

Article

Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms

Binh Thai Pham ¹, Ataollah Shirzadi ², Himan Shahabi ^{3,*}, Ebrahim Omidvar ⁴,
Sushant K. Singh ⁵, Meheebub Sahana ⁶, Dawood Talebpour Asl ³, Baharin Bin Ahmad ⁷,
Nguyen Kim Quoc ⁸ and Saro Lee ^{9,10,*}

¹ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

² Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran

³ Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran

⁴ Department of Rangeland and Watershed Management, Faculty of Natural Resources and Earth Sciences, University of Kashan, Kashan 87317-53153, Iran

⁵ Virtusa Corporation, 10 Marshall Street, Irvington, NJ 07111, USA

⁶ IGCMC, WWF-India, New Delhi-110003, India

⁷ Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia

⁸ Department of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam

⁹ Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124 Gwahak-ro, Yuseong-gu, Daejeon 34132, Korea

¹⁰ Department of Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Korea

* Correspondence: h.shahabi@uok.ac.ir (H.S.); leesaro@kigam.re.kr (S.L.); Tel.: +98-91-8665-8739 (H.S.); +82-42-8683-057 (S.L.)

Received: 10 July 2019; Accepted: 10 August 2019; Published: 13 August 2019



Abstract: Landslides have multidimensional effects on the socioeconomic as well as environmental conditions of the impacted areas. The aim of this study is the spatial prediction of landslide using hybrid machine learning models including bagging (BA), random subspace (RS) and rotation forest (RF) with alternating decision tree (ADTree) as base classifier in the northern part of the Pithoragarh district, Uttarakhand, Himalaya, India. To construct the database, ten conditioning factors and a total of 103 landslide locations with a ratio of 70/30 were used. The significant factors were determined by chi-square attribute evaluation (CSEA) technique. The validity of the hybrid models was assessed by true positive rate (TP Rate), false positive rate (FP Rate), recall (sensitivity), precision, F-measure and area under the receiver operating characteristic curve (AUC). Results concluded that land cover was the most important factor while curvature had no effect on landslide occurrence in the study area and it was removed from the modelling process. Additionally, results indicated that although all ensemble models enhanced the power prediction of the ADTree classifier ($AUC_{\text{training}} = 0.859$; $AUC_{\text{validation}} = 0.813$); however, the RS ensemble model ($AUC_{\text{training}} = 0.883$; $AUC_{\text{validation}} = 0.842$) outperformed and outclassed the RF ($AUC_{\text{training}} = 0.871$; $AUC_{\text{validation}} = 0.840$), and the BA ($AUC_{\text{training}} = 0.865$; $AUC_{\text{validation}} = 0.836$) ensemble model. The obtained results would be helpful for recognizing the landslide prone areas in future to better manage and decrease the damage and negative impacts on the environment.

Keywords: landslide; meta classifier; performance; goodness-of-fit; GIS; India

1. Introduction

Landslide is a local natural phenomenon, popularly understood as the “mass displacement of earth, debris, and rocks”, that can be triggered by hydrological, topographic, and geophysical reasons [1]. However, anthropogenic activities such as mining and construction, as well as natural events including heavy rainfall, earthquake, volcanic eruption, and marine erosion can also trigger landslides [2]. Landslides have multidimensional effects on the socioeconomic as well as environmental conditions of the impacted areas. Hilly areas are highly prone to landslides and the after effect may significantly change the topographic features as well as river course and patterns [3]. The environmental loss could be in the form of forest along with habitat and wildlife destruction, and can elicit mild to severe tsunami and floods. Human populations living in the landslide susceptible areas are the foremost victims of landslides and may experience loss of houses, cattle, fertile lands, and even lives of their families and friends.

The World Bank reported that nearly 3.7 million square kilometers of world’s land area is highly prone to landslides which could put at risk about 300 million human lives [4]. After nearly one and half decades later, these figures must have changed; however, because of the lack of consistency in the landslide reports, it is very difficult to come up with an exact number of the landslide incidences and the fatalities. These inconsistencies in the reports could be because of the varied nature of landslides; for example, landslides could be seismic only or triggered by rainfall, rock-slides, floods, or hurricane. In a recent study, the authors provide a good discussion of the discrepancies in landslide reports [2]. Furthermore, the authors report that between only 2004 and 2016, a total of 4862 landslide events occurred globally, with high impacts in the Central and South America, Caribbean Islands, East Africa, Turkey, Iran, European Alps, and Asia [2].

These seismic landslides caused 55,997 fatalities [2]. On the other hand, the National Aeronautics and Space Administration (NASA) reported that nearly 10,804 landslides occurred between 2007 and 2017, triggered by rainfall [5]. Petley captured the landslides events between 2004 and 2010, and reported 2620 landslides that caused 32,322 human losses [6]. The global infrastructure and economic damages due to landslides are daunting, costing about US\$20 billion, the highest in the USA (US\$12.1–4.3 billion) and Italy (US\$3.9 billion), followed by Japan (>US\$3.0 billion), India (US\$2.0 billion), China (>US\$1.0 billion), and Germany (US\$0.3 billion) [7]. These numbers have significantly risen in the last decades when compared with previous reports. For example, in 1992, China’s estimated annual economic loss due to landslides was nearly US\$500 million [8] and in 1998, USA’s estimated economic damages due to landslides was approximately US\$1–3.6 billion [9].

The Himalayan Arc across Indian and southeastern China has experienced the highest landslide events, followed by areas of Laos, Bangladesh, Myanmar, Indonesia, and the Philippines [2]. India, the scope of this study, has experienced severe naturally triggered landslides in the 21st century. In 2001, nearly 40 individuals died in the Amboori of Kerala state of India; in 2013, a landslide occurred in Kedarnath of Uttarakhand state of India and more than 5000 people died; and in 2014 in the Pune of Maharashtra state of India, over 100 individuals were found to be missing after the landslide (Figure 1) [10,11]. Figure 1 depicts the locations of landslide events in India and the fatalities that occurred at such locations—an estimated 12.6% of land area of India and approximately 4.5 million USD worth economic damage [12].

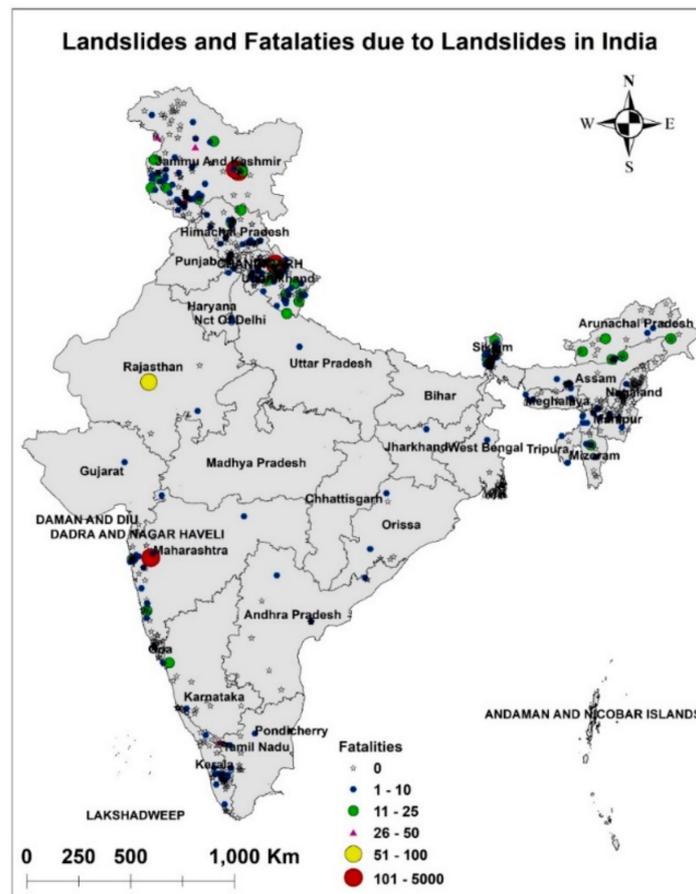


Figure 1. Landslides and fatalities due to landslide in India (This map is created based on the data using ArcGIS version 10.6.1) [11].

Based on our analysis of the NASA data, we found that a total of 958 landslide events occurred between 2007 and 2015 that caused 6779 human fatalities (Table 1).

Table 1. Landslide events and fatalities caused in India between 2007 and 2015 [11].

Indian State	Sum of Fatalities	Indian State	Sum of Fatalities
Uttarakhand	5228	Andhra Pradesh	19
Jammu and Kashmir	590	Manipur	12
Maharashtra	251	Orissa	12
Himachal Pradesh	114	Goa	8
Arunachal Pradesh	88	Uttar Pradesh	6
West Bengal	73	Gujarat	5
Assam	72	Nagaland	5
Sikkim	66	Tripura	4
Rajasthan	54	Jharkhand	2
Tamil Nadu	46	NCT	2
Kerala	38	Haryana	1
Meghalaya	30	Bihar	-
Karnataka	29	Odisha	-
Mizoram	24	Total	6779

The landslides may not be stopped or controlled; however, the losses can be reduced by establishing a decision support system to predict possible landslides or identifying landslide prone areas for management. Therefore, prediction of possible landslides at the local and regional levels is required for pro-active landslide mitigation policy creation and management [13]. Although

domain-knowledge-driven qualitative approach is advantageous in predicting landslides, data-driven quantitative methods are widely used because collecting field data from landslide areas are challenging and hard to acquire [3]. Pourghasemi et al. [14] reported that a variety of quantitatively-statistical, multi-criteria decision making, and machine learning-methods have been applied for predicting landslide susceptibility, of which logistical regression [15–18] is the most frequently used method, followed by the frequency ratio [19,20], weights-of-evidence [18,21], artificial neural networks [22,23], analytic hierarchy process [24,25], statistical index [26], index of entropy [27–30], and support vector machine [31,32]. Environmental data collected from fields as well as extracted from satellite images to develop landslide prediction models are diverse in nature, and therefore prone to inaccuracies [13]. To mitigate these challenges, researchers have applied various fuzzy-based techniques that are yet to accomplish satisfactory results [13].

Machine learning techniques have recently gained good attention among the environmental modeling research community as they are advantageous in efficiently capturing the complex relationship between the environmental predictors and the response, such as flood [33–41], wildfire [42], sinkhole [43], drought [44], gully erosion [45,46], groundwater [47–49] and land/ground subsidence [27], and landslide in this case [3,13,50–57]. In due course, researches have also attempted to improve the prediction accuracy and the interpretability of the models through applying various decision-trees machine learning algorithms such as chi-square automatic interaction detector; quick, unbiased and efficient statistical tree [58]; J48 decision trees [59]; ID3 decision trees [60]; random forests [61]; classification and regression trees [62]; alternating decision trees [63]; reduced error pruning trees [3]; naïve Bayes [35,53]; naïve Bayes tree [13,64]; kernel logistic regression [37]; logistic model tree [38,65]; Bayesian logistical regression [66] and support vector regression [67]. There is a consistent recommendation of applying hybrid or ensemble models as they have shown promising results in correctly predicting the landslide susceptibility on smaller, but complex datasets [3,13,53,62,68–70].

