Eng. Sci., 2022, **17**, 142–155

Engineered Science
DOI: https://dx.doi.org/10.30919/es8d600

Landslide Susceptibility Mapping Using Machine Learning Algorithms for Nainital, India

Pranav Rajendra Rane and Shweta Vincent*

Abstract

The district of Nainital in the state of Uttarakhand in north India has experienced several landslides in the previous decade due to its diverse geomorphology. Several lives have been lost and substantial damage has been caused to livestock and property due to these landslides. Machine Learning algorithms have been used in several studies earlier for the creation of landslide susceptibility maps (LSM) for the classification of regions prone to landslides. This study aims to creation of an LSM using the maximum likelihood (ML), ISO (International Organization for Standardization) clustering, and Random Forest (RF) algorithms. The landslide conditioning factors (LCF) viz., slope angle, slope aspect, curvature, distance from the stream, altitude, and land use land change (LULC) have been analyzed in detail for the generation of the LSM. According to the results of this case study, the highest positive predicted value (PPV%) of 83% is observed for the ISO algorithm with slope angle considered as LCF. The ISO algorithm with distance from stream LCF has the highest negative predicted value (NPV%) of 83%. The ISO algorithm with distance from stream LCF has the highest sensitivity of 84% and the ISO algorithm considering slope angle as LCF has the highest specificity of 84%. A Kappa index of 0.79 has been computed for ISO clustering using slope angle as LCF. It has also been observed that, of all the values of frequency ratio for the various classes of LCFs, the slope angle gives the highest value of 3.4 for the class of 25 – 35 degrees. Further, the ISO algorithm outperforms the ML and RF algorithms in terms of Area Under Curve (AUC) with a value of 0.817.

Keywords: Maximum Likelihood; ISO Clustering; Random Forest; Landslide Susceptibility Mapping; Landslide Conditioning Factors.
Received: 11 August 2021; Revised: 22 November 2021; Accepted: 23 November 2021.
Article type: Research article.

1. Introduction

Landslides are a natural phenomenon in which two layers of the earth slide over one another causing soil erosion and altering the landscape. Landslides could lead to a depletion in resources of soil[1,2] further leading to damages to human property, and agricultural fields and general degradation of the economics of the land.[3-5]

Consequently, it is very essential to prepare a landslide susceptibility map to cause early detection and mitigate the risks caused by landslides. Prior knowledge of methods to control land degradation helps in planning strategies for development in regions prone to landslides.[6,7]

Nainital is a hill station in the Northeastern state of Uttarakhand, India which is prone to several landslides.[8] Due to the mountainous terrain, and a rapid increase in urbanization, a modification of the landscape of Nainital has resulted in an increased instability of the slopes of the hills of Nainital (Fig. S1 in supplementary material). The problem has further escalated causing social and economic distress in the region along with alteration to the geomorphology of the environment. Table 1 illustrates the extent of the impact of landslides on Nainital over the years.

The modeling of susceptibility to a landslide is performed using different approaches viz., using physical models, statistical correlation models, and machine learning models. In the case of physical modeling of landslides, a detailed map of the geography of the region along with the geological aspect is required.[9,10] This is not practically feasible when the site under consideration is huge. In the case of statistical models, classification is performed using landslide conditioning factors which bear a key importance in the quality of the landslide susceptibility map which gets created.[11] Tilahun *et al.* have performed a bivariate statistical analysis of GIS (Geographical Information System) data on Shimada,

Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India.
**E-mail:* shweta.vincent@manipal.edu (S. Vincent)

Ethiopia.[12] They have made use of the parameters of frequency ratio and weight of evidence to compare the landslide causative factors and prepare a landslide susceptibility map for the area.

Table 1. The extent of the impact of landslides in Nainital from 2010 to 2017 (Source Geological Survey of India).

Loss Type	2010	2012	2013	2016	2017
Human deaths	220	176	225	119	84
Missing humans	-	-	4021	5	27
Injured humans	139	96	238	102	66
Farm animals	1798	997	11268	1391	1020

The era of Machine learning has brought in many algorithms which can help build a correlation model between the landslide conditioning factors and the actual occurrence of a landslide. Bui *et al*. have showcased the usage of support vector machines (SVM), artificial neural networks (ANN), and logistic regression (LR) for the accurate prediction of landslides in certain areas of Vietnam.[13] Their results showcase the superiority of ANN in the accurate prediction of landslides and the creation of susceptibility maps. An exhaustive survey has been performed on the application of machine learning for the early prediction and prevention of landslides.[14] Achour *et al.* have presented the usage of SVM, Random Forest (RF), and Boosted Regression Tree (BRT) for the prediction of landslides in the Ain Bouziane region of Algeria.[15] Their results show that the RF algorithm gives the highest prediction accuracy in comparison to the rest. Hence, their study reinforces the idea of the usage of machine learning algorithms for the effective and accurate prediction of landslides in various regions. Table 2 gives a detailed analysis of recent research in the prediction of landslides using various machine learning algorithms and their relative results.

Table 2. Summary of studies conducted using Machine Learning Algorithms for prediction of Landslides.

S. No.	Ref. No.	Location	Image Resolution	Dataset used	Evaluation parameter	Algorithm used	Parameter values	
1	Pham *et al.*[16]	Uttarakhand, India	20 m × 20 m	Google Earth images	Prediction accuracy, RMSE		RMSE	Prediction accuracy
						SVM	0.279	88.72
						LR	0.325	85.53
						Fisher's LDA	0.329	85.41
						BN	0.343	84.37
						NBC	0.355	83.89
2	Chen *et al.* [17]	Long county, Shaanxi province, China	30 m	Landsat 8, Google Earth Pro 7.1	Prediction Accuracy and AUC		AUC	Prediction accuracy
						LMT	75.2	71.7
						RF	78.1	79.5
						CART	74.2	74.6
3	Dehnavi *et al.*[18]	Iran	NA	NA	Prediction Accuracy and AUC	SWARA and ANFIS	AUC	Prediction accuracy
							84	80
4	Althuwaynee *et al.*[19]	Inje, Korea	5 m	Daum Web portal, UltraCAM X sensor	AUC		AUC	
						CAID	76	
						AHP	80	
5	Lee *et al.*[20]	PyeongChang, Korea Inje, Korea	10 m × 10 m	Daum Web portal, UltraCAM X sensor	AUC		AUC	
						SVM	81.36	
6	Kavzoglu *et al.*[21]	Duzkoy, Turkey	30 m	LANDSAT ETM+	Prediction Accuracy and AUC		AUC	Prediction accuracy
						SVM	98.5	94.43
						LR	98.4	93.85
						DT	98	89.82
						FR	92.1	85.07
						SI	91.1	82.48
						WoE	89.3	81.88

