Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Landslide Susceptibility Modeling and Mapping: A Comparison of Frequency Ratio and Decision Tree Models

Raghvendra Kumar[1], Nguyen Thanh Tuan[2,*]
[1]Department of CSE, GIET University, Gunupur-765022, Odisha, India; raghvendra@giet.edu
[2]Geotechnical and Artificial Intelligence research group, 54 Trieu Khuc, Thanh Liet, Hanoi 100000, Vietnam; tuannt94@utt.edu.vn

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.2.315-338

***Corresponding author:**
Email address:
tuannt94@utt.edu.vn

Received: 17/11/2025
Received in Revised Form: 10/03/2026
Accepted: 17/04/2026

Abstract: Landslides constitute one of the most damaging natural hazards in mountainous regions worldwide, and the Himalayas are particularly vulnerable due to their steep topography, complex geology, and intense monsoonal rainfall. This study develops a landslide susceptibility map for a part of Uttarakhand, India, using two comparatively simple yet robust modeling approaches—the bivariate Frequency Ratio (FR) method and the Decision Tree (DT) machine learning algorithm. A total of ten conditioning factors representing the key topographical and geo-environmental characteristics of the region were used, and a landslide inventory of 104 events was compiled for model training and validation. Model performance was evaluated using multiple statistical measures, including PPV, NPV, Sensitivity, Specificity, MAE, RMSE, and AUC. The results indicate that the DT model significantly outperforms the FR method, achieving an AUC of 0.848 compared to 0.578 for FR. The novelty of the work lies in the systematic comparison of these two widely used yet contrasting approaches in a Himalayan setting, demonstrating that the DT model provides a more reliable and transferable framework for susceptibility assessment in regions dominated by shallow landslides. These findings highlight the potential of simple machine-learning models to offer accurate and practical tools for landslide management in data-scarce mountainous terrains.

Keywords: Landslide susceptibility mapping; Decision Tree (DT); Frequency Ratio (FR); Machine learning; Himalayan terrain; Geohazard assessment; Remote sensing and GIS.

List of notations

$N_{pix}(SX_j)$	is the number of pixels of the landslide causal factor map layer
$N_{pix}(X_j)$	the number of pixels in the factor map layer to consider
E	is Entropy reduction of dataset
n	is the number of classes in the domain of the data set E
P_i	is the proportion of the number of class i elements over the total number of data set E
G(E, A)	is the information gain
A	is value that provides the domain of supporting attribute
E_v	is the subset of E

$|E_v|$ is the cardinality of E_v
$|E|$ is the cardinality of E

1. Introduction

Landslide is one of the most destructive disasters globally, affecting the socio-economy of countries and lives of people [1]. Landslide accounts for about 14% of natural disasters, especially in hilly areas [2]. The Himalayan region of India is a hilly and mountainous affected by natural disasters such as earthquake, flash floods and landslides. Most of the Uttarakhand state of the Himalaya is hilly and mountainous and is affected by numerous landslides every year [3]. Therefore, part of the landslide prone area of Uttarakhand has been selected for the landslide susceptibility mapping using statistical method, namely Frequency Ratio (FR) technique and Machine Learning (ML) method namely Decision Tree (DT) algorithm to identify best technique for the correct identification of landslide susceptible zones.

There are two main approaches used in the landslide studies: qualitative and quantitative. The qualitative methods are mainly based on knowledge and experiences of professionals; thus these approaches are subjective and influenced by individual judgment [4]. The hierarchical analysis process is one of the common qualitative methods, based on multi-criteria analysis, which is a tool applied to solve complex decision issues in many areas including landslide [5, 6]. The quantitative approach includes methods based on statistical analyses such as FR technique [7], Information Value Model [8] and Weight of Evidence method [9]. In particular, the FR method is considered a simple and popular approach in landslide susceptibility mapping with high accuracy [10, 11].

Recently, ML methods including DT methods are being used effectively in many fields including landslides prediction [5, 6, 8]. Other Artificial Intelligence (AI)/ ML methods applied in these studies include Na¨ıve Bayes (NB), Multilayer Perceptron (MLP) neural network classifier, and Alternating Decision Tree (ADT) [12].

The main objective of this study is to apply FR techniques and DT based ML method for developing landslide susceptibility zones as these approaches (FR and DT) have not been applied together in this area for landslide susceptibility mapping. In this study, published data of the Geological Survey of India (GSI) and data extracted from Google Earth Images and Digital Elevation Model (DEM) (30m spatial resolution) was used. Weka and ArcGIS software were used for the data analysis, modeling and visualization.

2. Material and methods

2.1. Study area

Pithoragarh district of Uttarakhand State, India is one of the landslide-prone areas of the Himalayas located between latitude 29°30'00" & 29°45'00" and longitude 80°00'00" & 80°15'00" has been selected as study area [3] (Fig. 1). Topography of the Pithoragarh area is hilly and mountainous dissected deep valleys. Hill ridges are aligned in East-West direction parallel to major faults. Major river in this area is Ramganga River and its tributaries. Geology of the area is complex comprising of meta-sedimentary rocks (slate, schist, phyllite, quartzite, limestone, marble, gneisses etc.) and meta-volcanics. Pithoragarh Formation comprising of two layers of different rocks occupy the major part of the district. The lower layer comprises of Dolomitic bearing stromatolite limestone, calcium phyllite, quartz and marble, whereas the upper layer is composed of limestone and dark gray limestone. Structurally rocks are folded and faulted and tectonically it is active. There is a great height difference in the study area varying elevation 429 m to 2427 m. Temperature in the area varies greatly from moderate at lower elevation to sub-zero at higher elevation. The average annual rainfall is 360 cm.

The district is thickly populated. Agriculture activities are confined in flat areas and on valley slopes. Forest hilly areas are covered with thick to sparse vegetation.