In order to correctly predict possible landslides, landslide prone areas have to be clearly understood and should be used for prediction model development. This study aims to fill the above research gaps by introducing a novel decision tree-based hybrid machine learning system to correctly predict the landslide susceptible areas. To achieve a landslide susceptibility map with reliable and high prediction accuracy, we ensembled a decision tree, ADTree, as a base classifier with several Meta classifier namely bagging (BA), random subspace (RS) and rotation forest (RF) ensemble at the northern part of the Pithoragarh district, Uttarakhand, Himalaya, India. An alternative decision tree (ADTree) classifier is one of the powerful algorithms among decision tree algorithms which is rarely used for spatial prediction modelling [71,72]. It combine decision rules by the boosting and decision tree algorithms in the classification problems and therefore it can produce a simpler structure and also its interpretation of classification rules are simple and more visualizable [73]. The modelling process was carried out using Arc GIS 10.3, and Weka 3.9.

2. Description of the Study Area

The study area is located in the northern part of the Pithoragarh region, Uttarakhand in India, between the latitudes 29°59'13" N and 29°48'2" N and longitudes 80°0'34" E and 80°12'28" E, respectively, covering an area of about 242 km² (Figure 2). Topographically, the region includes rugged hills and high mountain peaks which are dissected by long, narrow and deep valleys. The maximum elevation of the basin is 2713 m in the north and at least 757 m in the south at the outlet of the East Ramganga River from the basin. The East Ramganga River originates from the Namik glacier in the Himalayan Mountains, and flows into the Ganges River after passing 108 km in the Kumaon Region. The average slope is 28.61° and the maximum gradient is 75.50°. The areas with a steep slope are related to the slopes overlooking the river bed. Additionally, 9.6% of the basin area is more than 44° slope, and only 14.81% of the basin is less than 15° slope.

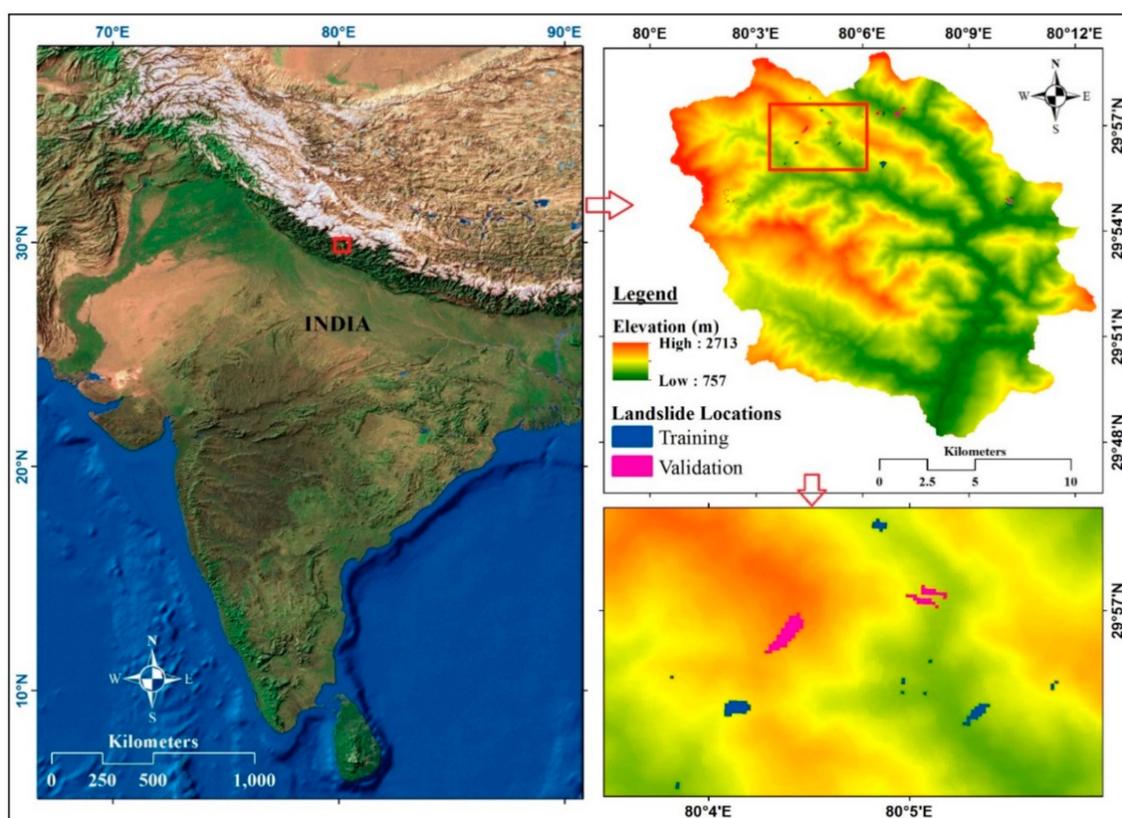


Figure 2. Location of landslides in the study area.

In terms of land cover, 57% of the study area is covered by moderate to high density of vegetation and 23.32% of the area is under the cultivated lands. The rest of the basin includes sparsely vegetated (10.57%), barren (6.84%), lakes and rivers (1.33%), settlement (0.31%) and extensive slope cut (0.31%). In recent years, the major part of the basin has been converted into low-density forests and land degradation due to the destruction of forests and land use change from forest to agricultural land. The very fertile lands are located on the riverside. The settlements are surrounded by vast areas of agricultural lands. Most soil mass movements occur along the river valleys and the periphery of roads that are drawn along the rivers, especially during the rainy season.

In terms of geology, this basin is occupied by metamorphic rocks found in the Dharamgarh Formation (biotite gneiss, chlorite schist, inter bands of schistose quartzite with meta-volcanics) and Bajinath Formation (quartzite and gneiss), meta-sedimentary rocks of the Pithoragarh Formation (dolomitic limestone with intercalations of talcose schist, carbonaceous phyllite, slate, limestone and quartzite) and the Bering Formation (quartzite/arenite/sericitie and phyllite intercalated with meta-volcanics) of Garhwal group and recent alluvium. In terms of lithology, the northern and northeastern part of the basin is mainly covered by slate, quartzite, talc and dolomite. This lithological unit covers an area about 56% of the basin that contained 90.27% of landslides. The remaining 9.73% of landslides are located in colluvium units covering 9% of the basin area. The southwestern part of the study area, consisting of in situ soil and quartzite and slate with basic metavolcanics, has covered an area about 24% of the basin but without any landslides. In terms of geological structure, the area is affected by friction and fault due to tectonic activity. Most likely, the instability of rock in the rock masses is fractured due to discontinuities caused by faults, cracks, fractures and seams.

3. Landside Inventory Map

The past landslide location of any given area gives valuable information on the patterns of spatial distribution of landslide events in the landslide susceptibility zonation [74]. The past landslide

locations help to understand the landslide behavior and relation between the landslide causative factors. On account of this, the making of a landslide inventory is an important step to landslide susceptibility assessment. Many scholars prepare a landslide inventory using high resolution remote sensing data or aerial photograph interpretation [27,75–77]. Every year the area was affected by several active landslides during rainy season or after a rainy season [78] and therefore, the Google Earth data were used to cover rainfall affected landslides. In the present study, the landslide inventory map was prepared by using the Google Earth digitization from the post-rainfall seasons and the locations were field verified. A total of 103 landslide polygons were delineated and converted into the raster format. Of the total delineated landslide locations, we have selected 70% of landslides as training dataset and the remaining 30% of landslide locations for validation datasets (Figure 2).

3.1. Landslide Conditioning Factors

A variety of landslide conditioning factors (LCF) have been used for developing landslide susceptibility prediction models including slope, lithology, aspect, land use, elevation, distance from river, distance from roads, distance from faults, plain curvature, profile curvature, precipitation, topographic wetness index, soil type, stream power index, normalized difference vegetation index, slope length, curvature, and drainage density [14]. The selection of these factors may vary based on the study area, scale of the study, and data availability [14]. Among the above-listed LCF, slope gradient has been the most frequently used LCF in the studies. The selection of LCF in this study is based on the previous research and our field observations.

3.1.1. Overburden Depth

The overburden depth captures the information of depth to the bedrock and has been linked to shallow translational debris landslides [79]. Furthermore, it is also influenced by slope and erosion. The study area is highly prone to erosion and has steep slopes; therefore, the overburden depth could play an important role in identifying landslide prone areas and developing prediction models. The overburden depth in the study area ranges between '0' and '4' m (Figure 3a).

3.1.2. Land Cover

The majority of landslides occur in forest-scant areas, as in densely vegetated areas, the plant roots hold the soil and rocks strongly and keep them stable at steeper slopes, reduces soil erosion, and therefore protects against landslides [80,81]. Henceforth, how various land covers impact the landslide become imperative in developing landslide prediction models. In this study, we have categorized land cover into barren, cultivated land, extensive slope cut, lakes and rivers, moderately vegetated, settlement areas, sparsely vegetated areas, thickly vegetated, and wasteland (Figure 3b).

3.1.3. Geomorphology

Geomorphology is one of the most important LCF as various geomorphological formations represent geomorphological phenomena including alluvial flood plain, colluvial footslop, denudational hill slope, highly dissected hills, lowly dissected hills, moderately dissected hills, ridge, river, and transportation midslope (Figure 3c) [79]. Geomorphology has also been found to have contributing effects on shallow and deep debris landslides [79].