Continued

7	Ermini *et al.*[22]	Riomaggiore Catchment, Italy	5m	NA	Standard Error		Standard Error	
						MLP	14	
						PNN	19	
8	Chen *et al.*[23]	Langao County	NA	NA	AUC		AUC	
						SWARA-ANFIS-PSO	89	
						SWARA-ANFIS-SFLA	89	
9	Kalantar *et al.*[24]	Dodangeh watershed, Mazandaran province, Iran	10 m	Satellite images	AUC		AUC	
						SVM	81.42	
						LR	79.82	
						ANN	70.2	
10	Aditian *et al.*[25]	Ambon Island, Indonesia	30 m	Google Earth	AUC		AUC	
						FR	66.8	
						LR	66.7	
						ANN	71.7	
11	Trigila *et al.*[26]	Giampieleri, Italy	1 m to 20 m	LIDAR survey	AUC		AUC	
						LR	85.1	
						RF	77.2	
12	Wang *et al.*[27]	Yunyang County, China	30 m	ASTER satellite	AUC		AUC	
						RF	99.2	
						FR	60	
13	Lee[28]	Sagimakri area, Korea	6.6 m	KOMPSAT -1	AUC		AUC	
						WoE	79.4	
14	Mersha *et al.*[29]	Simada area, Ethiopia	NA	Google Earth data	AUC		AUC	
						WoE	88.2	
						FR	84.8	
15	Jaafari *et al.*[30]	Caspian forest, Iran	NA	Aerial photos	AUC		AUC	
						IoE	75.9	
						FR	72.6	
16	Dickson *et al.*[31]	Auckland, New Zealand	NA	LiDAR	Prediction Accuracy		Prediction Accuracy	
						MEM	93.1	
						BRT	95.3	
						CART	89	
17	Hong *et al.*[32]	Linahua area, China	25 m × 25 m	NA	AUC		AUC	
						RF	81.2	
						EBF	81.3	
						FR	77.5	
						LR	71.7	
18	Youssef *et al.*[33]	Abha Basin, Saudi Arabia	0.5 m	GeoEye and Google Earth data	AUC, RMSE		AUC	RMSE
						RF	95.1	0.36
						LDA	94.1	0.37
						ANN	93.4	0.45
						SVM	93	0.39
						MARS	91.8	0.07
						NB	91.6	0.46
						QDA	89.9	0.46

Continued

19	Aghdam *et al.*[34]	Provinces near Alborz Mountains, Iran	90 m × 90 m	GeoEye and Google Earth data	AUC	Wi-ANFIS	AUC 89
20	Youssef *et al.*[35]	Fayfa mountains, Saudi Arabia	0.5 m and 0.6 m	GeoEye and QuickBird	AUC	FR LR Ensemble FR and LR	AUC 58 77 82
21	Deng *et al.*[36]	Pinggu district, Beijing	0.5 m	Google Earth data	AUC	FR CF	AUC 76.9 76.8
22	Park *et al.*[37]	Inje, Korea	0.5 m	NA	AUC	FR AHP LR ANN	AUC 79.4 78.9 79.4 80.6
23	Shahabi *et al.*[38]	Zab Basin, Iran	30 m	Landsat	AUC	FR LR Fuzzy logic	AUC 94.6 96.4 85.1 to 94.6
24	Devkota *et al.*[39]	Mugling-Narayanghat road, Nepal	NA	Aerial photos	AUC	IoE LR CF	AUC 90.1 86.2 83.5
25	Kumar *et al.*[40]	Tehri reservoir, Uttarakhand	30 m	ASTER, WorldView -2, Cartosat-1	AUC	AHP	AUC 78.7

2. Materials and methods

2.3.1 Description of the study area

The study area chosen is the district of Nainital, which has an area of about 4,251 sq. km. It is located in the southern part of the mountainous area of Uttarakhand and is a part of the Kumaon region. It lies approximately between 80º14' and 78º80' E and 29º00' and 29º05' N. It is surrounded by the Almora district to the north and the Udham Singh Nagar district to the south.

The region has a tight intersecting network of various rivers and smaller water bodies. Some of the major waterbodies are the Kosi, Gola, and Nandhaur rivers. The Kosi starts from the Almora district and flows in the southern direction towards Someswar and then towards the southeast of Almora and then join Nana Kosi and the Sual. From its junction with Sual, it forms a boundary between the Almora and Nainital districts. The Gola River starts in Almora and the Nandhaur river starts from the southern slopes of the lower hills in Patti Chandigarh. It flows south past Chorgallia through the Bhabar and Terai. Fig. S2 in the study's supplementary material showcases the map of the Nainital district.

There are various lakes present mostly in the Pargana Chhakhata district. The important lakes of the district are - Nainital, Bhimtal, Malwa Tal, Sat Tal, Naukuchhiya Tal, and Khurpa Tal. The lakes are one of the most attractive features of the Himalayan Region and some of them are used for irrigation purposes. The climate of the area depends upon the altitudes of the place; hence weather conditions and rain density vary from area to area. The summer usually lasts from March to June end. The monsoon season starts in July and lasts till mid-September. The period between September to November is the post-monsoon or autumn season and from December to February is considered the winter season. High humidity is found during the monsoon season and less in the winter months. The average daily maximum temperature varies between 23 ºC and 24 ºC in the mountainous terrain, while the mean daily maximum temperature is 39 ºC and the minimum is 27 ºC in the plains during the same season.

The land use and land cover (LULC) maps for Nainital have been generated using the ArcGIS (Aeronautical) Reconnaissance Coverage Geographic Information System tool with the aid of the ISO clustering algorithm. The aim of LULC map generation is to notice the changes in the

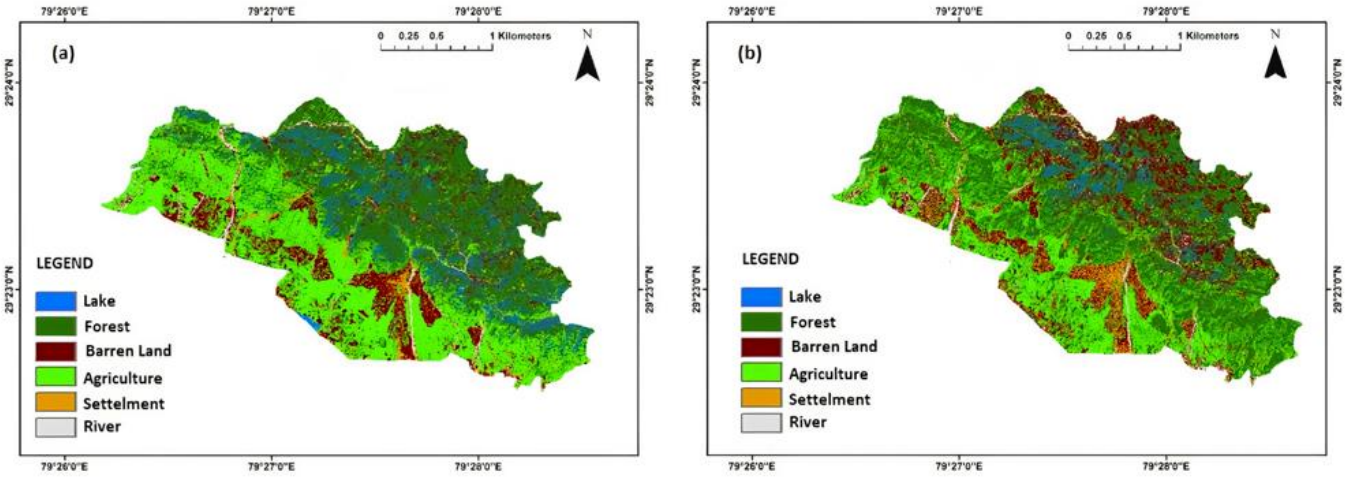

Fig. 1 (a) LULC map for 2010 (b) LULC map for 2015.

geological features of the region of interest. The LULC maps have been generated for 2010 (Fig. 1a) and 2015 (Fig. 1b). Both the maps have been generated to classify land features viz., Forest, Lake, Settlement, Barren land, Agriculture, and River. It is clear from both Figs. 1a, and 1b that within a span of five years, barren land has increased in the Nainital region with a corresponding increase in human settlement area and degradation in forest cover. This results in soil erosion during the monsoons and thereby leads to landslides. A major landslide occurred in 2013 (Refer to Table 1) which caused damages to both lives and property.