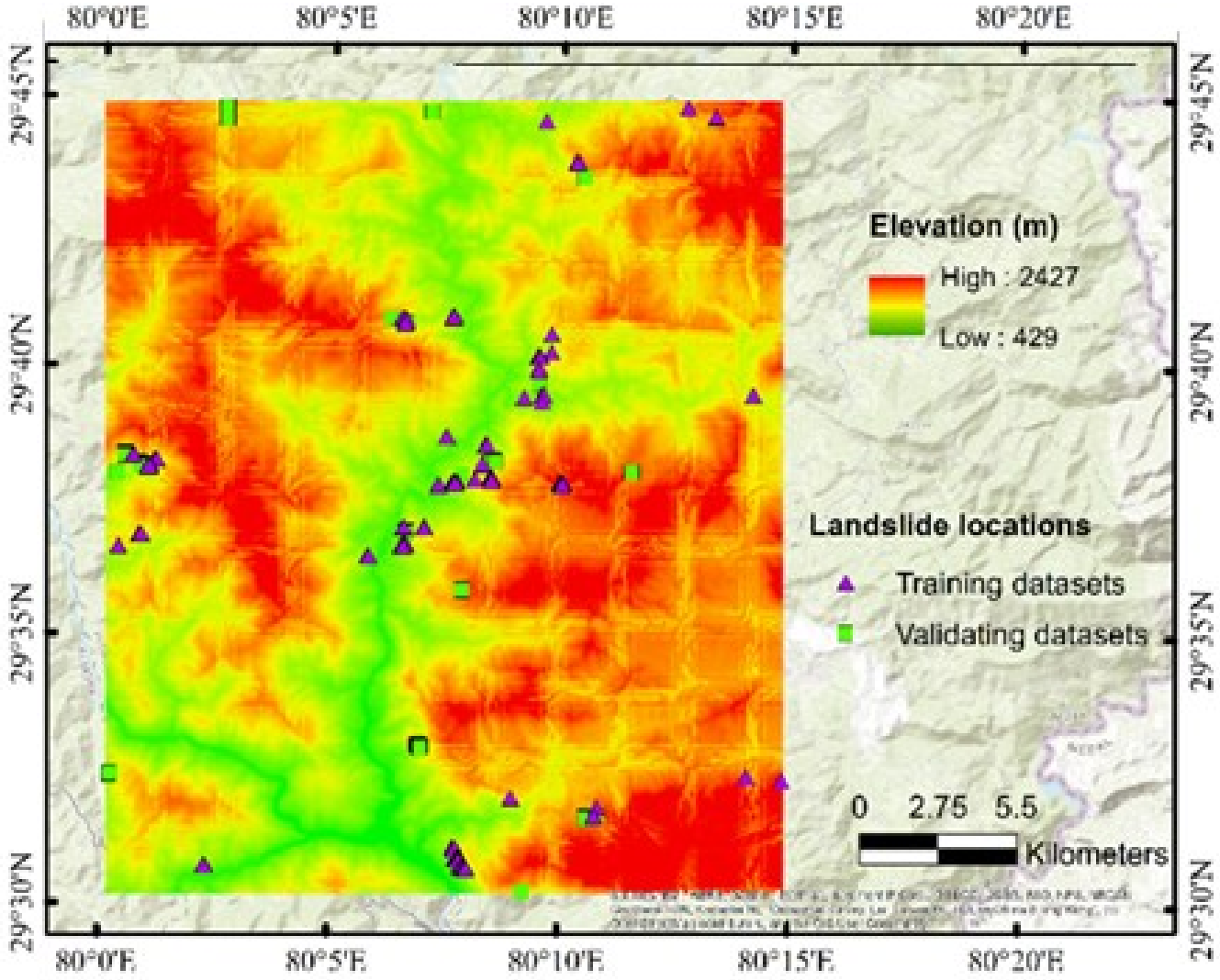

Fig. 1. Location of the study area and historical landslides

2.2. Methodology

The main steps of methodology adopted in this study are presented in Fig. 2. Landslide inventory was prepared from the available landslide records of the GSI and from Google Earth Images. Thematic maps of the landslide influencing factors were extracted from GSI report, DEM and Google Earth Images. These maps include Slope, Aspect, Curvature, Elevation, Geomorphology, Land cover, Slope Forming Material (SFM), Distance to river, Distance to road and Overburden depth. The accuracy of models was evaluated using statistical indices including Area Under the Receiver Operating Characteristic curve (AUC). Landslide susceptibility maps were generated using FR and DT (ML) models. Performance of the models was evaluated using standard statistical methods for the selection of the best model for the development of accurate landslide susceptibility maps.

2.3. Data used

2.3.1. Landslide inventory

Landslide inventory for 104 landslide events was created from the available record of GSI and Google Earth Images (Fig. 1). Most of the landslides occurring in the area are of shallow landslides (Fig. 3). Landslide area polygons are represented in maps by the points for visualization

and analysis. In addition, 104 non-landslide locations were identified from the areas where landslides have not occurred in the past and used for construction of the database for landslide susceptibility modeling. Landslide and non-landslide data was randomly split into 70:30 ratio for training (70%) and validation (30%) of the FR and DT models. This data splitting ratio is based on the experience of authors and other researches [3].