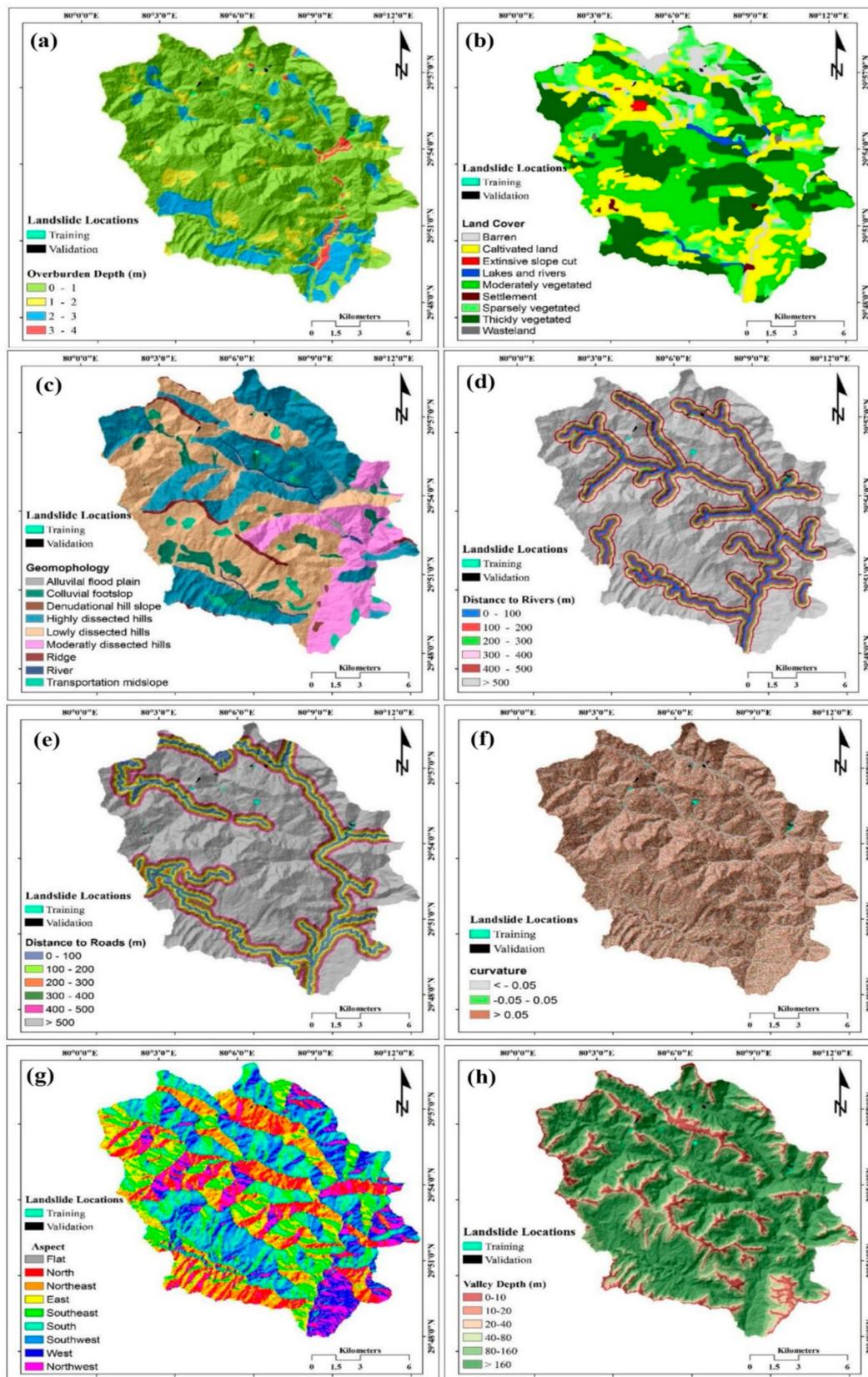


Figure 3. Cont.

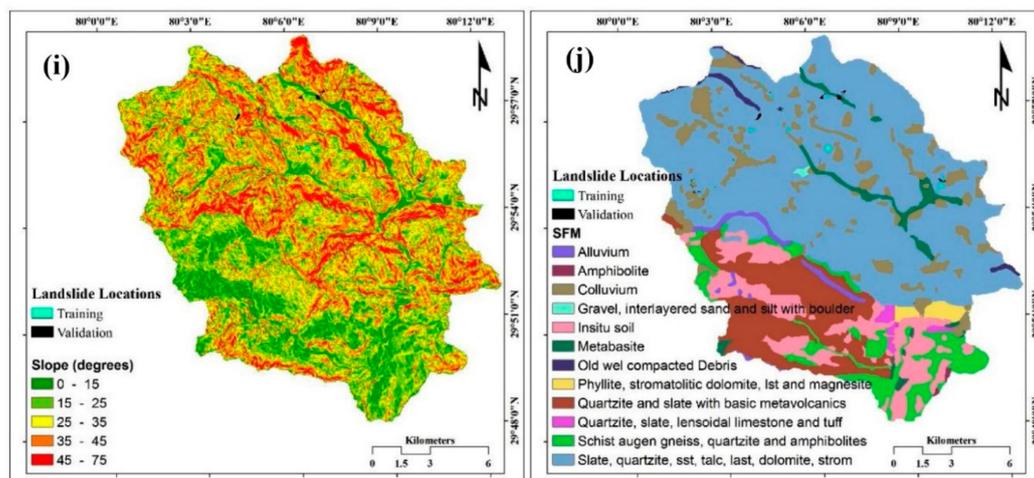


Figure 3. Landslide conditioning factors used in the study area: (a) overburden depth; (b) land cover; (c) geomorphology; (d) distance to rivers; (e) distance to roads; (f) curvature; (g) aspect; (h) valley depth; (i) slope; (j) SFM.

3.1.4. Distance to Rivers

River networks erode the catchment areas in their natural course through surface runoff, therefore making the hilly areas highly vulnerable to landslides. Consequently, distance to river has been an important LCF in those studies where the study areas have dense river networks such as in our case in Uttarakhand [82]. We have classified the distance to rivers into 0–100 m, 100–200 m, 200–300 m, 300–400 m, 400–500 m, and above 500 m from the landslide locations (Figure 3d).

3.1.5. Distance to Roads

As mentioned in the previous section, landslide could be induced by road construction; considering road network in the landslide prediction model development therefore becomes a necessity. As road networks negatively impact the slopes by loosening the slope materials, the distance from roads helps understand the landslide prone areas [83]. Like distance to rivers, we have classified the distance to roads into 0–100 m, 100–200 m, 200–300 m, 300–400 m, 400–500 m, and above 500 m from the landslide locations (Figure 3e).

3.1.6. Curvature

Erosion of riverbanks steepens the curvature, thus acting as a trigger point for landslide. Therefore, knowing whether the curvature is negative, zero, or positive for flat, concave, and convex surfaces is vital in identifying the landslide prone areas and so for developing landslide prediction models [84]. In this study, the curvature is classified into below -0.05 , between -0.05 and 0.05 , and above 0.05 (Figure 3f).

3.1.7. Aspect

Slope aspect is another important LCF that plays a significant role in inducing landslides in the study area as it influences the evapotranspiration by controlling the topographic moisture [82,85]. Slope aspect represents the course of extreme sloping of the terrain surface and moves clockwise starting at 00 (North) and ends to 3600 (West) [86]. The slope aspect in this study is categorized into flat, north, northeast, east, southeast, south, southwest, west, and northwest (Figure 3g).

3.1.8. Valley Depth

The valley depth above 160 m in the study area is highly prone to landslides and showed a positive association between the valley depth and landslides [53]. We have classified valley depth into six categories (Figure 3h).

3.1.9. Slope

Slope is one of the most significant LCF used in developing landslide susceptibility prediction models since with an increase in the slope angles, the likelihood of the occurrence of landslides increases [2,84,87]. Slope is also found to be associated with both shallow translational rockslides and debris slides and has the highest landslide susceptibility predictive capability [79,85]. The study area is precipitous, and the majority of the areas fall between 15–45 degrees and goes up to 700 m, making the area prone to landslide during heavy rainfall. A majority of the landslides in the area was found to have occurred in cut-slopes [82]. We have classified the slope of the study area into 0–15, 15–25, 25–35, 35–45, and 45–75 degrees (Figure 3i).

3.1.10. SFM

Slope forming material (SFM) defines the rock and the soil types of the area and has significant impact on both shallow translational rock and debris landslides [79]. In this study, we classified the SFM into twelve categories based on their rock and soil types (Figure 3j) and a majority of landslide events were reported in the study area with weak rock formed slopes [53].

4. Machine Learning Algorithms

Over time, landslide susceptibility modeling has been considered using both qualitative (inventory-based analysis) and quantitative or data driven models [88,89]. Development of geographical information system (GIS) and machine learning algorithm has provided alternative decision tree (ADTree), support vector machine, artificial neural network and kernel logistic regression (KLR) advanced techniques with precise model building [90]. Machine learning based data driven models with better performance than conventional models are quite appealing these days [88]. Machine learning-based landslide susceptibility models are more cost efficient and rapid than conventional models and can be extended to large area analysis [91]. Use of artificial neural network and support vector machine yielded high prediction accuracy but comparison with other models is still required to understand its precision.

4.1. Base Classifier: Alternating Decision Tree (ADTree)

Decision trees is one of the most advanced classification techniques with minimum probability of error, concomitant robustness, easy interpretation and precise classification, and has seamless applicability in solving real world situations [73]. This model has been built through data portioning in which each iteration data has been split according to the attribute values. Thus, the major goal of this analysis is to split data into subsets unless a subset contains homogenous target value or the predictable attribute. In each split, the impact of selected variables was examined on the predictable attribute. If the predictable attribute comprises discrete data, the resulting tree model is called a classification tree. This decision tree process is also called decision tree induction [92]. The training set inputs are divided into prediction node using split tests to obtain the prediction node values:

$$Z(c) = 2(\sqrt{W + (d)W - (d)} + \sqrt{W + (-d)W - (-d)}) + W' \quad (1)$$

where $W + (c)$ and $W - (c)$ refers to weighted sum of positive tuples and negative tuples meeting the demand of d . W' is other tuples' weighted sum except the tuple sets divided into p . Best split testing can be obtained by finding the minimum Z value.

The optimal construction algorithm of ADTree enunciated by [93] utilizes the Z_{pure} pruning technology as:

$$Z_{pure} = 2(\sqrt{W_+} + \sqrt{W_-}) + W, \quad (2)$$

where Z_{pure} represents the low limit of Z utilized for evaluating the predictive nodes.

4.2. Meta/Ensemble Classifiers

4.2.1. Bagging Ensemble Classifier

Ensemble model combines various base models to produce a more optimal predictive model than single decision tree classifier. The main idea of ensemble model is to combine several weak learners (bootstrapping) to a strong learner (aggregation) for enhancing the predictability of the model. This model helps to minimize the biasness, noise and variance errors. AdaBoost, random forest and bagging are some of the random subspaces used in ensemble models. These techniques have now been utilized for groundwater potential analysis, landslide and flood susceptibility analysis (Chen et al. 2019) [18]. In AdaBoost model inaccuracy arises as it ignores the remaining data by concentrating on the difficult one which leads to a large range of diversity in the performance of bagging [94]. However, bagging ensemble can effectively be utilized for landslide susceptibility and has better prediction power than the conventional models [95].

4.2.2. Random Subspace Ensemble Classifier

In the time of pattern recognition, machine learning classifier is one of the topics of interest among researchers [94]. Random subspace ensemble model comprises several classifiers in a data feature space. Random subspace ensemble classifier can be used by nearest neighbor, linear, support vector and by other classifiers [67]. The advantage of this model is that training data seems to be smaller for original data which is larger for subspace data.

4.2.3. Rotation Forest Ensemble Classifier

There are several methods used for landslide susceptibility analysis but none of them are perfect [70]. The accuracy of the landslide susceptibility can only be achieved using the combination of ensembles classifiers [63]. Rotation forest ensemble approach first introduced by Rodriguez et al. [94] focuses on inducing the diversity and individual accuracy within the ensembles [94]. For creation of the training set, principle component analysis (PCA) was used to extract the features. The success of this model is based on the rotation matrix which is formed by the base classifier and the transformation method [63] (Figure 4).

4.2.4. Selecting the Most Important Conditioning Factors Using Chi-Square Attribute Evaluation (CSEA) Technique

Feature selection techniques, which have been more widely used in artificial intelligence, select a small features set of the training dataset for reducing the cost and time of modelling process as well as producing acceptable results during the modelling process [96]. There are some feature selection techniques such as gain ratio (GI), information gain ratio (IGR), least square support vector machine (LSSVM), chi-square attribute evaluation (CSEA), correlation-based feature selection (CFS), fast correlation-based feature selection (FCBF), Euclidean distance, i -test, principal component analysis (PCA), and Markov blanket filter [97]. In this study, the chi-square attribute evaluation (CSEA) technique was used. The CSEA is calculated according to the following formula:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

where E is expected values and O is actual/observed values. The higher the value of the chi-square for a given conditioning factor in feature selection techniques, the more importance for landslide incidence.