2.3.2 Methodology followed

In our study of the region of Nainital, the landslide inventory map (LIM) has been downloaded for the region from the Bhukosh portal provided by the Geological Survey of India at a scale of 1:100000. Geological data of various regions in India can be downloaded from this portal. Fig. 2 illustrates the LIM of Nanital where the red dots represent the regions of active landslides.

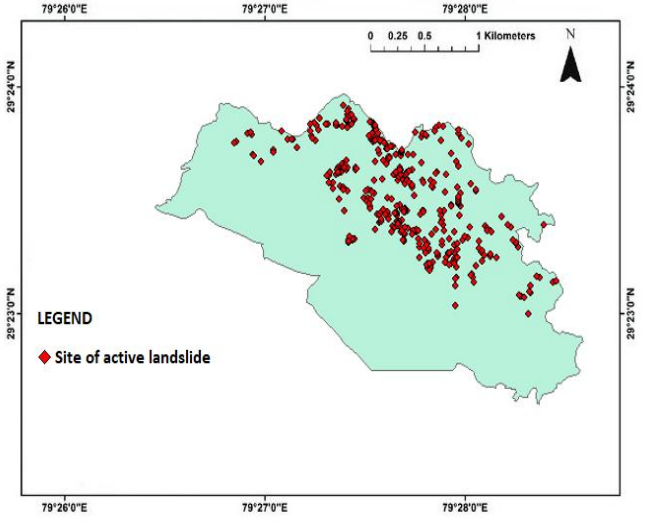

Fig. 2 Landslide Inventory Map of Nainital.

From the LIM obtained, the training and testing data are split based on the LCFs of slope angle, slope aspect, curvature, and distance from the stream. The corresponding datasets are trained using the ISO, ML, and RF algorithms for two classes of data i.e., Landslide positive and Landslide negative classes. The corresponding confusion matrices that are obtained are used to compute the classification parameters viz., accuracy, sensitivity, specificity, positive predicted value, and negative predicted value. Fig. 3 showcases the flowchart of the methodology followed.

2.3.3 Landslide conditioning factors and their analysis

The different causative factors which result in the triggering of landslides are termed Landslide Conditioning factors (LCFs). The identification of such factors which cause landslides is of prime importance for early detection and prediction. Major LCFs are slope aspect, slope angle, curvature, distance from the stream, distance to road, precipitation, and roughness to name a few. The following section describes primary landslide conditioning factors which play a vital role in triggering landslides.

2.3.4 Landslide conditioning factors

The development of an effective landslide susceptibility map (LSM) depends on the choice and evaluation of the impact of landslide causative/ conditioning factors. This section describes various landslide conditioning factors taken into consideration for our study.

The slope angle of the terrain: The slope angle of the terrain determines the angle of the presence of water below the surface which leads to the increased moisture content in the soil. This phenomenon indirectly leads to landslides.

Aspect angle of the terrain: The aspect angle specifies the direction of the region of interest in terms of the geographical directions of north, south, east, and west. Aspect determines factors such as exposure to sunlight, the intensity of winds, precipitation, *etc*. experienced by the region of interest.

The curvature of the terrain: Plan curvature of the terrain

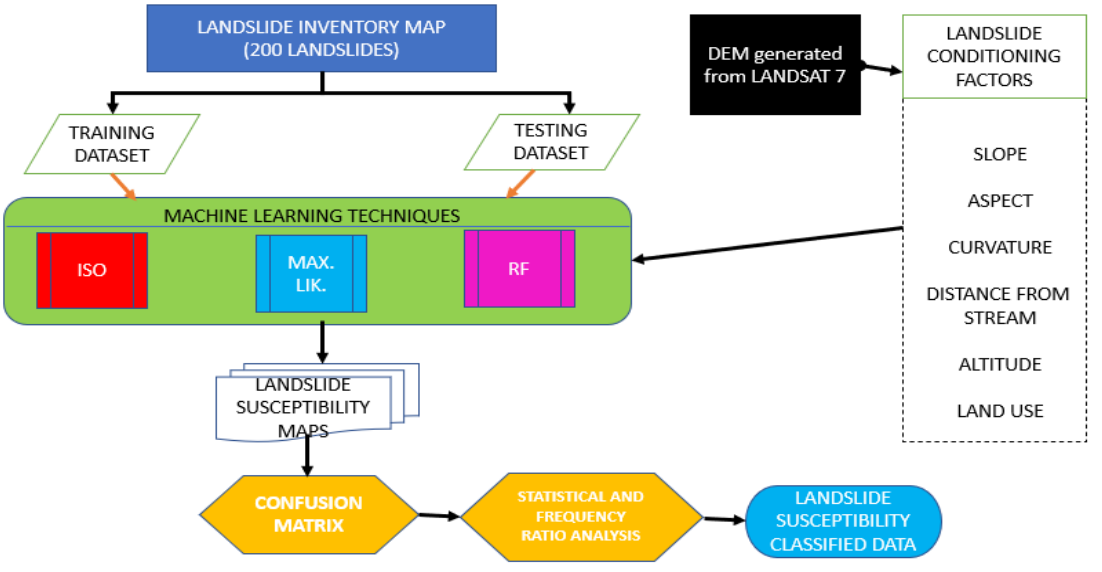

Fig. 3 Methodology followed.

signifies the counter line curvature made by a horizontal plane when it crosses the surface. The process of soil erosion due to the convergence of water at different profiles of curvature may lead to landslides.

Distance from the stream: This factor, as the name suggests, determines the distance of the region of interest from a stream/river.

Altitude: The altitude of the terrain with respect to mean sea level is affected by various geomorphologic and geological processes.

Land use: Land use is a vital factor the in determining the causes behind landslides. It is used to classify the land into separate entities such as lakes, rivers, vegetation, forests, barren land, *etc*. The gradual change in the land use land cover (LULC) map of Nainital of 2010 and 2015 has been shown in Figs. 1a and b respectively.

Lithology: The study of the strength of the rocks and soil is termed Lithology. It is also an important factor to be considered for the creation of LSM as different types of rocks in the region of interest have varied responses to landslide triggers.

Sediment transport index (STI): This index is used to create a relation between the slope of the region of interest and the water accumulation in the catchment area. It is denoted by Equation 1.

$$STI = \left(\frac{UA}{22.1}\right)^{0.6}\left(\frac{\sin SA}{0.09}\right)^{1.3} \quad (1)$$

where:

UA refers to the upper catchment area

SA refers to the slope angle in radians.

Wetness Index (WI): A reduction in the stability of soil is caused by the accumulation of moisture and precipitation in the soil. This is theoretically represented using the WI which is shown in Equation 2.

$$WI = log_e\left(\frac{FA}{w\ tanSA}\right) \quad (2)$$

where:

FA denotes the flow accumulation in meter square

w denotes the pixel width through which the water flows in meters SA denotes the slope angle in radians studies[41,42] that have been carried out to determine primary LCFs which however change from one region to another. In our study, the choice of these factors is based on Information Gain Ratio (IGR) with ten-fold cross-validation. Out of all the LCFs the ones which produced null predictive probabilities have been removed.