2.3.2. Landslide conditioning factors

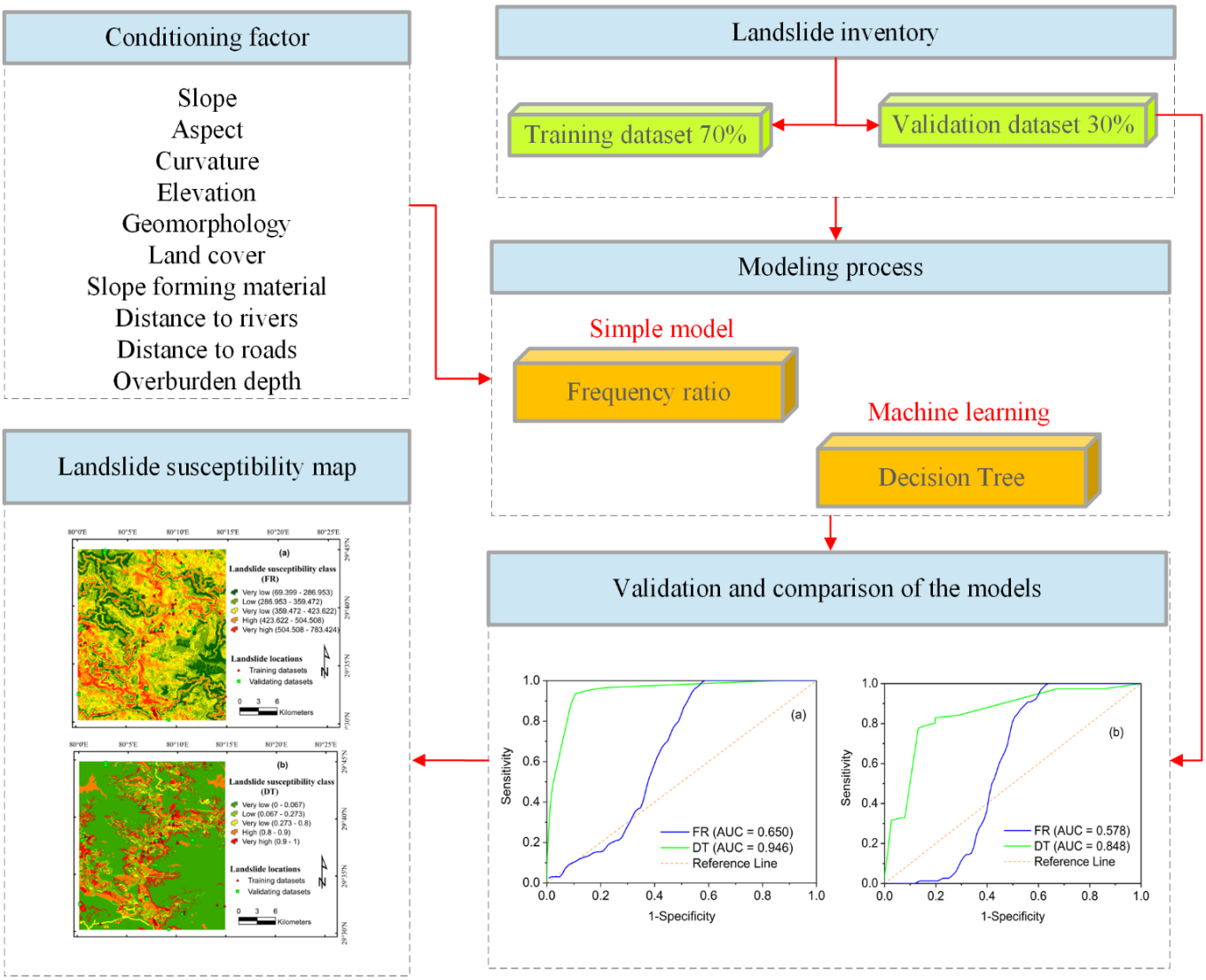

Fig. 2. Methodological framework for landslide susceptibility mapping

Selection of appropriate landslide affecting or conditioning factors is one of the most important steps in assessing landslide susceptibility of an area [13, 14]. In the present study, 10 important landslide affecting factors: Slope, Aspect, Curvature, Elevation, Land cover, Geomorphology, SFM, Distance to rivers, Distance to roads, Overburden depth were selected based on the experience of researchers considering local topographical and geo-environmental conditions for modeling [15]. Maps of topographical and hydrological factors were developed from the DEM.

Slope

Slope is one of the important factors for the development of a landslide susceptibility map [16, 17]. Most of the landslides occur between slope angles 30° and 45°. The slope map of the study area was prepared from the DEM and classified into five classes using the natural break method of ArcGIS [18]: 0 - 14.15; 14.15 - 23.29; 23.29 - 31.84; 31.84 - 41.57 and 41.57 - 75.19 (Fig. 4a).

Aspect

Fig. 3. Landslide photos of the study area (GSI)

The aspect is a direction of hill slope and is one of the important factors affecting landslide occurrence. The direction of a slope affects the process of landslides because it is related to the wind direction, the direction of rainfall and direction of solar radiation; and also affects moisture of the slope surface [19]. In this study, the slope direction map was extracted from DEM and classified into 9 classes using aspect tool in ArcGIS: Flat, north, northeast, east, southeast, south, southwest, west, and northwest (Fig. 4b).

Curvature

Curvature is one of the important factors affecting landslides [4]. Concave surface accumulates water and thus is more vulnerable to landslides in comparison to convex surface. Landslides are more common in concave surface areas and rare in low dipping flat areas. Curvature map was prepared from DEM and classified into 3 classes using curvature tool in ArcGIS: concave, flat and convex (Fig. 4c).

Elevation

Elevation is an important parameter that affects landslides because this parameter has a

link to the process of rock weathering on hill slopes and formation of soil [20]. Types of vegetation also depend on the elevation. Rainfall and snowfall affecting stability of slopes also depend on the elevation. Elevation map of the hill slopes was developed from DEM and classified into 9 levels (Fig. 4d).

Geomorphology

Landslides are downward movements of soil and rocks due to gravitational effects which are affected by geomorphology and geomorphology processes [21]. Geomorphology of the study area is a result of constant action of rivers and glaciers; and tectonic processes. Geomorphology depends on the geology (rock types and structures) of the area. The main geomorphological features of the area are dissected hills, intermontane valleys, alluvial plains and piedmont slope and rivers. Geomorphological map of the study area was extracted from the GSI report and Google Earth Images and classified into 12 classes (Fig. 4e).

Land cover

Land cover is one of the important factors in controlling shallow landslides [22]. In the study area Land cover map was extracted from the GSI report and Google Earth Images and classified into 11 classes including vegetated area, sparse vegetation, barren land, water bodies (river) etc. (Fig. 4f). Vegetation controls the infiltration and runoff of rainwater and thus affects landslide occurrences. Moreover, presence of sparse vegetation and barren land may make area vulnerable to landslide. Dense vegetation checks the erosion and thus occurrence of shallow landslide events.

Overburden depth

Overburden material plays important role in the shallow landslides [4]. Type of the material determines the infiltration of water and saturation of the groundmass. Moreover, jointed and weathered rock mass above solid fresh rocks also form overburden besides transported loose material underlain by rock mass [23]. Therefore, the depth of the overburden is very important in developing shallow landslides. In this study overburden depth map was extracted from the GSI report and classified into 4 classes (Fig. 4g).