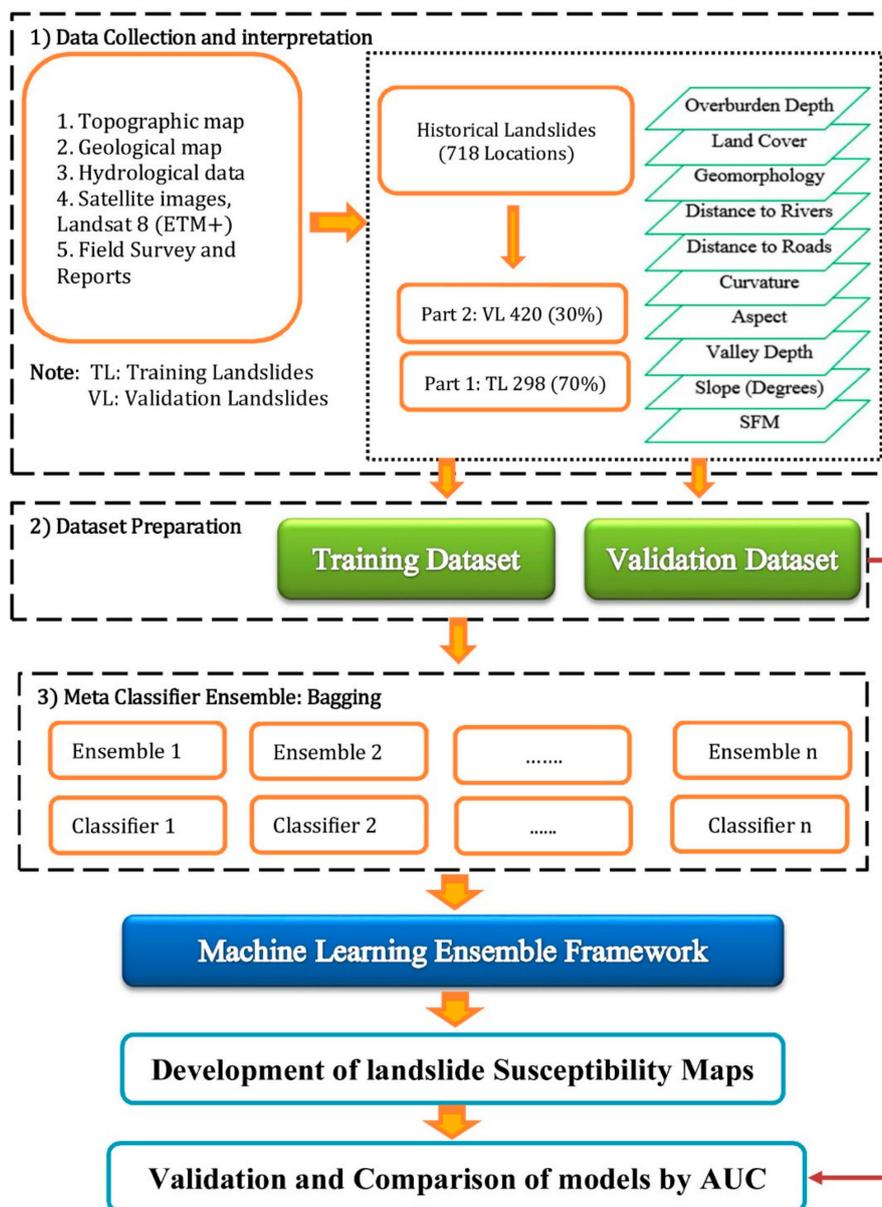


Figure 4. The flowchart of modeling process of the current study.

4.2.5. Model Validation and Comparison

Although there are some evaluation measures to validate the performance of the machine learning models, in this study TP rate, FP rate, recall, precision, F-measure and ROC were used. All these measures can be computed from the confusion matrix (Table 2). It consists of four elements including (A) true positive (TP); (B) false positive (FP); (C) false negative (FN); and (D) true negative (TN) [98].

$$TPRate = A/P = TP/P \quad (4)$$

$$FPRate = B/N = FP/N \quad (5)$$

$$Precision = A/(A + B) = TP/TP + FP \quad (6)$$

$$F\text{-measure} = 2/(1/Precision + 1)/Recall \quad (7)$$

Table 2. Confusion matrix and evaluation measures.

		Actual Class	
		1	0
Predicted Class	1	A (TP)	B (FP)
	0	C (FN)	D (TN)
Column total:		P	N

Another measure for evaluation of the performance of the models is receiver operating characteristic (ROC) curve. It is plotted by recall (sensitivity) and 100-specificity on the *x*- and *y*-axis, respectively [13,99]. According to the definition, specificity is the number of incorrectly classified landslide cells per total predicted non-landslide cells [55]. The area under the ROC curve (AUC) generally has been used to evaluate model performance. The AUC for an ideal and inaccurate model have the values of 1 and 0.5, respectively [69]. The AUC is calculated as follows:

$$AUC = (\sum TP + \sum TN) / (P + N) \tag{8}$$

where P and N are the total number of gullies and non-gullies, respectively.

5. Results and Analysis

5.1. The Most Significant Conditioning Factors

The predictive merit of the landslide susceptibility affecting factors with CSAE method is shown in Figure 5. The conditioning factors with higher than zero average merit (AM) values indicate contribution to landslide models. Conditioning factor selection findings revealed that the curvature factor had no effect on landslide susceptibility modelling, because its average merit (AM) was zero; hence, it was not entered in the modelling process. The CSAE method also showed that other nine conditioning factors were capable of landslide susceptibility modelling (AM > 0). Land cover has the highest predictive merit for landslide susceptibility modelling (AM = 234.285). It is followed by geomorphology (AM = 150.886), valley depth (AM = 131.336), SFM (AM = 116.89), aspect (AM = 75.479), distance to river (AM = 67.457), slope (AM = 43.585), depth (AM = 19.609), and distance to road (AM = 11.627).

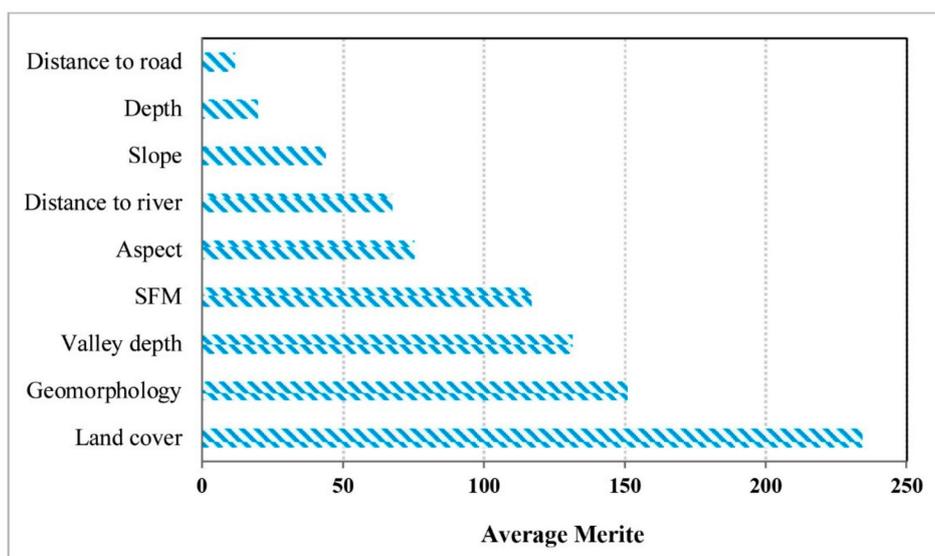


Figure 5. The most important factors by chi-square attribute evaluation (CSEA) technique.

5.2. Landslide Modelling, Evaluation and Comparison

The number of seeds and iterations can affect the landslide model performance. In order to select the optimal values a trial and error procedure has been carried out with varying numbers of seeds and iterations versus AUROC using both the training and validation data. The results showed that the best performances of RSADT model (AUC = 0.915) for the validation dataset were obtained with the number of iterations and seeds equal to 14 and 7, respectively (Figure 6a,b). Also, it can be concluded that the maximum performance (AUC) of BAADT model in the validation step was determined as 0.919 since the number of iteration equal to 20 (Figure 6c) and for the number of seed equal to 3 (Figure 6d). From Figure 6e and f. It can be observed that the highest AUC value (0.931) of RFADT model for the validation dataset was obtained with number of iterations and seeds equal to 3 and 12, respectively.

The ADTree, BAADT, RSADT and RFADT models were constructed using training data sets. According to statistical performance analysis of models in Table 3, all of the models have shown acceptable performance for landslide position prediction in the training step. Among the four models, the RFADT model has the best performance in term of TP (0.911) and FP (0.100) rate, precision (0.911), AUC (0.972), Kappa (0.815) and RMSE (0.305). It is followed by the BAADT and RSADT models. In addition, the ADTree model was shown to have the lowest performance with TP, FP, precision, Kappa, AUC and RMSE equal to 0.863, 0.131, 0.867, 0.939, 0.722 and 0.326, respectively.

Table 3. Model validation and comparison by training dataset.

	ADTree	BAADT	RSADT	RFADT
TP Rate/Recall	0.863	0.880	0.881	0.911
FP Rate	0.131	0.131	0.112	0.100
Precision	0.867	0.880	0.885	0.911
Kappa	0.722	0.752	0.751	0.815
RMSE	0.326	0.314	0.325	0.305
AUC	0.939	0.954	0.949	0.972

The results of statistical performance criteria in the validation step showed that all of the landslide susceptibility models had acceptable values (Table 4). Out of these, like the training stage, the RFADT model was the best performing model (TP rate = 0.717, FP rate = 0.285, precision = 0.771, AUC = 0.931, Kappa = 0.433, and RMSE = 0.397) and the ADTree model showed the lowest performance (TP rate = 0.717, FP rate = 0.285, precision = 0.771, AUC = 0.931, Kappa = 0.433, and RMSE = 0.397). Therefore, both BAADT and RSADT models had intermediate efficiency between the RFADT and ADTree models.

Table 4. Model validation and comparison by validation dataset.