2.3.5 Factor selection based on correlation analysis

The concept of multicollinearity exists in the study of landslides where the datasets have LCFs with high correlation. This could result in an erroneous analysis of the data.[43] The Pearson coefficient of correlation between two LCFs (*e.g.* slope angle (S) and slope aspect (A) is computed as shown in Equation 3. High collinearity is indicated by $r_{S.A}$ having a value greater than 0.7.[44] Here, $\bar{S}$ and $\bar{A}$ indicate the mean values of slope angle and slope aspect, respectively.

$$r_{S.A} = \sum_{i=1}^{n} \frac{S_i - \bar{S}}{\sqrt{\sum_{k=1}^{n}(S_k - \bar{S})^2}} \cdot \frac{A_i - \bar{A}}{\sqrt{\sum_{k=1}^{n}(A_k - \bar{A})^2}} \quad (3)$$

Multicollinearity among the various LCFs cited in Section 3.2.1 was identified by computing the metric of variance inflation factors (VIF) and tolerances (Table 3). The Pearson's

Table 3. Analysis of multicollinearity of LCFs.

S. No.	LCF	Tolerance	VIF
1	Slope	0.985	2.45
2	Aspect	0.894	3.7
3	Curvature	0.684	2.57
4	Distance from stream	0.245	1.45
5	Altitude	0.689	2.34
6	Land use	0.774	1.56
7	Lithology	0.256	2.76
8	STI	0.367	3.49
9	WI	0.968	1.72

Table 4. Pearson's correlation between pairs of LCFs.

	Slope	Aspect	Curv.	Dist. Str.	Altitude	Land use	Lith.	STI	WI
Slope	1								
Aspect	0.345	1							
Curv.	0.370	- 0.016	1						
Dist. St.	0.044	0.072	0.539	1					
Altitude	0.245	0.062	0.067	0.578	1				
Land use	0.174	0.120	- 0.078	0.175	0.234	1			
Lith.	0.286	0.314	0.093	0.066	0.564	0.783	1		
STI	0.448	0.175	0.144	0.068	0.321	0.632	0.335	1	
WI	0.359	0.068	0.149	0.098	0.743	0.564	0.452	0.734	1

the correlation coefficient was computed for the various LCFs and is presented in Table 4. The result shows that the largest VIF is 3.7 and the lowest tolerance is 0.245. These values satisfy the VIF < 10 and tolerance > 0.1 criteria which indicates that there is no multicollinearity among the 9 LCFs considered.

2.3.6 Factor selection based on information gain ratio (IGR)

All LCFs do not serve equally in the correct prediction of landslides. Therefore, to quantify the ability to the prediction of landslides using various LCFs, the technique of Information Gain Ratio (IGR) has been used in our study. The concept of information gain stems from information theory which is used to find the decrease in overall entropy of the system to find the feasibility of the usage of a LCF for the prediction of landslides. However, due to the bias that information gain has towards datasets that contain many sample values, the IGR is used. A high value of IGR indicates a higher predictive ability[45] of the LCF. For an input dataset of m samples in the training dataset T, and the output class L_i of landslide or non-landslide, the information entropy to classify T is given by equation 4.

$$\text{InfoEnt}(T) = -\sum_{i=1}^{2} \frac{m(L_i,T)}{|T|} \log_2 \frac{m(L_i,T)}{|S|} \quad (4)$$

The amount of information required to disintegrate T into $(T_1, T_2, \ldots T_n)$ with respect to a LCF X is given by equation 5.

$$\text{InfoEnt}(T, X) = \sum_{j=1}^{n} \frac{T_j}{|T|} \text{InfoEnt}(T) \quad (5)$$

The Information Gain Ratio (IGR) for a certain LCF X is given by equation 6.

$$\text{IGR}(T, X) = \frac{\text{InfoEnt}(T) - \text{InfoEnt}(T, X)}{\text{SplitInfo}(T, X)} \quad (6)$$

Here, $SplitInfo(T, X)$ represents the potential information generated by dividing the training dataset T into n subsets and is given by equation 7.

$$\text{SplitInfo}(T, X) = -\sum_{k=1}^{m} \frac{T_k}{|T|} \log_2 \frac{T_k}{|T|} \quad (7)$$

The details of the LCF selection based on IGR method have been specified in Table 5. It has been observed that the largest merit is for slope (0.568). The LCFs of lithology, STI, and WI showcase an average merit of zero and hence indicate no ability to predict the occurrence of landslides. Therefore, the LCFs of lithology, STI, and WI have been excluded from further analysis for landslide prediction.

Based on the parameters for the selection of LCFs for our region of study, the LCFs of slope angle, slope aspect, slope curvature, distance from the stream, altitude, and land use have been considered for our analysis. The slope angle map,[46] aspect angle map,[47] slope curvature map,[48] distance from stream map,[49] altitude map, and land use map have been showcased in Figs. 4(a-f).

Table 5. Details of the selection of Landslide Conditioning factors (LCF) using the Information Gain Ratio (IGR) technique.

S. No.	LCF	Average Merit	Standard Deviation
1	Slope	0.568	± 0.027
2	Aspect	0.100	± 0.023
3	Curvature	0.389	± 0.028
4	Distance from stream	0.377	±0.152
5	Altitude	0.038	± 0.084
6	Land use	0.028	± 0.088
7	Lithology	0	0
8	STI	0	0
9	WI	0	0

2.4 Machine learning algorithms used for landslide prediction and their parameters for analysis

As described in Table 2, several machine learning algorithms are available in the existing literature to classify landslides. This section of the study describes the Maximum Likelihood algorithm (ML), the ISO clustering algorithm (ISO), and the Random Forest (RF) algorithm used in our study.

2.4.1 Maximum likelihood (ML) algorithm

The ML algorithm is a method to fit models to available data and find out which is the best fit. Alternatively, it can be defined as a technique to find the best set of parameters viz. mean (μ) and variance (σ) which could best be used to describe a distribution from which a data sample 'x' has been pulled out. The methodology followed by the ML algorithm is to determine the best parameters that fit the data by maximizing the log-likelihood function comprising of these parameters and then using information theory to compare all these models to find the best-fit model.