Distance to roads

Table 1. Characteristics of groups of SFM of the study area

No.	SFM groups	Characteristics of SFM groups
1	Group 1	Alluvium
2	Group 2	Chlorite, Schist and Massive Amphibolite
3	Group 3	Dolomite
4	Group 4	Garnet, Mica, Schist and Micaceous Quartzite
5	Group 5	Granite, Gneiss, Garnetiferous Schist & Amphibolite
6	Group 6	In-situ Soil
7	Group 7	Limestone with intercalations of Shale
8	Group 8	Limestone, Dolomite, Shale and Cherty Quartzite
9	Group 9	Metabasite
10	Group 10	Older well compacted Debris
11	Group 11	Phyllite, Stromatoliphyllite, and Stromatolitic
12	Group 12	Phyllite, Stromatolitic Dolomite, Limestone, and Cu & Mg Min
13	Group 13	Schist, Augen Gneiss, Quartzite and Amphibolites
14	Group 14	Slate, Lenses of Quartzite and Dolomite, and Epidorite Dyke
15	Group 15	Slate, Quartzite, and Dolomite with Epidiorite Dyke
16	Group 16	Transported Soil
17	Group 17	Transported Soil and Colluvium

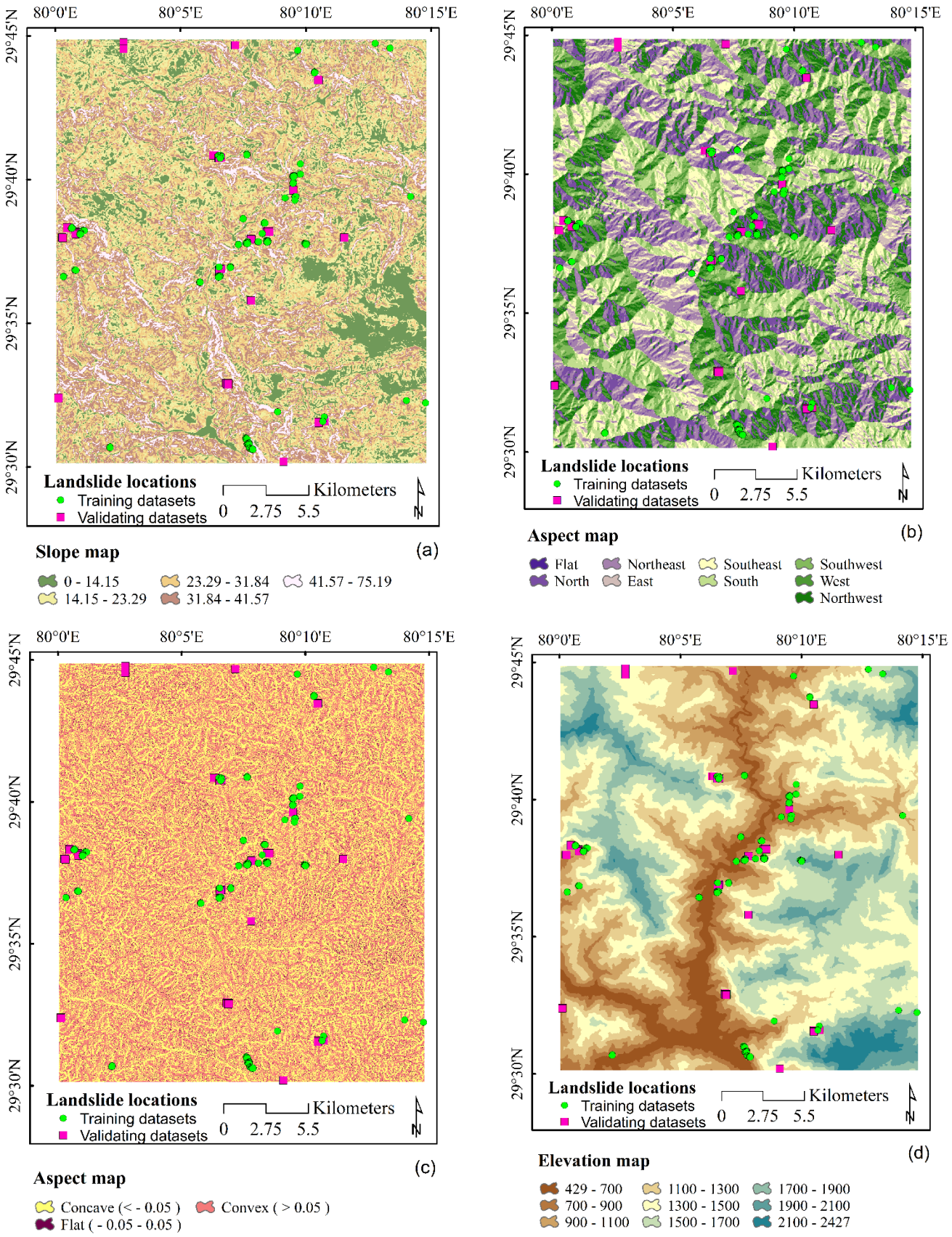

Fig. 4. Thematic maps of the study area: (a) slope, (b) aspect, (c) curvature, (d) elevation, (e) geomorphology, (f) land cover, (g) overburden depth, (h) distance to roads, (i) distance to rivers, and (j) SFM