	ADTree	BAADT	RSADT	RFADT
TP Rate/Recall	0.711	0.714	0.717	0.717
FP Rate	0.291	0.288	0.276	0.285
Precision	0.734	0.773	0.759	0.771
Kappa	0.421	0.427	0.451	0.433
RMSE	0.404	0.400	0.398	0.397
AUC	0.897	0.919	0.915	0.931

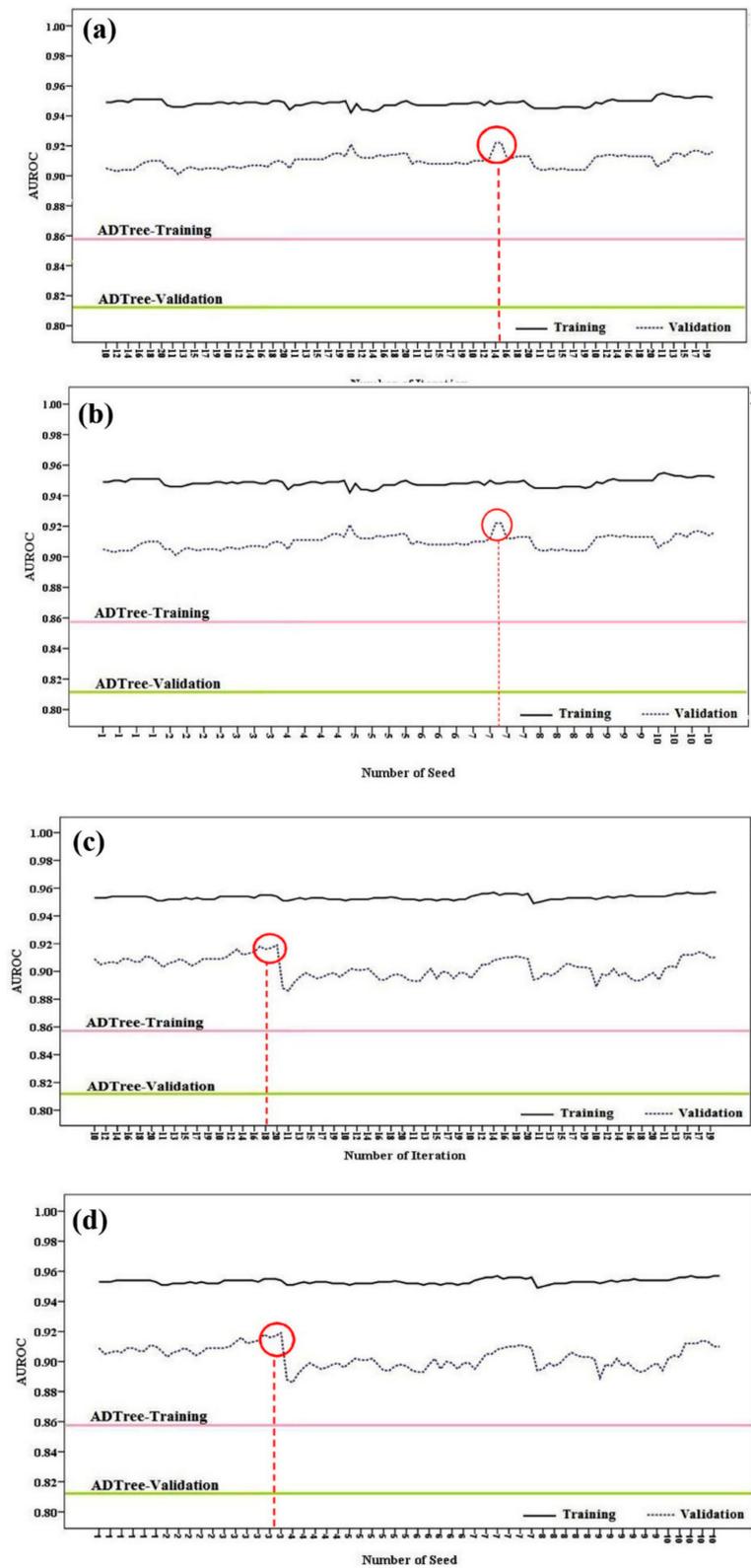


Figure 6. Cont.

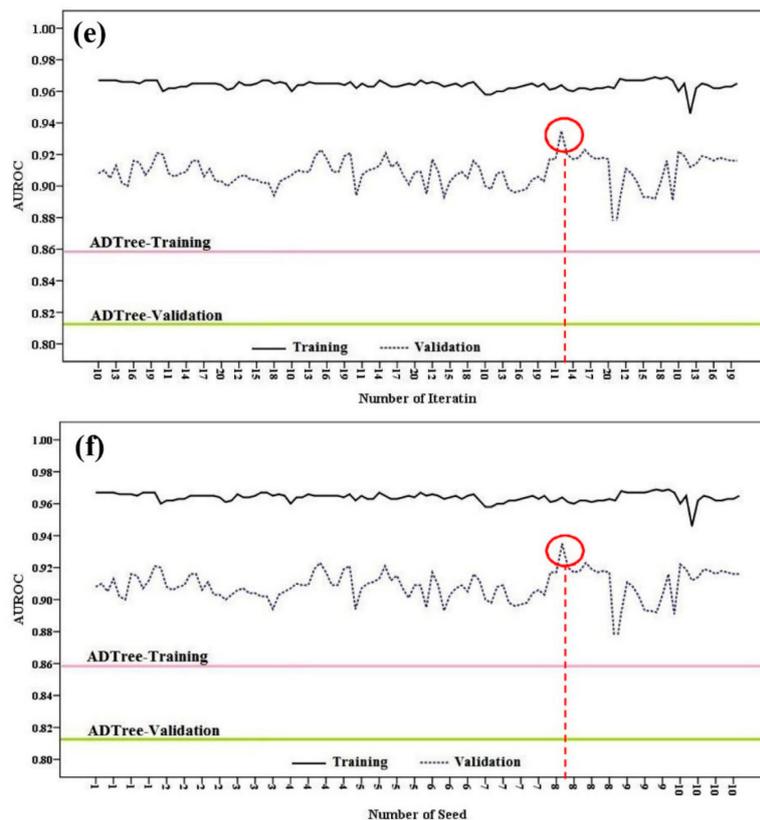


Figure 6. Determination of optimal values of parameters (seed and iteration) used in the new hybrid models. (a) RS-Iteration; (b) RS-Seed; (c) Bagging-Iteration; (d) Bagging-Seed; (e) RF-Iteration; (f) RF-Seed.

5.3. Development of Landslide Susceptibility Maps

After determining the landslide susceptibility index using different models, the entire study area was classified into five susceptibility classes (very low (VLS), low (LS), moderate (MS), high (HS) and very high (VHS)) based on the geometrical interval, natural break and quantile classification schemes. The relative distribution of the susceptible classes in the study area and the contribution of classes in the recorded landslides are shown in Figure 7. Generally, the histograms of all models for different classification methods revealed that most of the recorded landslides are located in very high (VHS) susceptibility classes, except for ADTree model in which high class (HS) had the highest proportion of the recorded landslides. In the case of ADTree model, the very high susceptibility class determined by geometrical interval, natural breaks and quantile schemes cover 15.1, 16.9, and 16.9 percentages of the whole watershed pixels and, 27.4, 29.8, and 29.8 percentages of the recorded landslide pixels, respectively. Therefore, the natural break and quantile schemes were the best methods; however, the quantile was selected as the most appropriate method for landslide susceptibility classification. Accordingly, the quantile method was selected as the best classification method for the BAADT. However, both quantile and geometrical interval were best for the RFADT susceptibility maps for which the quantile was used. Finally, the geometrical method was the most appropriate for classification of the RSADT susceptibility maps.

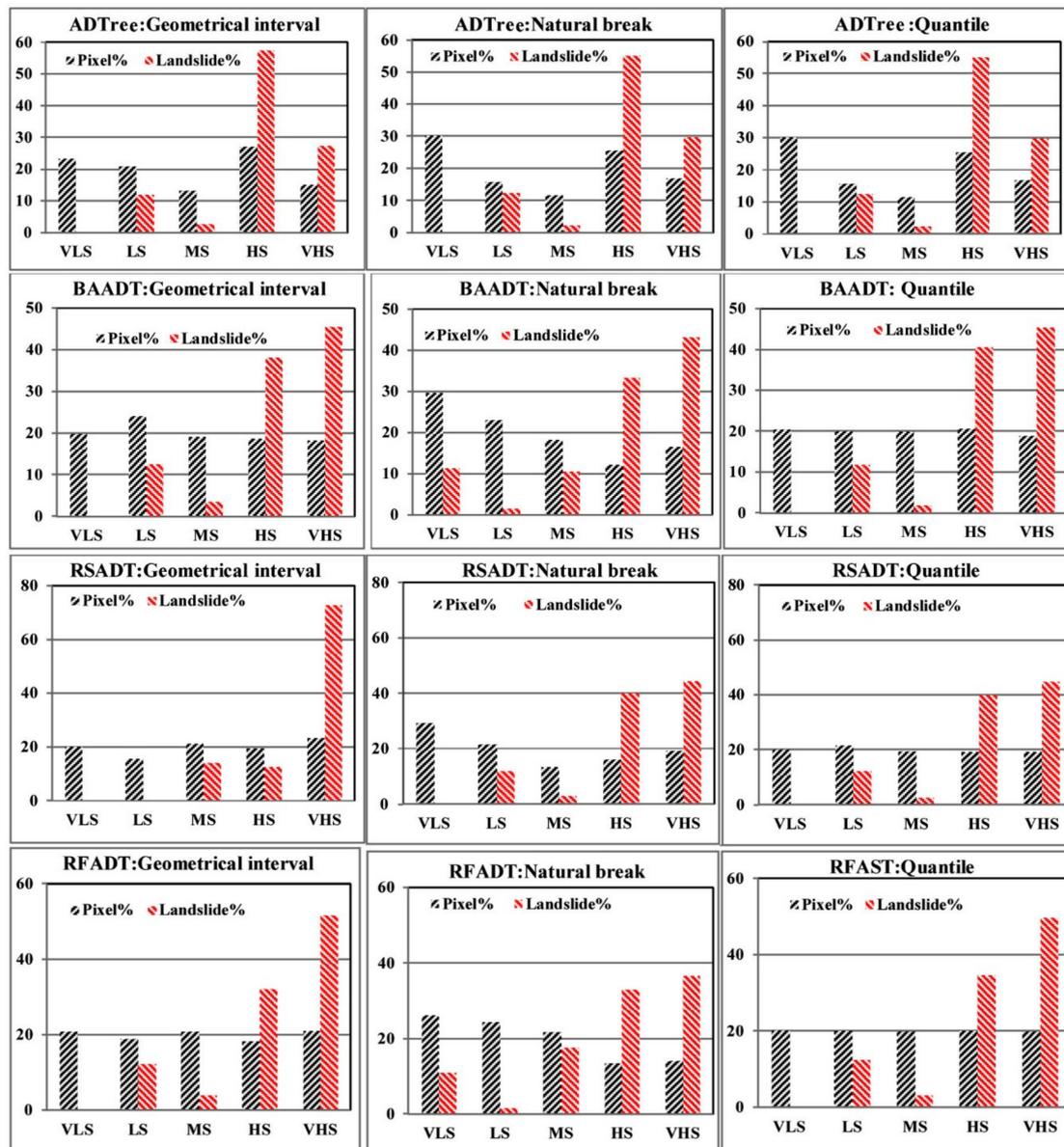


Figure 7. Selecting the best method to classify the landslide susceptibility maps.

The final landslide susceptibility maps were prepared using the selected classification schemes. In the case of ADTree model, the VLS class has the largest area (30.27%) followed by HS (25.45%), VHS (16.9%), LS (15.8%), and MS (11.61%) (Figure 8a). Regarding RFADT as the best performing model, the determining percentage of the study area in VLS, LS, MS, HS and VHS classes were 26.24%, 24.38%, 21.80%, 13.49% and 14.08%, respectively (Figure 8b). For BAADT, the VLS, LS, MS, HS, and VHS classes covered 29.70, 23.13, 13.59, 16.14 and 19.17 percentages of the whole study area, respectively (Figure 8c). Also, based on RSADT model, the VHS has the largest area (23.46%), followed by MS (21.22%), VLS (20.32%), HS (19.29%), and LS (15.71%) (Figure 8d).