The likelihood function is defined in equation 6 as the true

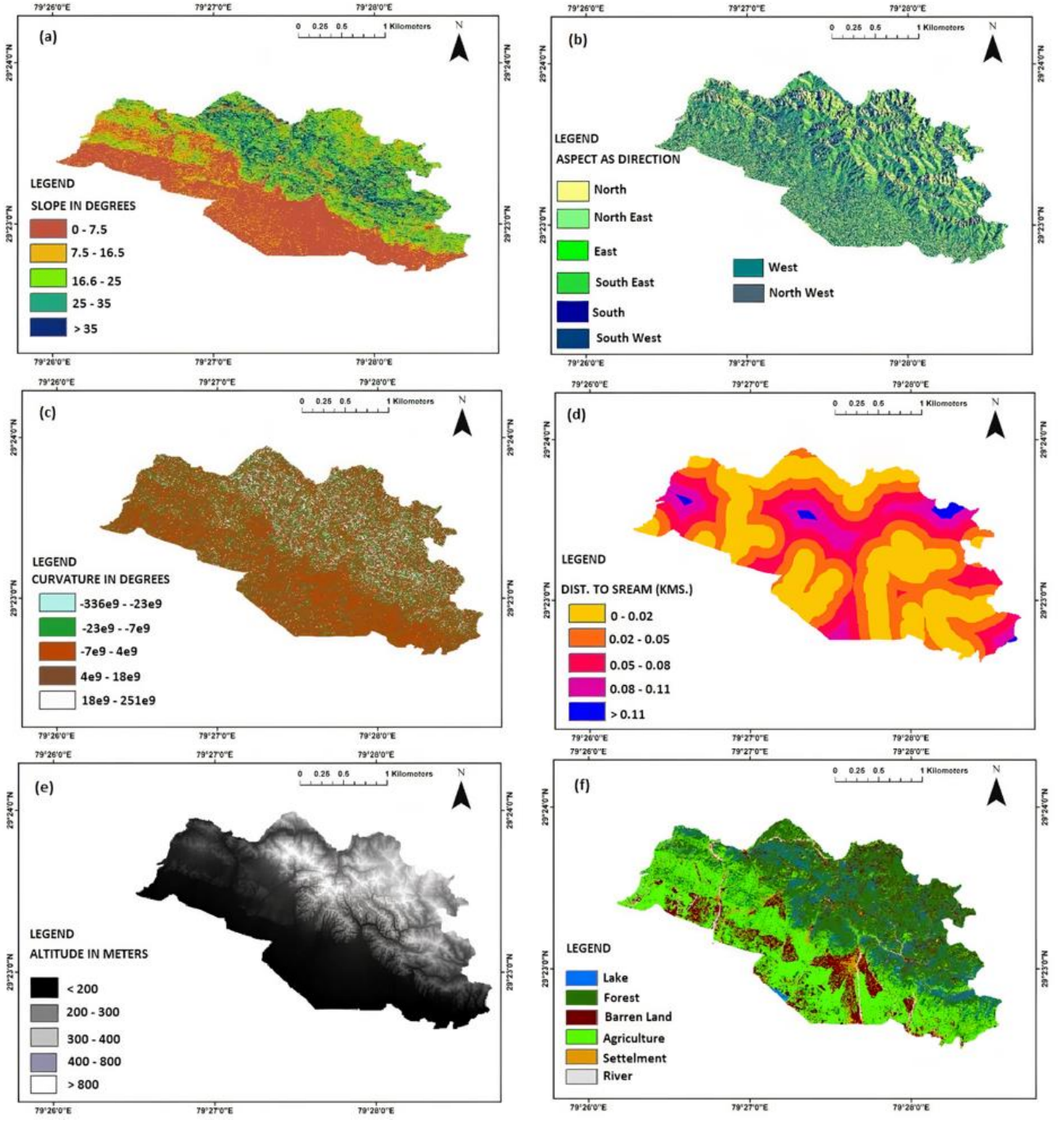

Fig. 4 (a) Slope angle LCF map, (b) Aspect angle LCF map, (c) Curvature LCF map, (d) Distance from stream LCF map, (e) Altitude map (f) LULC map.

parameters of a given data are the same as observing the data given the parameter values. Maximizing the values of these functions are cumbersome. Hence the log of equation 8 is taken and is termed as the Log-likelihood function shown in equation 9.

$$L(\theta|x_i) = \prod f(x_i|\theta) \quad (8)$$

$$\ln L(\theta|x_i) = \sum f(x_i|\theta) \quad (9)$$

where:

$L(\theta|x_i)$ = Likelihood of sample x_i being drawn from a distribution represented by parameters θ. $f(x_i|\theta)$ = probability of sample x_i being drawn given a distribution with parameters θ is encountered.

The justification for using the ML algorithm for our classification is that this is a form of unsupervised classification where the dataset of images available does not have any information about which distribution it has been taken from. Therefore, the ML estimation algorithm is best suited to estimate the distribution and thereby the characteristics of the landscape such as rivers, lakes barren land, agricultural land, *etc*. by comparing various distributions

and choosing the best fit from them using the principles of information theory. The hyperparameter values assigned for the ML algorithm are showcased in Table 3.

The ML algorithm of classification is available in ArcGIS and it follows two principles which are - the sample classes are normally distributed and use Bayes' theorem for making decisions. The variance and covariance are taken into account in this classification. The cells are assigned to each class by statistical probability. By assigning an equal option to a priori probability, the cells are classified into the class based on maximum probability.

2.4.2 ISO clustering (ISO) algorithm

The ISO algorithm is an iterative self-organizing process of unsupervised classification and uses Euclidean distance for allotting each point to a cluster. The cluster is assigned by the user. Each cell is allocated to the nearest mean by Euclidean distance. After the completion of the first iteration, new means are calculated for all clusters, this process is continued till the specified number of iterations is met. Each cluster is expected to have enough cells for the proper representation of each class.

In the case of a wide variety of ranges of input band values for various classes, normalization of the input band values is performed as shown in equation 10.

$$OR = \frac{(IR-min_{IR})\times(max_{OR}-min_{OR})}{(max_{IR}-min_{IR})} + min_{OR} \quad (10)$$

where:
OR = Output Raster
IR = Input Raster
min_{IR} = Minimum value of Input Raster
max_{IR} = Maximum value of Input Raster
min_{OR} = Minimum value of Output Raster
max_{OR} = Maximum value of Output Raster

2.4.3 Random forest (RF) algorithm

The RF algorithm is a supervised learning technique that uses multiple decision trees to classify data using the technique of averaging to improve the overall accuracy of classification. It uses the principle of bagging to obtain the best results in classification. The method of splitting the decision tree could either be using the Gini index (equation 11) obtained by subtracting the summation of the square of all class probabilities from 1, or Entropy (equation 12) which signifies the randomness in a sample and needs to be reduced during the process of classification.

$$Gini(P(X)) = 1 - \sum_{n=1}^{N}(x_n)^2 \quad (11)$$

$$Entropy(P(X)) = -\sum_{n=1}^{N} x_n log_2(x_n) \quad (12)$$

where:
$P(X) = (x_1, x_2, \ldots . x_n)$ and x_i is the probability of classification of an object into a particular class.
The overall size of the forest can be controlled by reducing the number of leaf nodes and the maximum depth of the tree. Table 6 tabulates the hyperparameter space used for the implementation of the ML, ISO, and RF algorithms.

Table 6. Hyperparameters of ML, ISO, and RF Algorithms.

Name of Algorithm	Hyperparameters
ML	Reject fraction = 0.01 This indicates to reject classifications that have a confidence value lesser than 0.01 A priori Probability weighting = Equal This indicates that each pixel of the image is assigned to the class to which it has the highest probability of being a part
ISO	Number of classes = 6 This indicates the number of output land feature classes which is 6 in this case Sample interval = 10 This indicates the sampling interval to be chosen which has a default value of 10
RF	Number of estimators = 5000 This indicates the number of Decision trees in the Random Forest Splitting criterion = Gini This indicates the criterion for splitting the nodes of the Decision trees Maximum depth = 3 This indicates the amount of levels/ depth of the Decision tree Class weight = Balanced This indicates that all classes are given equally probable priority

2.4.4 Classification assessment

The assessment of the accuracy of the ML, ISO, and RF algorithms is performed by computing five parameters viz., accuracy, sensitivity, specificity, positive predicted value, and negative predicted value. These parameters are computed based on the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) that are generated by the algorithms and are defined using the following equations 13, 14, 15, 16, and 17.

$$\text{Accuracy (Acc.)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

$$\text{Sensitivity (Sens.)} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{Specificity (Spec.)} = \frac{TN}{TN+FP} \quad (15)$$

$$\text{Positive Predicted Value (PPV)} = \frac{TP}{TP+FP} \quad (16)$$

$$\text{Negative Predicted Value (NPV)} = \frac{TN}{TN+FN} \quad (17)$$

The Kappa index (κ)[45,50-52] is a measure to express the ability of landslide classification algorithms to classify the pixels corresponding to landslides. It is termed as the proportion of observed agreement beyond that expected by chance. A Kappa magnitude of 0.8 to 1 indicates almost perfect agreement, 0.6 to 0.8 indicates substantial agreement, 0.4 to 0.6 indicates moderate agreement, 0.2 to 0.4 indicates fair agreement, 0 to 0.2 indicates slight agreement and less than 0 indicates no agreement.[53]

To validate the results obtained using the ML, ISO, and RF

algorithms, frequency ratio (FR) analysis has been carried out for the four land conditioning factors. FR analysis is a technique that quantitatively performs landslide susceptibility mapping by making use of GIS techniques and spatial data. It builds an association between the landslide inventory and landslide causative factors.[54] The computation of FR is performed using Equation 18.[55]

$$FR = \frac{Y}{X} \quad (18)$$

where:

Y = % of landslide pixels in a particular class and X = % of total pixels in a particular class.