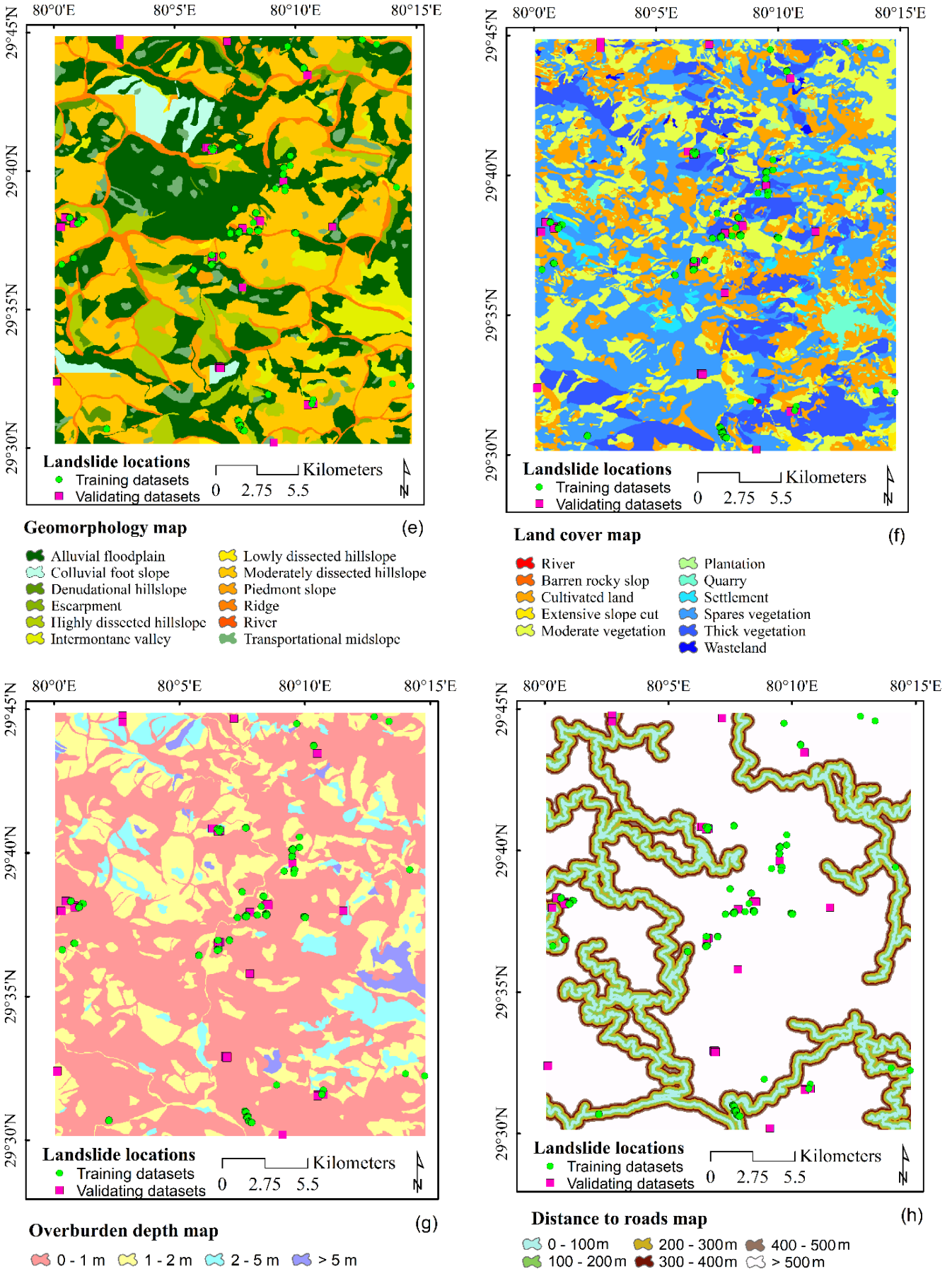

Fig. 4. (continued)

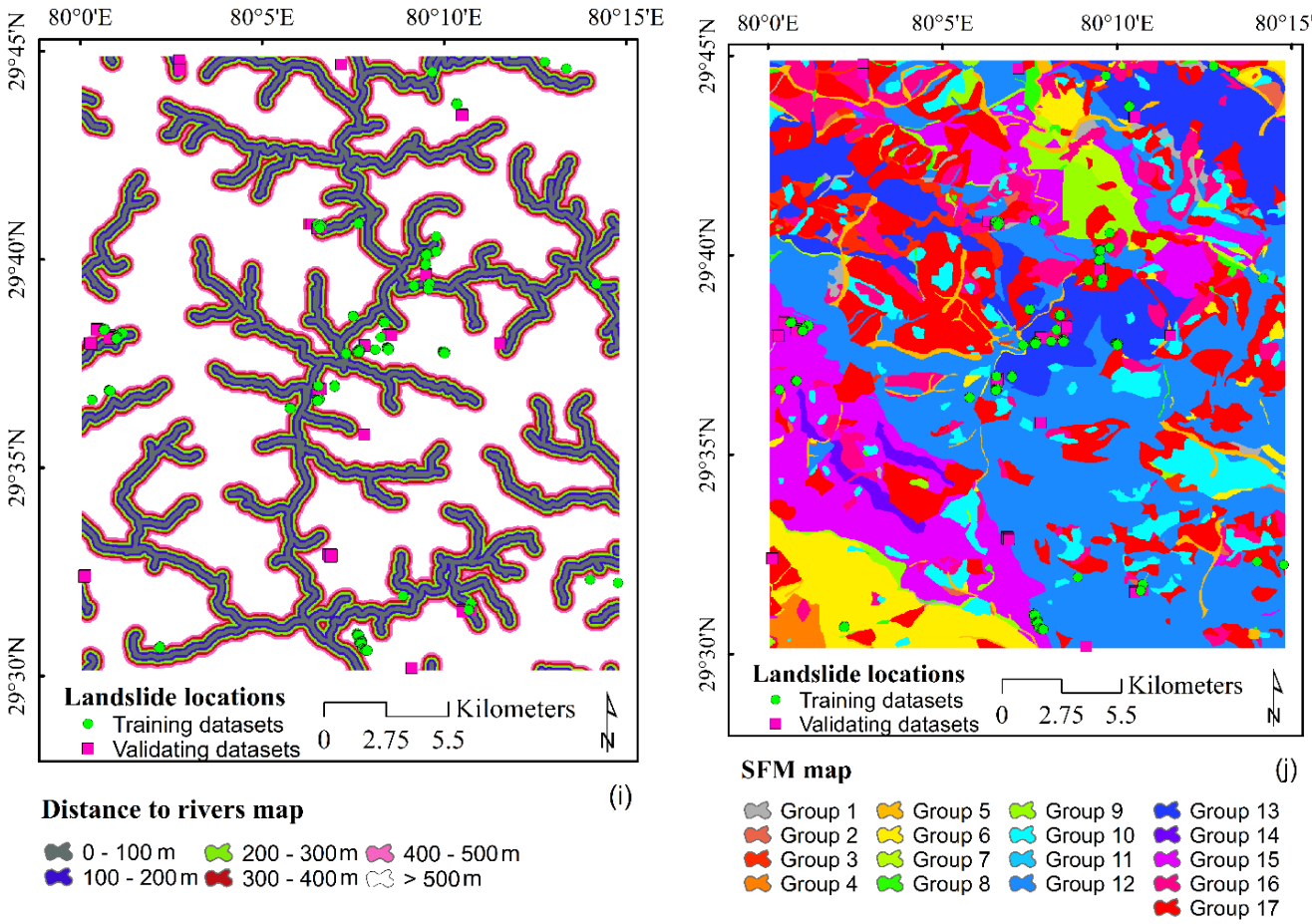

Fig. 4. (continued)

Distance to roads is one of the important factors affecting landslides. Excavation of roads through rock mass/ groundmass for the road construction creates instability of slopes, thus causing landslides along and adjacent roads [24, 25]. The road network map of the study area was created from Google Earth Images and distance road map was classified into 6 classes (Fig. 4h).