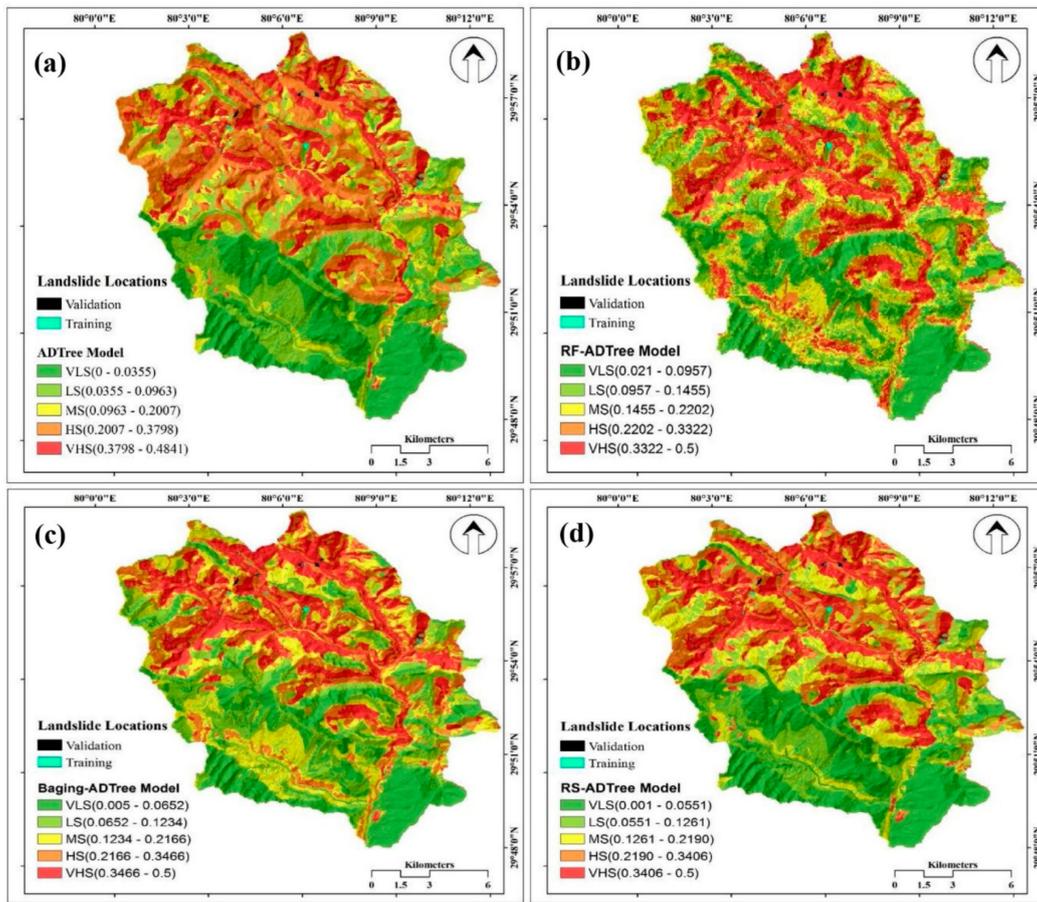


Figure 8. Landslide susceptibility maps: (a) ADTree; (b) RF-ADTree; (c) Baging-ADTree; (d) RS-ADTree.

5.4. Evaluation of the Landslide Susceptibility Maps

The performance of ensemble models in the prediction of landslide susceptibility were compared using the area under the ROC curve (AUC) for both training and validation datasets.

Figure 9a shows the ROC curves of the four landslide susceptibility maps prepared by ADTree, RSADT, RFADT and BAADT models in the training step. The result showed that the highest degree of fit has the RSADT (AUC = 0.883), followed by the RFADT (AUC = 0.871), BAADT (AUC = 0.865), and ADT (AUC = 0.859). From Figure 9b, it can be observed that for the validation step RSADT has the highest area under the curve, with AUC value of 0.842. It is followed by RFADT, BAADT and ADTree with AUC values of 0.840, 0.836, and 0.813, respectively.

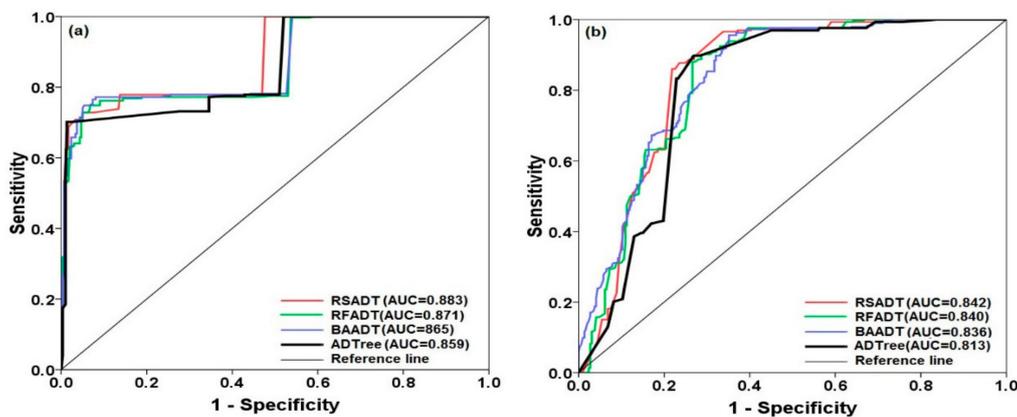


Figure 9. Model validation and comparison by AUC: (a) training datasets, (b) validation dataset.

Comparison of the ROC curves between training and validation steps showed that the AUC values of the training dataset were higher than the validation dataset. This is because of the same landslides that have already been used to construct the landslide models and used for performance analysis in the training step.

6. Discussion

Recognizing regions that are prone to landslide occurrence is one of the most important issues in land management and allocation strategies. Although different methods and techniques have been explored for spatial prediction of landslides over the world, the aims of all these methods are the same. Indeed, achieving a reasonable and reliable susceptibility map of landslides is a debate and controversial subject among landslides researchers. Researchers earlier have been mainly focused on the individual models for spatial prediction of natural hazards such as landslides. On the other hand, recently, most of them focused on the application of the ensemble/hybrid models due to some of their advantages. Basically, the aim of this study is to introduce a new hybrid artificial intelligence for landslide susceptibility mapping at the Pithoragarh region, Uttarakhand in India. This study has been focused on application of some meta-classifier/ensemble algorithms including BA, RS and RF based on a decision tree classifier such as ADTree for spatial prediction of landslide. In the modelling process and analyzing the goodness-of-fit and performance of the ensemble models, ten factors were selected. According to the chi-square attribute evaluation (CSAE) technique, all factors except curvature were effective and used for the final process. Feature selections are intelligent techniques that along with selecting the unimportant factors helps to increase the goodness-of-fit and performance of the models [55]. In this study, the curvature, known to be an ineffective factor, was removed from the final modelling process due to creating noise and over-fitting problems. Hybrid models are powerful techniques for considering the appropriate factors and enhancing the power prediction of base/individual classifiers while decreasing noise and over-fitting problems. Indeed, their results were better visualized and considered when compared with other cutting-edge/soft computing individual algorithms [55].

The ADTree can be considered for classifying binary classes and enhancing the accuracy such that it has produced promising results in spatial prediction of landslide over the world [63,67,90,100]. It is a known fact that the ADTree is an interpretable and robust algorithm against noise in order to provide significant improvement in classification error in comparison to the individual/base decision tree stump classifiers [73]. The ADTree in addition to a classification scheme has a measure of confidence that is known as classification margin. It based on very simple/weak rules represents a majority vote for classification issues. Based on this majority vote, ADTree using the Adaboost/boosting algorithm easily learn alternating trees from the training dataset [73].

After comparing the goodness-of-fit and performance of the ensemble models using TP rate/recall, FP rate, precision, kappa index, RMSE, and ROC indexes, the RS-ADTree model was known as the best model in predicting of landslide modelling. The RS is more efficient in reduction of both variance and bias compared to other ensemble methods. The obtained results are in agreement with Shirzadi et al. [13] and Pham et al. [101] who reported the ability of RS for spatial prediction of landslides and enhancing the accuracy of the base classifier used in their study. However, other ensemble models including RFADT and BAADT were also powerful techniques with higher prediction accuracy than the ADTree as an individual/base classifier. In this study, ADTree was selected as a weak classifier (a classifier with a poor performance), for modelling process of landslide susceptibility. Basically, we developed some novel ensemble models to enhance and improve the performance of the ADTree classifier by developing powerful decision rules.

It is remarkable that the ensemble models may have a different result in combination with decision tree individual/base classifiers. For example, bagging may be useful for perceptrons neural network algorithms and linear discriminant analysis (LDA) for weak and unstable classifiers; bagging and RS may be advantageous for k-nearest neighbors classification rules; and boosting and bagging are

advantageous for linear classifiers [102]. Accordingly, it is possible that in a region not all these ensembles enhance the prediction accuracy of single-based classifiers. For example, Bui et al. [103] have used of functional tree (FT) as a base/weak classifier for developing some ensemble model such as bagging-FT, Adaboost-FT, and multiboost-FT for landslide susceptibility modelling. Their result concluded that Adaboost-FT had lower prediction accuracy than the FT algorithm. Such results also were shown in Bui et al. [104] that indicated Adaboost-DT had the lower prediction accuracy than the DT and bagging-DT models. In this study, all ensemble models performed well and the prediction accuracy were better obtained than the ADTree classifier. Hong et al. [98], Pham et al. [99], and Pham et al. [100] achieved the same results in which their applied ensembles had a better accuracy than the based classifier. Skurichina and Duin [102] demonstrated that the RS may have a better performance than the boosting and bagging algorithms which are useful for unstable classifiers [105]. They also confirmed that bagging is not useful for linear classifiers because they are mainly stable. Additionally, they reported that bagging for very small and also for very large training sample sizes is not usually appropriate.

The advantage of bagging is the shifting effect on the generalization error of the base classifier in the direction of generalization error computed on smaller training datasets. Therefore, it is applicable for classifiers that having a decreasing learning curve. On the other hand, the RF is a robust classifier with low bias and noise that causes an enhancement of the accuracy of individual/base classifiers and also the diversity in the ensemble at the same time [94]. The RS are useful meta-classifiers for weak linear classifiers which have been obtained from a small and critical training dataset. However, the efficient dimensionality (disadvantages) of RS depends on the level of redundancy in the feature space of the training dataset [102]. The above-mentioned advantages of the ensemble models prepared reasonable landslide susceptibility maps with high prediction accuracy in comparison to use of a weak classifier.