In order to perform a comparative analysis of the three machine learning algorithms, the area under the ROC curves is computed on the validation dataset.

2.4.5 Bagging strategy employed

Bagging which is the short form of Bootstrap Aggregation refers to an ensemble machine learning algorithmic technique to find the best possible classification or regression model

study outlines the results achieved and the discussion pertaining to the same.

3. Results and discussion

The LANDSAT 7 dataset with a spatial resolution of 30 meters is used. The DEM is extracted from the images obtained from the LANDSAT 7 sensor. From the DEM, the LCF maps have been extracted using the ArcGIS software. The ArcGIS software is used to classify the regions of landslide occurrence and non-occurrence with location data from the LIM (Fig. 2) obtained from the Geological Survey of India. Initially the results. In our study, we have presented the usage of the ML, ISO, and RF techniques for the creation of LSMs for the region of Nainital. Each of these algorithms provides varied results for the parameters of concern *i.e.* accuracy, sensitivity, specificity, positive predicted value, and negative predicted value. In order to obtain the best possible results for our classification, we have performed bagging by drawing several bootstrap samples from the training data set pixels and fitting the ML, ISO, and RF algorithms on the same.

This section gave a brief description of the algorithms we have used for our classification. The following section of our

Table 7. Comparative performance of ML, ISO, and RF models based on Confusion matrix parameters and Kappa index.

Para m.	ML						ISO						RF					
	S	A	C	D	AL	L	S	A	C	D	AL	L	S	A	C	D	AL	L
TP	20	46	34	39	38	44	55	39	39	42	61	59	59	59	64	61	37	32
TN	22	45	46	43	42	43	62	34	35	37	38	53	54	54	27	38	35	43
FP	8	14	10	10	11	13	12	16	15	13	15	21	22	22	16	15	17	10
FN	10	15	25	14	12	14	19	11	11	8	38	26	26	26	54	38	9	25
PPV (%)	72	77	77	78	80	77	83	71	71	77	80	72	73	73	80	81	71	77
NPV (%)	69	75	65	76	76	75	77	76	76	83	51	69	68	68	33	50	76	65
Sensitivity (%)	67	76	58	74	74	76	75	78	78	84	61	71	70	70	55	62	78	58
Specificity (%)	74	76	82	81	81	76	84	68	70	74	71	71	71	71	63	72	68	81
Accuracy (%)	70	76	70	77	77	76	84	73	74	79	65	70	70	70	58	65	73	69
Kappa Index	0.30	0.69	0.42	0.60	0.60	0.69	0.79	0.47	0.48	0.60	0.67	0.58	0.57	0.57	0.70	0.68	0.46	0.41

points with landslide occurrence are extracted from the LCF. These points are classified with a training-testing split of 80:20 using the ML, ISO, and RF algorithms. The results of the comparative performances of the three algorithms based on the confusion matrix parameters and Kappa index are showcased in Table 7.

Due to lack of space, the following naming convention abbreviations have been followed for Table 7; Slope angle is denoted as S, Aspect angle is denoted as A, Curvature is denoted as C, Distance from the stream is denoted as D, Altitude is denoted as AL and Land use is denoted as L.

From Table 7 several conclusions have been drawn. Firstly, the highest positive predicted value (PPV%) is for the ISO algorithm with a slope angle considered LCF (83%). This indicates that the probability that, the ISO algorithm with the slope as LCF, correctly classifies the pixels into the class of landslide is 83%. The ISO algorithm with distance from stream LCF has the highest negative predicted value (NPV %) of 83% which indicates the probability of correctly classifying the pixels in the non-landslide class. The ISO algorithm with distance from stream LCF has the highest sensitivity of 84% indicating that 84% of the pixels are correctly classified into the landslide class. Further, the ISO algorithm considering slope angle as LCF has the highest specificity of 84% indicative of the correct classification of non-landslide pixels into the non-landslide class. The highest accuracy of classification is also exhibited by the ISO algorithm which is 84%. The highest Kappa indices for all three algorithms range between 0.69 to 0.79 which indicates substantial agreement between the classification models and the ground truth. The highest Kappa index is computed for ISO clustering using slope angle as LCF with a value of 0.79.

Considering this, it has been concluded that the ISO clustering algorithm which uses slope angle as a land conditioning factor, has the highest PPV, specificity, and accuracy in the classification of landslides in the Nainital region. Its other parameters also lie within permissible limits of acceptance. Therefore, the ISO clustering algorithm outperforms the Maximum likelihood algorithm and random forest algorithm in terms of PPV, specificity, and accuracy.

Table 8. Mean Accuracy and Standard Deviation of K-fold Cross-validation over ML, ISO, and RF algorithms.

Algorithm	Mean Accuracy	Standard Deviation
ML	79 %	± 0.023
ISO	82 %	± 0.039
RF	76 %	± 0.013

We have used 1000 samples of the input image to classify it into two output classes viz., landslide occurrence and no landslide occurrence, based on the features of slope angle, slope aspect, curvature, distance to stream, altitude, and LULC. The algorithms are evaluated using the k-fold cross-validation technique used for Bagging. For using this technique, the input image has been split into 10 sets and a three-fold cross-validation of the classification results has been performed. The mean accuracy and standard deviation accuracy scores observed after the k-fold cross-validation for the algorithms used for classification are shown in Table 8. The frequency ratio analysis for the LCFs obtained from the LCF maps has been tabulated in Table 9.

Table 9. Frequency ratio analysis of LCFs.

LCF	Class	% of pixels in class (X)	% of landslide pixels in class (Y)	Frequency Ratio = (Y/X)
Slope (degree)	0 – 7.5	28.82	3.2	0.11
	7.5 – 16.5	20.20	5.5	0.27
	16.5 – 25	21.5	24.5	1.13
	25 – 35	10.95	37.2	3.4
	> 35	18.53	29.5	1.6
Aspect (direction)	N	9.54	27	2.83
	NE	41.5	8.2	0.2
	E	11.77	34.2	2.9
	SE	5.8	4.91	0.84
	S	6.2	5	0.8
	SW	6.3	5.1	0.8
	W	9.5	7.62	0.8
	NW	9.4	7.7	0.81
Curvature (degree)	-336 e^9 - -23 e^9	32.6	9.5	0.3
	-23 e^9 - -7e^9	21.63	24.75	1.2
	-7e^9- 4e^9	15.71	11	0.7
	4e^9- 18e^9	13.7	42.5	3.1
	18e^9- 251e^9	15.62	12.5	0.8
Distance to Stream (km)	0 – 0.02	23.3	38.25	1.6
	0.02 – 0.05	24.66	23	0.94
	0.05 – 0.08	22.5	9.75	0.44
	0.08 – 0.11	18.8	24.75	1.32
	> 0.11	10.83	4.25	0.4
Altitude (m)	< 200	25.71	16.1	0.62
	200 – 300	19.4	12.7	0.65
	300 – 400	17.3	5	0.29
	400 - 800	23.63	29.75	1.3
	> 800	13.9	37.25	2.67
Land Use	Lake	19.5	11.7	0.6
	Forest	21.63	24.75	1.14
	Barren Land	20.23	34.75	1.71
	Agriculture	13.9	17.25	1.24
	Settlement	13.62	17.5	1.28
	River	10.83	4.25	0.4

From Table 9 it has been observed that of all the values of frequency ratio for the various classes of LCFs, the slope angle gives the highest value of 3.4 for the class of 25 – 35 degrees. This is followed by an aspect of East with a value of 2.9 further followed by the LCF of altitude for > 800 meters with a value of 2.67. The frequency of landslide occurrence is almost equal in terms of LULC for barren lands, agricultural lands, and

settlements.

