France, 1930.

30. Confalonieri, R.; Bellocchi, G.; Tarantola, S.; Acutis, M.; Donatelli, M.; Genovese, G. Sensitivity analysis of the rice model WARM in Europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environ. Model. Softw.* **2010**, *25*, 479–488. [CrossRef]
31. GSI.IR. Available online: https://gsi.ir/en (accessed on 31 July 2020).
32. Douran Portal. Available online: http://www.areo.ir/en-US/AREEO/7747/page/Soil-Conservation-and-Watershed-Management-Researc (accessed on 31 July 2020).
33. Davoodi, R.; Brown, I.E.; Loeb, G.E. Advanced modeling environment for developing and testing FES control systems. *Med. Eng. Phys.* **2003**, *25*, 3–9. [CrossRef]
34. Lei, X.; Chen, W.; Avand, M.; Janizadeh, S.; Kariminejad, N.; Shahabi, H.; Costache, R.; Shahabi, H.; Shirzadi, A.; Mosavi, A. GIS-Based Machine Learning Algorithms for Gully Erosion Susceptibility Mapping in a Semi-Arid Region of Iran. *Remote Sens.* **2020**, *12*, 2478. [CrossRef]
35. Cornforth, D.H.; Cornforth, D. *Landslides in Practice: Investigation, Analysis, and Remedial/Preventative Options in Soils*; Wiley: Hoboken, NJ, USA, 2005; ISBN 0-471-67816-3.
36. Zhang, T.; Han, L.; Chen, W.; Shahabi, H. Hybrid Integration Approach of Entropy with Logistic Regression and Support Vector Machine for Landslide Susceptibility Modeling. *Entropy* **2018**, *20*, 884. [CrossRef]
37. Jacobs, L.; Dewitte, O.; Poesen, J.; Sekajugo, J.; Nobile, A.; Rossi, M.; Thiery, W.; Kervyn, M. Field-based landslide susceptibility assessment in a data-scarce environment: The populated areas of the Rwenzori Mountains. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 105–124. [CrossRef]
38. San, B.T. An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The Candir catchment area (western Antalya, Turkey). *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 399–412. [CrossRef]
39. Lombardo, L.; Cama, M.; Maerker, M.; Rotigliano, E. A test of transferability for landslides susceptibility models under extreme climatic events: Application to the Messina 2009 disaster. *Nat. Hazards* **2014**, *74*, 1951–1989. [CrossRef]
40. Mahalingam, R.; Olsen, M.J.; O'Banion, M.S. Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study). *Geomat. Nat. Hazards Risk* **2016**, *7*, 1884–1907. [CrossRef]
41. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artifcial Intelligence (IJCAI), Montreal, QC, Canada, 20–25 August 1995; Volume 14, pp. 1137–1145.
42. Wiens, T.S.; Dale, B.C.; Boyce, M.S.; Kershaw, G.P. Three way k-fold cross-validation of resource selection functions. *Ecol. Model.* **2008**, *212*, 244–255. [CrossRef]
43. Václavík, T.; Meentemeyer, R.K. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecol. Model.* **2009**, *220*, 3248–3258. [CrossRef]
44. Boria, R.A.; Olson, L.E.; Goodman, S.M.; Anderson, R.P. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Model.* **2014**, *275*, 73–77. [CrossRef]
45. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sens.* **2020**, *12*, 1737. [CrossRef]
46. Tien Bui, D.; Ho, T.-C.; Pradhan, B.; Pham, B.-T.; Nhu, V.-H.; Revhaug, I. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ. Earth Sci.* **2016**, *75*, 1101. [CrossRef]
47. Wilson, J.P.; Gallant, J.C. *Terrain Analysis: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000; ISBN 0-471-32188-5.
48. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [CrossRef]
49. Horton, R.E. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **1945**, *56*, 275–370. [CrossRef]
50. Ding, Q.; Chen, W.; Hong, H. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 619–639. [CrossRef]
51. Gallant, J.C.; Austin, J.M. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Res.* **2015**, *53*, 895–906. [CrossRef]