7. Conclusions

According to the obtained results from the hybrid machine learning algorithms in literature, they are more strongly and robustness than other methods and techniques for spatial prediction of landslide and hence are more favorable among landslide researchers. Since each classifier/algorithm has a different probability distribution function and structure, the output from modelling will be different due to uncertainties from the model and inputs. In the landslide modelling by machine learning, the performance and prediction accuracy will be enhanced when the proper meta/ensemble classifier is tested and selected. This result will be obtained when a training dataset with low noise and over-fitting problems and high performance and goodness-of-fit is selected. In this study, among ten conditioning factors, curvature had error and noise in the training dataset and was removed from the modeling, while land cover was the most significant factor for landslide occurrence in the study area. Three meta-classifiers including BA, RS and RF in this study were used for combination with ADTree as a weak base classifier to construct hybrid models. Our findings based on several statistical metrics pointed out that the RS-ADTree hybrid model outperformed the BA-ADTree and RF-ADTree models. This model was more able to overcome bias and over-fitting problems, resulting in higher prediction accuracy. Therefore, we conclude that the RS-ADTree ensemble model can be used as a new promising technique for spatial prediction of landslides in the study area. The RF-ADTree and the BA-ADTree models are other proper models for landslide susceptibility mapping. We suggest that to check the applicability and efficiency of these models, more case studies with different climate and geo-environmental factors should be used and validated. We believe that achieving a landslide susceptibility map with reliable and high prediction accuracy, which is the main aim of landslide researchers, may be useful and constructive for decision making, enabling better management of landslide prone areas.

Author Contributions: B.T.P., A.S., H.S., E.O., S.K.S., M.S., D.T.A., B.B.A., N.K.Q., and S.L. contributed equally to the work. B.T.P., A.S., H.S., and N.K.Q. collected field data and conducted the landslide susceptibility mapping

and analysis. B.T.P., A.S., H.S., E.O., S.K.S., D.T., and N.K.Q. wrote the manuscript. B.T.P., H.S., M.S., B.B.A., and S.L. provided critical comments in planning this paper and edited the manuscript. All the authors discussed the results and edited the manuscript.

Funding: This research was supported by the Basic Research Project of the Korea Institute of Geoscience, Mineral Resources (KIGAM) funded by the Minister of Science and ICT and Universiti Teknologi Malaysia (UTM) based on Research University Grant (Q.J130000.2527.17H84).

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

References

1. Cruden, D.M. A suggested method for a landslide summary. *Bull. Int. Assoc. Eng. Geol.* **1991**, *43*, 101–110.
2. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [[CrossRef](#)]
3. 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](#)]
4. Dilley, M.; Chen, R.S.; Deichmann, U.; Lerner-Lam, A.L.; Arnold, M. *Natural Disaster Hotspots: A Global Risk Analysis*; The World Bank: Washington, DC, USA, 2005.
5. Kirschbaum, D.; Stanley, T. Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. *Earth's Futur.* **2018**, *6*, 505–523. [[CrossRef](#)]
6. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [[CrossRef](#)]
7. Klose, M.; Maurischat, P.; Damm, B. Landslide impacts in Germany: A historical and socioeconomic perspective. *Landslides* **2016**, *13*, 183–199. [[CrossRef](#)]
8. Li, T.; Wang, S. *Landslide Hazards and Their Mitigation in China*; Science Press: Beijing, China, 1992.
9. Highland, L.M.; Godt, J.; Howell, D.; Savage, W. *El nino 1997-98; Damaging Landslides in the San Francisco Bay Area*; US Dept. of the Interior, US Geological Survey, National Landslide: Denver, CO, USA, 1998; pp. 2327–6932.
10. Kuriakose, S.L.; Sankar, G.; Muraleedharan, C. History of landslide susceptibility and a chorology of landslide-prone areas in the western Ghats of Kerala, India. *Environ. Geol.* **2009**, *57*, 1553–1568. [[CrossRef](#)]
11. NASA. Global Landslide Catalog. Available online: <https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog/h9d8-neg4#About> (accessed on 30 March 2019).
12. Kaur, H.; Gupta, S.; Parkash, S. Comparative evaluation of various approaches for landslide hazard zoning: A critical review in Indian perspectives. *Spat. Inf. Res.* **2017**, *25*, 389–398. [[CrossRef](#)]
13. Shirzadi, A.; Bui, D.T.; Pham, B.T.; Solaimani, K.; Chapi, K.; Kavian, A.; Shahabi, H.; Revhaug, I. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* **2017**, *76*, 60. [[CrossRef](#)]
14. Pourghasemi, H.R.; Yansari, Z.T.; Panagos, P.; Pradhan, B. Analysis and evaluation of landslide susceptibility: A review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arab. J. Geosci.* **2018**, *11*, 193. [[CrossRef](#)]
15. Mousavi, S.Z.; Kavian, A.; Soleimani, K.; Mousavi, S.R.; Shirzadi, A. GIS-based spatial prediction of landslide susceptibility using logistic regression model. *Geomat. Nat. Hazards Risk* **2011**, *2*, 33–50. [[CrossRef](#)]
16. Shirzadi, A.; Saro, L.; Joo, O.H.; Chapi, K. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. *Nat. Hazards* **2012**, *64*, 1639–1656. [[CrossRef](#)]
17. Shahabi, H.; Khezri, S.; Ahmad, B.B.; Hashim, M. Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena* **2014**, *115*, 55–70. [[CrossRef](#)]
18. Chen, W.; Sun, Z.; Han, J. Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl. Sci.* **2019**, *9*, 171. [[CrossRef](#)]
19. Shirzadi, A.; Chapi, K.; Shahabi, H.; Solaimani, K.; Kavian, A.; Ahmad, B.B. Rock fall susceptibility assessment along a mountainous road: An evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. *Environ. Earth Sci.* **2017**, *76*, 152. [[CrossRef](#)]

20. Shahabi, H.; Hashim, M.; Bin Ahmad, B. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environ. Earth Sci.* **2015**, *73*, 8647–8668. [[CrossRef](#)]
21. Bourenane, H.; Guettouche, M.S.; Bouhadad, Y.; Braham, M. Landslide hazard mapping in the Constantine city, Northeast Algeria using frequency ratio, weighting factor, logistic regression, weights of evidence, and analytical hierarchy process methods. *Arab. J. Geosci.* **2016**, *9*, 154. [[CrossRef](#)]
22. Shirzadi, A.; Shahabi, H.; Chapi, K.; Bui, D.T.; Pham, B.T.; Shahedi, K.; Ahmad, B.B. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* **2017**, *157*, 213–226. [[CrossRef](#)]
23. Tian, Y.; Xu, C.; Hong, H.; Zhou, Q.; Wang, D. Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ann) models: An example of the 2013 Minxian (China) mw 5.9 event. *Geomat. Nat. Hazards Risk* **2019**, *10*, 1–25. [[CrossRef](#)]
24. Yan, F.; Zhang, Q.; Ye, S.; Ren, B. A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model. *Geomorphology* **2019**, *327*, 170–187. [[CrossRef](#)]
25. Mandal, S.; Mondal, S. Weighted overlay analysis (woa) model, certainty factor (cf) model and analytical hierarchy process (ahp) model in landslide susceptibility studies. In *Statistical Approaches for Landslide Susceptibility Assessment and Prediction*; Springer: Berlin, Germany, 2019; pp. 135–162.
26. Liu, J.; Duan, Z. Quantitative assessment of landslide susceptibility comparing statistical index, index of entropy, and weights of evidence in the Shangnan area, China. *Entropy* **2018**, *20*, 868. [[CrossRef](#)]
27. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.; Panahi, M.; Hong, H. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, Malaysia. *Remote Sens.* **2018**, *10*, 1527. [[CrossRef](#)]
28. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Xie, X. Gis-based landslide susceptibility evaluation using certainty factor and index of entropy ensembled with alternating decision tree models. In *Natural Hazards Gis-Based Spatial Modeling Using Data Mining Techniques*; Springer: Berlin, Germany, 2019; pp. 225–251.
29. Shadman Roodposhti, M.; Aryal, J.; Shahabi, H.; Safarrad, T. Fuzzy shannon entropy: A hybrid gis-based landslide susceptibility mapping method. *Entropy* **2016**, *18*, 343. [[CrossRef](#)]
30. 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](#)]
31. Hong, H.; Shahabi, H.; Shirzadi, A.; Chen, W.; Chapi, K.; Ahmad, B.B.; Roodposhti, M.S.; Hesar, A.Y.; Tian, Y.; Bui, D.T. Landslide susceptibility assessment at the Wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards* **2018**, *1–40*, 173–212. [[CrossRef](#)]
32. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196. [[CrossRef](#)]
33. Wang, Y.; Hong, H.; Chen, W.; Li, S.; Panahi, M.; Khosravi, K.; Shirzadi, A.; Shahabi, H.; Panahi, S.; Costache, R. Flood susceptibility mapping in Dingnan county (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. *J. Environ. Manag.* **2019**, *247*, 712–729. [[CrossRef](#)] [[PubMed](#)]
34. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamawoski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L.; et al. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J. Hydrol.* **2019**, *573*, 311–323. [[CrossRef](#)]
35. Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Bin Ahmad, B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [[CrossRef](#)]
36. Tien Bui, D.; Khosravi, K.; Shahabi, H.; Daggupati, P.; Adamowski, J.F.; Melesse, A.M.; Thai Pham, B.; Pourghasemi, H.R.; Mahmoudi, M.; Bahrami, S. Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sens.* **2019**, *11*, 1589. [[CrossRef](#)]

37. Bui, D.T.; Panahi, M.; Shahabi, H.; Singh, V.P.; Shirzadi, A.; Chapi, K.; Khosravi, K.; Chen, W.; Panahi, S.; Li, S. Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci. Rep.* **2018**, *8*, 15364. [[CrossRef](#)] [[PubMed](#)]
38. Tien Bui, D.; Khosravi, K.; Li, S.; Shahabi, H.; Panahi, M.; Singh, V.; Chapi, K.; Shirzadi, A.; Panahi, S.; Chen, W. New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. *Water* **2018**, *10*, 1210. [[CrossRef](#)]
39. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [[CrossRef](#)] [[PubMed](#)]
40. 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](#)]
41. Rahmati, O.; Samadi, M.; Shahabi, H.; Azareh, A.; Rafiei-Sardooi, E.; Alilou, H.; Melesse, A.M.; Pradhan, B.; Chapi, K.; Shirzadi, A. Swpt: An automated gis-based tool for prioritization of sub-watersheds based on morphometric and topo-hydrological factors. *Geosci. Front.* **2019**, *8*, 47–62. [[CrossRef](#)]
42. Jaafari, A.; Zenner, E.K.; Panahi, M.; Shahabi, H. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric. For. Meteorol.* **2019**, *266*, 198–207. [[CrossRef](#)]
43. Taheri, K.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Gutiérrez, F.; Khosravi, K. Sinkhole susceptibility mapping: A comparison between bayes-based machine learning algorithms. *Land Degrad. Dev.* **2019**, *30*, 730–745. [[CrossRef](#)]
44. Roodposhti, M.S.; Safarrad, T.; Shahabi, H. Drought sensitivity mapping using two one-class support vector machine algorithms. *Atmos. Res.* **2017**, *193*, 73–82. [[CrossRef](#)]
45. Azareh, A.; Rahmati, O.; Rafiei-Sardooi, E.; Sankey, J.B.; Lee, S.; Shahabi, H.; Ahmad, B.B. Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models. *Sci. Total Environ.* **2019**, *655*, 684–696. [[CrossRef](#)]
46. Tien Bui, D.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Omidavr, E.; Pham, B.T.; Talebpour Asl, D.; Khaledian, H.; Pradhan, B.; Panahi, M. A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors* **2019**, *19*, 2444. [[CrossRef](#)]
47. Miraki, S.; Zanganeh, S.H.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Pham, B.T. Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
48. 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](#)]
49. Rahmati, O.; Choubin, B.; Fathabadi, A.; Coulon, F.; Soltani, E.; Shahabi, H.; Mollafar, E.; Tiefenbacher, J.; Cipullo, S.; Ahmad, B.B. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and unec methods. *Sci. Total Environ.* **2019**, *688*, 855–866. [[CrossRef](#)] [[PubMed](#)]
50. Singh, S.K.; Taylor, R.W.; Rahman, M.M.; Pradhan, B. Developing robust arsenic awareness prediction models using machine learning algorithms. *J. Environ. Manag.* **2018**, *211*, 125–137. [[CrossRef](#)] [[PubMed](#)]
51. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.-X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using gis-based machine learning techniques for Chongren county, Jiangxi province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [[CrossRef](#)] [[PubMed](#)]
52. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
53. Thai Pham, B.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Trung Tran, H.; Minh Le, T.; Tran, V.P.; Kim Khoi, D.; Shirzadi, A. A novel hybrid approach of landslide susceptibility modeling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**, 1–25. [[CrossRef](#)]
54. 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](#)]

55. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378.
56. Shafizadeh-Moghadam, H.; Minaei, M.; Shahabi, H.; Hagenauer, J. Big data in geohazard; pattern mining and large scale analysis of landslides in Iran. *Earth Sci. Inform.* **2019**, *12*, 1–17. [[CrossRef](#)]
57. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M. Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* **2019**, *10*, 157. [[CrossRef](#)]
58. Park, I.; Lee, S. Spatial prediction of landslide susceptibility using a decision tree approach: A case study of the Pyeongchang area, Korea. *Int. J. Remote Sens.* **2014**, *35*, 6089–6112. [[CrossRef](#)]
59. Bui, D.T.; Pradhan, B.; Revhaug, I.; Tran, C.T. A comparative assessment between the application of fuzzy unordered rules induction algorithm and j48 decision tree models in spatial prediction of shallow landslides at Lang Son city, Vietnam. In *Remote Sensing Applications in Environmental Research*; Springer: Berlin, Germany, 2014; pp. 87–111.
60. Tsangaratos, P.; Ilija, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* **2016**, *13*, 305–320. [[CrossRef](#)]
61. 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**, *164*, 135–149. [[CrossRef](#)]
62. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Int. Assoc. Eng. Geol.* **2018**, 1–23. [[CrossRef](#)]
63. Pham, B.T.; Bui, D.T.; Prakash, I. Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. *Geotech. Geol. Eng.* **2017**, *35*, 2597–2611. [[CrossRef](#)]
64. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total. Environ.* **2018**, *627*, 744–755. [[CrossRef](#)] [[PubMed](#)]
65. Chen, W.; Zhao, X.; Shahabi, H.; Shirzadi, A.; Khosravi, K.; Chai, H.; Zhang, S.; Zhang, L.; Ma, J.; Chen, Y.; et al. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int.* **2019**, 1–25. [[CrossRef](#)]
66. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Bui, D.T. A Novel Hybrid Approach of Bayesian Logistic Regression and Its Ensembles for Landslide Susceptibility Assessment. *Geocarto Int.* **2018**, 1–44. [[CrossRef](#)]
67. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hoang, N.D.; Pham, B.; Bui, Q.T.; Tran, C.T.; Panahi, M.; Bin Ahmad, B. A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. *Remote Sens.* **2018**, *10*, 1538. [[CrossRef](#)]
68. Pham, B.T.; Shirzadi, A.; Bui, D.T.; Prakash, I.; Dholakia, M. 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](#)]
69. Shirzadi, A.; Soliamani, K.; Habibnejhad, M.; Kavian, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B. Novel gis based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [[CrossRef](#)] [[PubMed](#)]
70. Chen, W.; Shirzadi, A.; Shahabi, H.; Bin Ahmad, 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](#)]
71. Shirzadi, A.; Solaimani, K.; Roshan, M.H.; Kavian, A.; Chapi, K.; Shahabi, H.; Keesstra, S.; Ahmad, B.B.; Bui, D.T. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *Catena* **2019**, *178*, 172–188. [[CrossRef](#)]
72. 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](#)] [[PubMed](#)]

73. Freund, Y.; Mason, L. *The Alternating Decision Tree Learning Algorithm*; ICML: New Jersey, NY, USA, 1999; pp. 124–133.
74. He, Q.; Shahabi, H.; Shirzadi, A.; Li, S.; Chen, W.; Wang, N.; Chai, H.; Bian, H.; Ma, J.; Chen, Y.; et al. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Sci. Total. Environ.* **2019**, *663*, 1–15. [[CrossRef](#)] [[PubMed](#)]
75. Mohammadi, A.; Shahabi, H.; Bin Ahmad, B. Integration of insartechique, google earth images and extensive field survey for landslide inventory in a part of Cameron highlands, Pahang, Malaysia. *Appl. Ecol. Environ. Res.* **2007**, *16*, 8075–8091. [[CrossRef](#)]
76. An, K.; Kim, S.; Chae, T.; Park, D. Developing an accessible landslide susceptibility model using open-source resources. *Sustainability* **2018**, *10*, 293. [[CrossRef](#)]
77. Lee, C.F.; Huang, W.K.; Chang, Y.L.; Chi, S.Y.; Liao, W.C. Regional landslide susceptibility assessment using multi-stage remote sensing data along the coastal range highway in northeastern Taiwan. *Geomorphology* **2018**, *300*, 113–127. [[CrossRef](#)]
78. Martha, T.R.; Roy, P.; Govindharaj, K.B.; Kumar, K.V.; Diwakar, P.; Dadhwal, V. Landslides triggered by the june 2013 extreme rainfall event in parts of Uttarakhand state, India. *Landslides* **2015**, *12*, 135–146. [[CrossRef](#)]
79. Ghosh, S.; Carranza, E.J.M.; van Westen, C.J.; Jetten, V.G.; Bhattacharya, D.N. Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). *Geomorphology* **2011**, *131*, 35–56. [[CrossRef](#)]
80. Prandini, L.; Guidiini, G.; Bottura, J.A.; Pançano, W.L.; Santos, A.R. Behavior of the vegetation in slope stability: A critical review. *Bull. Int. Assoc. Eng. Geol.* **1977**, *16*, 51–55. [[CrossRef](#)]
81. Varnes, D.J. Slope movement types and processes. *Spec. Rep.* **1978**, *176*, 11–33.
82. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M. 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](#)]
83. Yalcin, A.; Reis, S.; Aydinoglu, A.; Yomralioglu, T. A gis-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, ne Turkey. *Catena* **2011**, *85*, 274–287. [[CrossRef](#)]
84. Nefeslioglu, H.A.; Duman, T.Y.; Durmaz, S. Landslide susceptibility mapping for a part of tectonic kelkit valley (eastern black sea region of Turkey). *Geomorphology* **2008**, *94*, 401–418. [[CrossRef](#)]
85. Pham, B.T.; Pradhan, B.; Bui, D.T.; Prakash, I.; Dholakia, M. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
86. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: A case study of Duzkoy district. *Nat. Hazards* **2015**, *76*, 471–496. [[CrossRef](#)]
87. Dehnavi, A.; Aghdam, I.N.; Pradhan, B.; Varzandeh, M.H.M. A new hybrid model using step-wise weight assessment ratio analysis (swara) technique and adaptive neuro-fuzzy inference system (anfis) for regional landslide hazard assessment in Iran. *Catena* **2015**, *135*, 122–148. [[CrossRef](#)]
88. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the three gorges reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23–37. [[CrossRef](#)]
89. 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](#)]
90. 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](#)]
91. Kavzoglu, T.; Colkesen, I.; Sahin, E.K. Machine learning techniques in landslide susceptibility mapping: A survey and a case study. In *Landslides: Theory, Practice and Modelling*; Springer: Berlin, Germany, 2019; pp. 283–301.
92. Nefeslioglu, H.A.; Sezer, E.A.; Gokceoglu, C.; Bozkir, A.S.; Duman, T.Y.; Bozkır, A.S. Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey. *Math. Probl. Eng.* **2010**, *2010*, 1–15. [[CrossRef](#)]

93. Pfahringer, B.; Holmes, G.; Kirkby, R. Pacific-Asia Conference on Knowledge Discovery and Data Mining. In *Optimizing the Induction of Alternating Decision Trees*; Springer: Berlin, Germany, 2001; pp. 477–487.
94. Rodriguez, J.J.; Kuncheva, L.; Alonso, C.J. Rotation Forest: A New Classifier Ensemble Method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [[CrossRef](#)] [[PubMed](#)]
95. Truong, X.L.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, X.Q.; Do, T.H.; Bui, D.T.; Lee, S. Enhancing Prediction Performance of Landslide Susceptibility Model Using Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree. *Appl. Sci.* **2018**, *8*, 1046. [[CrossRef](#)]
96. Vafaie, H.; Imam, I.F. Feature selection methods: Genetic algorithms vs. Greedy-like search. In *International Conference on Fuzzy and Intelligent Control Systems*; Walt Disney World: Orlando, FL, USA, 1994; p. 28.
97. Karegowda, A.G.; Manjunath, A.; Jayaram, M. Comparative study of attribute selection using gain ratio and correlation based feature selection. *Int. J. Inf. Technol. Knowl. Manag.* **2010**, *2*, 271–277.
98. Fawcett, T. An introduction to roc analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
99. Shahabi, H.; Hashim, M. Landslide susceptibility mapping using gis-based statistical models and remote sensing data in tropical environment. *Sci. Rep.* **2015**, *5*, 9899. [[CrossRef](#)]
100. Chen, W.; Xie, X.; Peng, J.; Wang, J.; Duan, Z.; Hong, H. Gis-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomat. Nat. Hazards Risk.* **2017**, *8*, 950–973. [[CrossRef](#)]
101. Pham, B.T.; Bui, D.T.; Pham, H.V.; Le, H.Q.; Prakash, I.; Dholakia, M. Landslide hazard assessment using random subspace fuzzy rules based classifier ensemble and probability analysis of rainfall data: A case study at Mu Cang Chai district, Yen Bai province (Vietnam). *J. Indian Soc. Remote Sens.* **2017**, *45*, 673–683. [[CrossRef](#)]
102. Skurichina, M.; Duin, R.P.W. Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Pattern Anal. Appl.* **2002**, *5*, 121–135. [[CrossRef](#)]
103. Bui, D.T.; 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.
104. Bui, D.T.; Ho, T.C.; Revhaug, I.; Pradhan, B.; Nguyen, D.B. Landslide susceptibility mapping along the national road 32 of Vietnam using gis-based j48 decision tree classifier and its ensembles. In *Cartography from Pole to Pole*; Springer: Berlin, Germany, 2014; pp. 303–317.
105. Breiman, L. Arcing classifier (with discussion and a rejoinder by the author). *Ann. Stat.* **1998**, *26*, 801–849. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).