The quantitative performance of the three machine learning models has been compared using the receiver operating characteristics curve (ROC). The ROC is computed by drawing a graph based on Sensitivity and Specificity using different threshold values. The area under ROC (AUC) gives the comprehensive performance of the machine learning models used for creating the LSMs (Table 10). A perfect model gives an AUC value of 1 which signifies that all the landslide pixels have been correctly classified as landslides.

Table 10. AUC values for Machine Learning algorithms used for the creation of LSMs.

S. No.	Machine Learning algorithm	AUC
1	ML	0.743
2	ISO	0.817
3	RF	0.721

From Table 10 it can be interpreted that, the ISO algorithm outperforms the ML and RF algorithms in terms of AUC. Therefore, it has been concluded that the ISO algorithm using the LCF of slope angle along with altitude generates the best LSM (Fig. 5) for the region of Nainital.

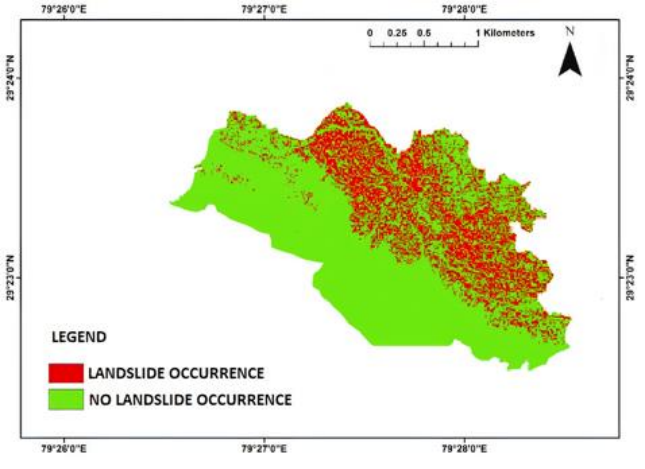

Fig. 5 LSM map for Nainital.

4. Conclusion

This study presents a comprehensive comparison and evaluation of three machine learning algorithms viz., Maximum Likelihood estimate, ISO clustering, and Random Forest for the classification of landslides based on land conditioning factors of slope angle, slope aspect, curvature, and distance from the stream, for the region of Nainital in India. The results presented are:

1. Presentation of collinearity among the various LCFs and selection of appropriate LCFs for construction of the LSM using the Information Gain Ratio.
2. Assessment of the ML, ISO, and RF algorithms and their comparison using the Confusion matrix, Kappa index, Frequency ratio, and AUC.

According to the results of this case study, the highest PPV of 83% is observed for the ISO algorithm with slope angle considered as LCF. The ISO algorithm with distance from stream LCF has the highest negative NPV of 83%. The ISO algorithm with distance from stream LCF has the highest sensitivity of 84% and the ISO algorithm considering slope angle as LCF has the highest specificity of 84%. A Kappa index of 0.79 has been computed for ISO clustering using slope angle as LCF.

It has been observed that of all the values of frequency ratio for the various classes of LCFs, the slope angle gives the highest value of 3.4 for the class of 25 – 35 degrees. It has also been observed that the ISO algorithm outperforms the ML and RF algorithms in terms of AUC with a value of 0.887.

The various results of this study may prove to be of use for policy and decision-making in the areas prone to landslides in Naintal, India.

Conflict of interest

There are no conflicts to declare.

Supporting information

Applicable.

References

[1] S. Grammenos, Office for Official Publications of the European Communities; Luxembourg: 2003, *Feasibility study. Comparable statistics in the area of care for dependent authors in the European Union*, 2003.

[2] S. Keesstra, G. Mol, A. Zaal, J. Wallinga, B. Jansen, Wageningen Soil Conference 2015: Soil Science in a Changing World, *The Netherlands, Wageningen*, 2015.

[3] J. Corominas, C. van Westen, P. Frattini, Recommendations for the quantitative assessment of landslide risk (No. IKEEART-2015-1153), *Aristotle University of Thessaloniki*, 2014, **73**, 209-263, doi: 10.1007/s10064-013-0538-8.

[4] M. Papathoma-Köhle, A. Zischg, S. Fuchs, T. Glade, M. Keiler, *Environmental Modelling & Software*, 2015, **63**, 156-169, doi: 10.1016/j.envsoft.2014.10.003.

[5] L. Schilirò, L. Montrasio, G. Scarascia Mugnozza, *Science of the Total Environment*, 2016, **569-570**, 134-144, doi: 10.1016/j.scitotenv.2016.06.124.

[6] P. M. Atkinson, R. Massari, *Computers & Geosciences*, 1998, **24**, 373-385, doi: 10.1016/s0098-3004(97)00117-9.

[7] K. Mertens, L. Jacobs, J. Maes, C. Kabaseke, M. Maertens, J. Poesen, M. Kervyn, L. Vranken, *Science of the Total Environment*, 2016, **550**, 1032-1043, doi: 10.1016/j.scitotenv.2016.01.171.

[8] S. Khanduri, *International Journal of Engineering Applied Sciences and Technology*, 2019, **3**, 42-49, doi: 10.33564/IJEAST.2019.v03i12.007.

[9] L. Thanh, F. De Smedt, *Landslides*, 2014, **111**, 897-907, doi: 10.1007/s10346-013-0437-x.

[10] C. J. Van Westen, M. J. T. Terlien, *Earth Surface Processes and Landforms*, 1996, **21**, 853-868, doi: 10.1002/(sici)1096-9837(199609)21: 9853: aid-esp676>3.0.co;2-c.

[11] D. Costanzo, E. Rotigliano, C. Irigaray, J. D. Jiménez-Perálvarez, J. Chacón, *Natural Hazards and Earth System*