Distance to rivers

The distance to rivers and streams affects the process of landslides, because flow of water can erode rocks and soil and thus can create instability of slopes at and near the banks [23]. Moreover, groundmass is likely to be more saturated near rivers, thus may enhance landslides. Moreover, seepage water will flow towards the rivers causing erosion of ground material and also may exert water pressure on the slopes, causing landslides. The river network map was created from DEM and distance to river map was developed and classified into 6 levels (Fig. 4i).

Slope forming material (SFM)

Different types of slopes forming material depend on the lithology of the area and affect occurrence of landslides [26]. The SFM map was extracted from the GSI report and classified into 17 classes (Group 1 to Group 17) (Fig. 4j). The characteristics of SFM material influencing landslides are summarized in Table 1 [12, 27].

In this study, rainfall was not included as a conditioning factor because the objective is to develop spatial landslide susceptibility maps, which evaluate the inherent spatial predisposition of terrain to landsliding rather than temporal triggering processes. Although rainfall is the primary trigger in the Himalayan region, it is treated as a triggering variable and therefore falls outside the scope of susceptibility modeling. Tectonic and structural factors such as faults, folds, and bedding planes were also not incorporated, as these

features exhibit strong local variability and require detailed, high-resolution field data that are not uniformly available across the entire study area. Moreover, the focus of this work is on regional-scale shallow landslides, for which topography, geomorphology, land use/land cover, and SFM-derived terrain attributes generally provide more consistent and robust explanatory power. These considerations guided the selection of conditioning parameters used in the present analysis.

2.4. Methods used

2.4.1. Frequency Ratio (FR)

FR technique is based on the FR value defined as the ratio between the probability of occurrence and the probability of not occurring landslide for specific attributes (causal parameters) [28]. The FR model was used to assess the susceptibility of landslides considering the relationships between past landslides and the influence factor in a given area [13]. The FR approach is one of the easiest to apply and the results are easy to understand [28]. The FR value is determined by the equation:

$$FR = \left[\frac{\frac{N_{pix}(SX_j)}{\sum_{j=1}^{m} N_{pix}(SX_j)}}{\frac{N_{pix}(X_j)}{\sum_{j=1}^{n} N_{pix}(X_j)}} \right] \quad (1)$$

where, $N_{pix}(SX_j)$ is the number of pixels of the landslide causal factor map layer. Also, $N_{pix}(X_j)$ is the number of pixels in the factor map layer to be considered. In the FR method, values higher than one indicate a greater relationship between landslides and attribute-specific factors [29]

2.4.2. Decision Tree (DT)

DT is a type of Supervised Learning algorithm that uses a structured hierarchical tree to classify objects based on a set of criteria. DT is used to solve both classification and regression problems. Classification and Regression Trees (CART), J48, Iterative Dichotomiser 3 (ID3) are some of the DT techniques that have been presented [30, 31]. Quinlan demonstrated the C4.5 algorithm (Class 4.5), which is an advanced variant of ID3 [32].

DT is like a diagram of tree structure in which each leaf node represents a result and each inner node represents a feature (or attribute). The root node is the top node in the DT. It learns to partition based on the value of an attribute. In DT construction, there are two tasks: building the tree and pruning it [33]. The initial step is to create a model, then choose the optimal attribute to split the records using the Attribute Selection Measure (ASM). In this method, dataset is split into smaller subsets by making that attribute a decision node. For each child Node, the process is repeated recursively until all occurrences in the training dataset are assigned to either loss of node or no variable. Pruning is accomplished by deleting unneeded nodes while maintaining classification accuracy [34].

The entropy reduction evaluation criteria were used to identify the input variables, and the information obtained was determined by the formula [30]:

$$\text{Entropy}(E) = -\sum_{i=1}^{n} P_i \log_2 P_i \quad (2)$$

where, n is the number of classes in the domain of the data set E; P_i is the proportion of the number of class i elements over the total number of data set E.

The information gain is used to calculate the expected decrease in entropy at the lower levels of the hierarchy, where data sets are corrected using a different supporting attribute, which is determined by the formula below [30]:

$$\text{Gain}(E,A) = \text{Entropy}(E) - \sum_{v \in \text{Values}(A)} \frac{|E_v|}{|E|} \text{Entropy}(E_v) \quad (3)$$

where, values (A) provide the domain of supporting attribute (A); E_v denotes the subset of E where the corresponding value A is v for each record and $|E_v|$

and |E| denote the cardinality of E_v and E, respectively.

2.4.3. Validation methods

2.4.3.1. Area Under the Receiver Operating Characteristic curve (AUC-ROC)

The ROC curve is used to evaluate the results of model prediction. The best possible prediction method will produce a curve that is a point in the upper left corner of the ROC space, e.g., sensitivity (all true positives) and 1-specificity (no false positives at all). To compare the sensitivity and accuracy of two or more models, Area Under the Curve (AUC) can be used. In the other words, ROC is a probability curve and AUC represents the classification level of the model. The meaning of AUC-ROC can be interpreted as follows: Is the probability that a randomly selected positive sample will rank higher than a randomly selected negative sample. The better model has a higher AUC. An AUC value of 1 indicates a perfect model while a value of 0.5 indicates an incorrect model [35, 36]. The AUC-ROC curve is used as a popular validation method for landslide prediction models [36, 37]. First, the values of 100% sensitivity and specificity were plotted, and then the AUC was used for quantitative evaluation of performance of the models [16].