52. Mokarram, M.; Roshan, G.; Negahban, S. Landform classification using topography position index (Case study: Salt dome of Korsia-Darab plain, Iran). *Model. Earth Syst. Environ.* **2015**, *1*, 40. [CrossRef]
53. Jebur, M.N.; Pradhan, B.; Tehrany, M.S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sens. Environ.* **2014**, *152*, 150–165. [CrossRef]
54. Ozdemir, A. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.* **2011**, *405*, 123–136. [CrossRef]
55. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [CrossRef]
56. Chen, W.; Yan, X.; Zhao, Z.; Hong, H.; Bui, D.T.; Pradhan, B. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). *Bull. Eng. Geol. Environ.* **2019**, *78*, 247–266. [CrossRef]
57. Alin, A. Multicollinearity. *WIREs Comput. Stat.* **2010**, *2*, 370–374. [CrossRef]
58. Jensen, D.R.; Ramirez, D.E. Revision: Variance inflation in regression. *Adv. Decis. Sci.* **2013**, *2013*, 671204. [CrossRef]
59. Liao, D.; Valliant, R. Variance inflation factors in the analysis of complex survey data. *Surv. Methodol.* **2012**, *38*, 53–62.
60. Roy, P.; Chakrabortty, R.; Chowdhuri, I.; Malik, S.; Das, B.; Pal, S.C. Development of Different Machine Learning Ensemble Classifier for Gully Erosion Susceptibility in Gandheswari Watershed of West Bengal, India. *Mach. Learn. Intell. Decis. Sci.* **2020**, 1–26. [CrossRef]
61. 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 of evidential belief function–logistic regression algorithm. *Land Degrad. Dev.* **2018**, *29*, 4035–4049. [CrossRef]
62. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
63. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
64. Kim, J.-C.; Lee, S.; Jung, H.-S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **2018**, *33*, 1000–1015. [CrossRef]
65. Abellán, J.; Moral, S. Building classification trees using the total uncertainty criterion. *Int. J. Intell. Syst.* **2003**, *18*, 1215–1225. [CrossRef]
66. Mantas, C.J.; Abellán, J. Analysis and extension of decision trees based on imprecise probabilities: Application on noisy data. *Expert Syst. Appl.* **2014**, *41*, 2514–2525. [CrossRef]
67. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 57–72.
68. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976; ISBN 978-0-691-10042-5.
69. Walley, P. Inferences from Multinomial Data: Learning About a Bag of Marbles. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 3–34. [CrossRef]
70. Freund, Y.; Mason, L. The alternating decision tree learning algorithm. In Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99), Bled, Slovenia, 27–30 June 1999; Volume 99, pp. 124–133.
71. 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]
72. Cheung, D.; Williams, G.J.; Li, Q. (Eds.) *Advances in Knowledge Discovery and Data Mining, Proceedings of the 5th Pacific-Asia Conference (PAKDD 2001), Hong Kong, China, 16–18 April 2001*; Springer: Berlin/Heidelberg, Germany, 2001; ISBN 978-3-540-41910-5.
73. Rahmati, O.; Tahmasebipour, N.; Haghizadeh, A.; Pourghasemi, H.R.; Feizizadeh, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* **2017**, *298*, 118–137. [CrossRef]

74. Chakrabortty, R.; Pal, S.C.; Chowdhuri, I.; Malik, S.; Das, B. Assessing the Importance of Static and Dynamic Causative Factors on Erosion Potentiality Using SWAT, EBF with Uncertainty and Plausibility, Logistic Regression and Novel Ensemble Model in a Sub-tropical Environment. *J. Indian Soc. Remote Sens.* **2020**, *48*, 765–789. [CrossRef]
75. Chowdhuri, I.; Pal, S.C.; Chakrabortty, R. Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv. Space Res.* **2020**, *65*, 1466–1489. [CrossRef]
76. Angileri, S.E.; Conoscenti, C.; Hochschild, V.; Märker, M.; Rotigliano, E.; Agnesi, V. Water erosion susceptibility mapping by applying stochastic gradient treeboost to the Imera Meridionale river basin (Sicily, Italy). *Geomorphology* **2016**, *262*, 61–76. [CrossRef]
77. Yesilnacar, E.K. *The Application of Computational Intelligence to Landslide Susceptibility Mapping in Turkey*; University of Melbourne: Melbourne, VIC, Australia, 2005.
78. Cama, M.; Lombardo, L.; Conoscenti, C.; Rotigliano, E. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* **2017**, *288*, 52–65. [CrossRef]
79. Vander Heyden, Y.; Nijhuis, A.; Smeyers-Verbeke, J.; Vandeginste, B.G.M.; Massart, D.L. Guidance for robustness/ruggedness tests in method validation. *J. Pharm. Biomed. Anal.* **2001**, *24*, 723–753. [CrossRef]
80. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy). *Geomorphology* **2014**, *204*, 399–411. [CrossRef]
81. Rahmati, O.; Naghibi, S.A.; Shahabi, H.; Bui, D.T.; Pradhan, B.; Azareh, A.; Rafiei-Sardooi, E.; Samani, A.N.; Melesse, A.M. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *J. Hydrol.* **2018**, *565*, 248–261. [CrossRef]
82. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2014**, *11*, 425–439. [CrossRef]
83. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [CrossRef]
84. Chen, Z.; Wang, J. Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. *Nat. Hazards* **2007**, *42*, 75–89. [CrossRef]
85. Goetz, J.N.; Guthrie, R.H.; Brenning, A. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* **2011**, *129*, 376–386. [CrossRef]
86. Kadavi, P.R.; Lee, C.-W.; Lee, S. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. *Remote Sens.* **2018**, *10*, 1252. [CrossRef]
87. Chen, W.; Shirzadi, A.; Shahabi, H.; Ahmad, B.B.; Zhang, S.; Hong, H.; Zhang, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [CrossRef]
88. Saha, S.; Saha, A.; Hembram, T.K.; Pradhan, B.; Alamri, A.M. Evaluating the Performance of Individual and Novel Ensemble of Machine Learning and Statistical Models for Landslide Susceptibility Assessment at Rudraprayag District of Garhwal Himalaya. *Appl. Sci.* **2020**, *10*, 3772. [CrossRef]
89. Mokhtari, M.; Abedian, S. Spatial prediction of landslide susceptibility in Taleghan basin, Iran. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1297–1325. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.