Sciences, 2012, **12**, 327-340, doi: 10.5194/nhess-12-327-2012.
[12] M. Tilahun, M. Matebie, *Geoenvironmental Disasters*, 2020, **7**, 1-22, doi: 10.1186/s40677-020-00155-x.
[13] D. Tien Bui, T. A. Tuan, H. Klempe, B. Pradhan, I. Revhaug, Landslides, 2016, **13**, 361-378, doi: 10.1007/s10346-015-0557-6.
[14] Z. Ma, G. Mei, *Neural Computing and Applications*, 2020, **20**, 1-27, doi: 10.1007/s00521-020-05529-8.
[15] Y. Achour, H. R. Pourghasemi, *Geoscience Frontiers*, 2020, **11**, 871-883, doi: 10.1016/j.gsf.2019.10.001.
[16] B. T. Pham, B. Pradhan, D. Tien Bui, I. Prakash, M. B. Dholakia, *Environmental Modelling & Software*, 2016, **84**, 240-250, doi: 10.1016/j.envsoft.2016.07.005.
[17] W. Chen, X. Xie, J. Wang, B. Pradhan, H. Hong, D. T. Bui, Z. Duan, J. Ma, *Catena*, 2017, **151**, 147-160, doi: 10.1016/j.catena.2016.11.032.
[18] A. Dehnavi, I. N. Aghdam, B. Pradhan, M. H. Morshed Varzandeh, *Catena*, 2015, **135**, 122-148, doi: 10.1016/j.catena.2015.07.020.
[19] O. F. Althuwaynee, B. Pradhan, S. Lee, *International Journal of Remote Sensing*, 2016, **37**, 1190-1209, doi: 10.1080/01431161.2016.1148282.
[20] S. Lee, S.-M. Hong, H.-S. Jung, *Sustainability*, 2017, **9**, 48, doi: 10.3390/su9010048.
[21] T. Kavzoglu, E. Kutlug Sahin, I. Colkesen, *Natural Hazards*, 2015, **76**, 471-496, doi: 10.1007/s11069-014-1506-8.
[22] L. Ermini, F. Catani, N. Casagli, *Geomorphology*, 2005, **66**, 327-343, doi: 10.1016/j.geomorph.2004.09.025.
[23] W. Chen, M. Panahi, P. Tsangaratos, H. Shahabi, I. Ilia, S. Panahi, S. Li, A. Jaafari, B. Bin Ahmad, *Catena*, 2019, **172**, 212-231, doi: 10.1016/j.catena.2018.08.025.
[24] B. Kalantar, B. Pradhan, S. A. Naghibi, A. Motevalli, S. Mansor, Geomatics, *Natural Hazards and Risk*, 2018, **9**, 49-69, doi: 10.1080/19475705.2017.1407368.
[25] A. Aditian, T. Kubota, Y. Shinohara, *Geomorphology*, 2018, **318**, 101-111, doi: 10.1016/j.geomorph.2018.06.006.
[26] A. Trigila, C. Iadanza, C. Esposito, G. Scarascia-Mugnozza, *Geomorphology*, 2015, **249**, 119-136, doi: 10.1016/j.geomorph.2015.06.001.
[27] Y. Wang, D. Sun, H. Wen, H. Zhang, F. Zhang, *International Journal of Environmental Research and Public Health*, 2020, **17**, 4206, doi: 10.3390/ijerph17124206.
[28] S. Lee, *Environmental Earth Sciences*, 2013, **70**, 3197-3215, doi: 10.1007/s12665-013-2385-0.
[29] T. Mersha, M. Meten, *Geoenvironmental Disasters*, 2020, **7**, 20, doi: 10.1186/s40677-020-00155-x.
[30] A. Jaafari, A. Najafi, H. R. Pourghasemi, J. Rezaeian, A. Sattarian, *International Journal of Environmental Science and Technology*, 2014, **11**, 909-926, doi: 10.1007/s13762-013-0464-0.
[31] M. E. Dickson, G. L. W. Perry, *Environmental Modelling & Software*, 2016, **76**, 117-127, doi: 10.1016/j.envsoft.2015.10.029.
[32] H. Hong, H. R. Pourghasemi, Z. S. Pourtaghi, *Geomorphology*, 2016, **259**, 105-118, doi: 10.1016/j.geomorph.2016.02.012.
[33] A. M. Youssef, H. R. Pourghasemi, *Geoscience Frontiers*, 2021, **12**, 639-655, doi: 10.1016/j.gsf.2020.05.010.
[34] I. N. Aghdam, M. H. M. Varzandeh, B. Pradhan, *Environmental Earth Sciences*, 2016, **75**, 553, doi: 10.1007/s12665-015-5233-6.
[35] A. M. Youssef, B. Pradhan, M. N. Jebur, H. M. El-Harbi, *Environmental Earth Sciences*, 2015, **73**, 3745-3761, doi: 10.1007/s12665-014-3661-3.
[36] X. Deng, L. Li, Y. Tan, *International Journal of Geo-Information*, 2017, **6**, 103, doi: 10.3390/ijgi6040103.
[37] S. Park, C. Choi, B. Kim, J. Kim, *Environmental Earth Sciences*, 2013, **68**, 1443-1464, doi: 10.1007/s12665-012-1842-5.
[38] H. Shahabi, M. Hashim, B. B. Ahmad, *Environmental Earth Sciences*, 2015, **73**, 288-304, doi: 10.1038/srep09899.
[39] K. C. Devkota, A. Deep Regmi, H. R. Pourghasemi, K. Yoshida, B. Pradhan, I. C. Ryu, M. R. Dhital, O. F. Althuwaynee, *Natural Hazards*, 2013, **65**, 135-165, doi: 10.1007/s11069-012-0347-6.
[40] R. Kumar, R. Anbalagan, *The Journal of Geology*, 2016, **80**, 271-286, doi: 10.1007/s12040-019-1159-9.
[41] H. A. H. Al-Najjar, B. Kalantar, B. Pradhan, V. Saeidi, *Proceedings of SPIE Remote Sensing conference. Strasbourg, France*, 2019.
[42] W. Chen, H. R. Pourghasemi, S. B. Naghibi, *Frontiers in Earth Science,* 2017, **77**, 611-629, doi: 10.3389/feart.2021.701837.
[43] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, S. Lautenbach, *Ecography*, 2013, **36**, 27-46, doi: 10.1111/j.1600-0587.2012.07348.x.
[44] G. D. Booth, M. J. Niccolucci, E. G. Schuster, *U. S. Department. of Agricultural and Forest Services, Ogden, Utah*, 1994.
[45] H. Saito, D. Nakayama, H. Matsuyama, *Geomorphology*, 2009, **109**, 108-121, doi: 10.1016/j.geomorph.2009.02.026.
[46] P. Magliulo, A. Di Lisio, F. Russo, A. Zelano, *Natural Hazards*, 2008, **47**, 411-435, doi: 10.1007/s11069-008-9230-x.
[47] B. Neuhäuser, B. Damm, B. Terhorst, *Landslides*, 2012, **9**, 511-528, doi: 10.1007/s10346-011-0305-5.
[48] I. S. Evans, Final report on Grand DA-ERO-591-73-G004, University of Durham, England, 1979.
[49] H. R. Pourghasemi, A. G. Jirandeh, B. Pradhan, C. Xu, C. *Journal of Earth System Science*, 2013, **122**, 349-369, doi: 10.1007/s12040-013-0282-2.
[50] J. Cohen, *Guilford Press*, 1960, **20**, 37–46, doi: 10.1177/001316446002000104.
[51] B. D. Tien, B. Pradhan, O. Lofman, I. Revhaug, *Geomorphology*, 2012, **45**, 199–211, doi: 10.1155/2012/974638.
[52] M. Van Den Eeckhaut, P. Reichenbach, F. Guzzetti, M. Rossi, J. Poesen, *Natural Hazards and Earth System Sciences*, 2009, **9**, 507-521, doi: 10.5194/nhess-9-507-2009.
[53] J. R. Landis, G. G. Koch, *Biometrics*, 1977, **33**, 159, doi: 10.2307/2529310.
[54] S. Reis, A. Yalcin, M. Atasoy, R. Nisanci, T. Bayrak, M. Erduran, C. Sancar, S. Ekercin, *Environmental Earth Sciences*,

2012, **66**, 2063-2073, doi: 10.1007/s12665-011-1432-y.
[55] S. Mondal, R. Maiti, *International Journal of Disaster Risk Science*, 2013, **4**, 200-212, doi: 10.1007/s13753-013-0021-y.

Author information

Mr. Pranav Rajendra Rane *is pursuing his Bachelor of Technology degree in Mechatronics from Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Karnataka, India. His research interests include Machine learning, Remote sensing and Robotics.*

Dr. Shweta Vincent *is a Professor at the Department of Mechatronics, from Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Karnataka, India. She obtained her Ph.D. in Antenna design from Karunya Institute of Technology and Sciences, Coimbatore, India in 2020. Her research interests include Antenna Design, Algorithm Design, Machine Learning, and Remote Sensing.*