2.4.3.2. Statistical indices

The landslide susceptibility models were validated using standard statistical indices: Kappa (k), Specificity (SPF), Sensitivity (SST), Negative Predictive Value (NPV), Positive Predictive Value (PPV), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Accuracy (ACC) [38, 39]. The landslide prediction models' dependability was assessed using a k value ranging from '0' to '1' [40]. If k approaches 1, the landslide models are more reliable. SPF and NPV are used to assess how successfully landslide models classify non-slide pixels, whereas SST and PPV are used to classify landslide pixels. RMSE and MAE also indicates model error value, while ACC is used to confirm the model's overall accuracy [39, 41].

3. Results

3.1. Prediction of landslide susceptibility using Frequency Ratio

The results of FR analysis of landslides and influencing factors are presented in Table 2. In the case of slope factor, the highest probability of landslide occurrences is on the slope class 31.84° - 41.57° (FR = 1.685) and in class 41.57° - 75.19° (FR = 3.061). This indicates that areas with moderate to higher slope gradients are more prone to landslides. The results of the slope orientation show that the value of the largest FR index belongs to the slopes oriented in West direction (2.739), which means that landslides in the study area mainly occur on the western facing slopes. As expected, more landslide events occur on concave surfaces with a higher FR value (1.086) in comparison to convex surfaces (FR = 0.972). FR analysis on the Elevation factor indicates that occurrence of landslides is more between 700 and 900 m (3.714) in comparison to other elevations.

The FR value of 3.473 of Extensive slope cut class is the highest in comparison to other land cover classes indicating significant influence from anthropogenic activities. The FR value of SFM Group 11 is 10.848 which is highest of other SFM groups. The FR analysis on geomorphological layer indicates that they can be seen that the FR value of the Piedmont slope layer is 2.501 indicating high occurrence of landslide on Piedmont slope layer.

The analysis results of the distance to rivers show that the FR value of the distance one from 400 to 500m is 1.550 which is highest in comparison to other distances class. The reason at relatively larger distances of landslide occurrences may be attributed to other topographical, geological and locations of alignment of roads. The results of the distance to the road show that the FR value is highest at 2.500 within 0 - 100m as expected due to instability created in ground mass by road cutting within this buffer zone. The FR value of overburden depth at

the top layer 0 - 1 m is 1.142 which only indicates presence of thin slope-wash loose material.

3.2. Prediction of landslide susceptibility using Decision Tree

Table 2. FR analysis of factors causing landslides

Influencing factors	Class	Number of total cells	Number of landslides	Percentage of pixels	Percentage of landslide	Frequency ratio
Slope (°)	0 - 14.15	115596	16	16.070	9.410	0.586
	14.15 - 23.29	193908	14	26.960	8.240	0.305
	23.29 - 31.84	202396	39	28.140	22.940	0.815
	31.84 - 41.57	150676	60	20.950	35.290	1.685
	41.57 - 75.19	56675	41	7.880	24.120	3.061
Aspect	Flat	23	0	0.000	0.000	0.000
	North	102044	9	14.190	5.290	0.373
	Northeast	94824	3	13.180	1.760	0.134
	East	75758	12	10.530	7.060	0.670
	Southeast	87820	20	12.210	11.760	0.964
	South	101979	20	14.180	11.760	0.830
	Southwest	92214	35	12.820	20.590	1.606
	West	78786	51	10.950	30.000	2.739
	Northwest	85803	20	11.930	11.760	0.986
Curvature	Concave (<-0.05)	346678	89	48.200	52.350	1.086
	Flat (-0.05 - 0.05)	33074	3	4.600	1.760	0.384
	Convex (>0.05)	339499	78	47.200	45.880	0.972
Elevation	< 700 m	40514	4	5.630	2.350	0.418
	700 - 900 m	66073	58	9.190	34.120	3.714
	900 - 1100 m	99423	47	13.820	27.650	2.000
	1100 - 1300 m	124702	39	17.340	22.940	1.323
	1300 - 1500 m	141618	10	19.690	5.880	0.299
	1500 - 1700 m	125186	2	17.410	1.180	0.068
	1700 - 1900 m	76329	1	10.610	0.590	0.055
	1900 - 2100 m	33448	9	4.650	5.290	1.138
	2100 - 2427 m	11958	0	1.660	0.000	0.000
SFM	Group 1	6089	0	0.850	0.000	0.000
	Group 2	2651	0	0.370	0.000	0.000
	Group 3	10933	0	1.520	0.000	0.000
	Group 4	4760	0	0.660	0.000	0.000
	Group 5	18037	1	2.510	0.590	0.235
	Group 6	41910	3	5.830	1.760	0.303
	Group 7	5053	0	0.700	0.000	0.000
	Group 8	4049	0	0.560	0.000	0.000
	Group 9	18153	1	2.520	0.590	0.233
	Group 10	46256	0	6.430	0.000	0.000
	Group 11	390	1	0.050	0.590	10.848
	Group 12	202083	30	28.100	17.650	0.628
	Group 13	81422	35	11.320	20.590	1.819
	Group 14	6565	0	0.910	0.000	0.000
	Group 15	91681	75	12.290	28.630	2.825
	Group 16	51515	20	7.160	11.760	1.643
	Group 17	130519	21	18.150	12.350	0.681

Table 2. (continued)

Influencing factors	Class	Number of total cells	Number of landslides	Percentage of pixels	Percentage of landslide	Frequency ratio
Land cover	River	99	0	0.010	0.000	0.000
	Barren rocky slope	274	0	0.040	0.000	0.000
	Cultivated land	160402	23	22.300	13.530	0.607
	Extensive slope cut	9745	8	1.350	4.710	3.473
	Moderate vegetation	155600	65	21.630	38.240	1.767
	Plantation	2778	1	0.390	0.590	1.523
	Settlement	14482	0	2.010	0.000	0.000
	Quarry	14984	8	2.080	4.710	2.259
	Sparse vegetation	257412	51	35.790	30.000	0.838
	Thick vegetation	101464	14	14.110	8.240	0.584
	Wasteland	2011	0	0.280	0.000	0.000
Distance to roads	0 - 100 m	65993	39	9.180	22.940	2.500
	100 - 200 m	56067	8	7.800	4.710	0.604
	200 - 300 m	48168	0	6.700	0.000	0.000
	300 - 400 m	42474	0	5.910	0.000	0.000
	400 - 500 m	38149	0	5.300	0.000	0.000
	> 500 m	468400	123	65.120	72.350	1.111
Overburden depth	0-1 m	463042	125	64.380	73.530	1.142
	1-2 m	196487	38	27.320	22.350	0.818
	2-5 m	47429	7	6.590	4.120	0.624
	> 5 m	12293	0	1.710	0.000	0.000
Distance to rivers	0 - 100 m	66102	7	9.190	4.120	0.448
	100 - 200 m	63794	20	8.870	11.760	1.326
	200 - 300 m	61801	16	8.590	9.410	1.095
	300 - 400 m	59681	8	8.300	4.710	0.567
	400 - 500 m	57324	21	7.970	12.350	1.550
	> 500 m	410549	98	57.080	57.650	1.010
Geomorphology	Alluvial floodplain	257437	54	35.790	31.760	0.887
	Colluvial foot slope	19761	0	2.750	0.000	0.000
	Denudational hillslope	12178	3	1.690	1.760	1.042
	Escarpment	5702	0	0.790	0.000	0.000
	Highly dissected hillslope	47073	24	6.540	14.120	2.157
	Intermontane valley	30714	0	4.270	0.000	0.000
	Lowly dissected hillslope	2570	0	0.360	0.000	0.000
	Moderately dissected hillslope	262462	76	36.490	44.710	1.225
	Piedmont slope	11840	7	1.650	4.120	2.501
	Ridge	46194	6	6.420	3.530	0.550
	River	79	0	0.010	0.000	0.000
	Transportation mid-slope	23241	0	3.230	0.000	0.000

Table 3. Accuracy analysis of the DT model for training and testing datasets

Parameters	PPV (%)	NPV (%)	SST (%)	SPF (%)	ACC (%)	Kappa (k)	MAE	RMSE
Training dataset	89.41	83.53	82.93	89.83	91.47	0.829	0.147	0.271
Testing dataset	84.21	78.95	80.00	83.33	81.58	0.632	0.229	0.399

In landslide modeling using DT model, we have assigned landslide label as "1" and non-landslide label as "0" in the training and testing dataset. Statistical indices: PPV, NPV, SST, SPF, Kappa, ACC, MAE and RMSE were used to evaluate performance of the DT model (Table 3). The results show that classification probability of the PPV is high for both landslide class on training dataset (89.41%) and testing dataset (84.21%). Similarly, PPV value for correctly classification of non-landslide pixels in the area on training dataset is 83.53% and on testing dataset is 78.95%. The model with high SST value on training set (82.93%) and on testing dataset (80.00%) demonstrates that the landslide points are correctly classified as landslide. High SPF for training data (89.83%) and testing data (83.33%) show correctly classified non-slide area. The Kappa Index for training data (0.829) and for testing data (0.632), represents a significant agreement as a combination of predictions and reality. Low MAE values for training data (0.147) and for testing data (0.229) and high values of ACC for training (91.47%) data and for testing data (81.58%) show that the model has high accuracy. RMSE value of prediction of DT model for training data is 0.271 and testing data is 0.399 (Fig. 5). The actual values (shown by the blue line) are very close to the predicted values (shown by the red and orange line). It indicates that DT model has good predictive capability.

3.3. Validation and comparison of the models

The more the ROC curve is skewed to the top and to the left, the higher the probability of correct prediction. At the same time, the area under the AUC curve will be larger. As shown in Fig. 6, the DT model (AUC>0.8) gives much higher accuracy than FR (AUC<0.65). On the training data, the DT model with AUC=0.946 was 45.53% higher than the FR model with AUC=0.650, whereas on the testing data, the DT model with AUC=0.848 was 46.71% higher than the FR model with AUC=0.578. Furthermore, the FR model shows less accuracy when the ROC curve is partially below the Reference line (AUC=0.5). The AUC-ROC results indicate that DT exhibits superior quality with much higher accuracy in prediction than FR model.

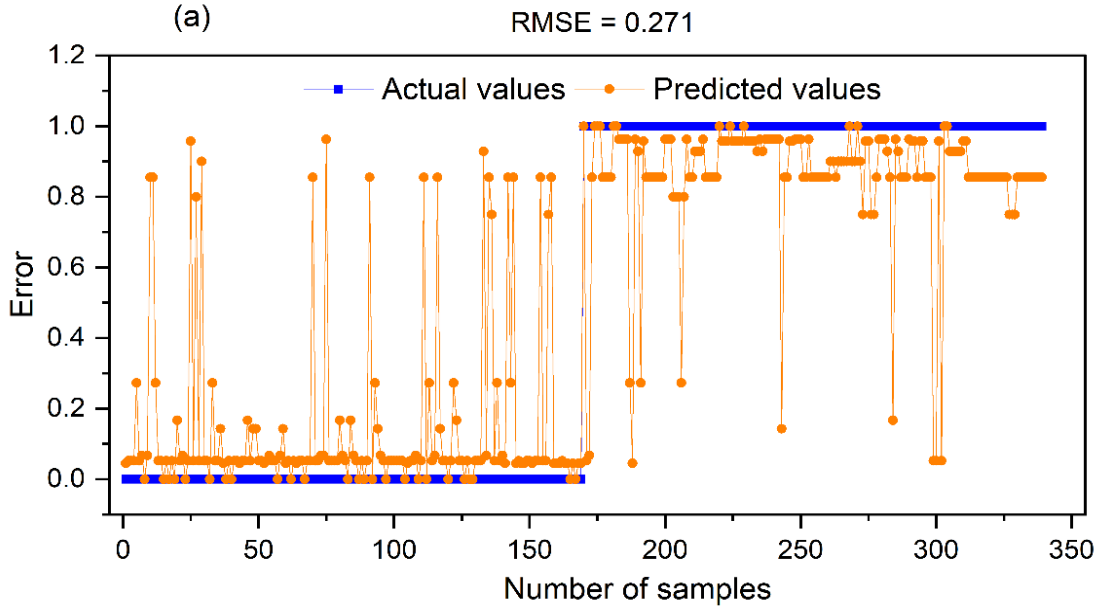

Fig. 5. RMSE analysis of the models: (a) training, (b) testing

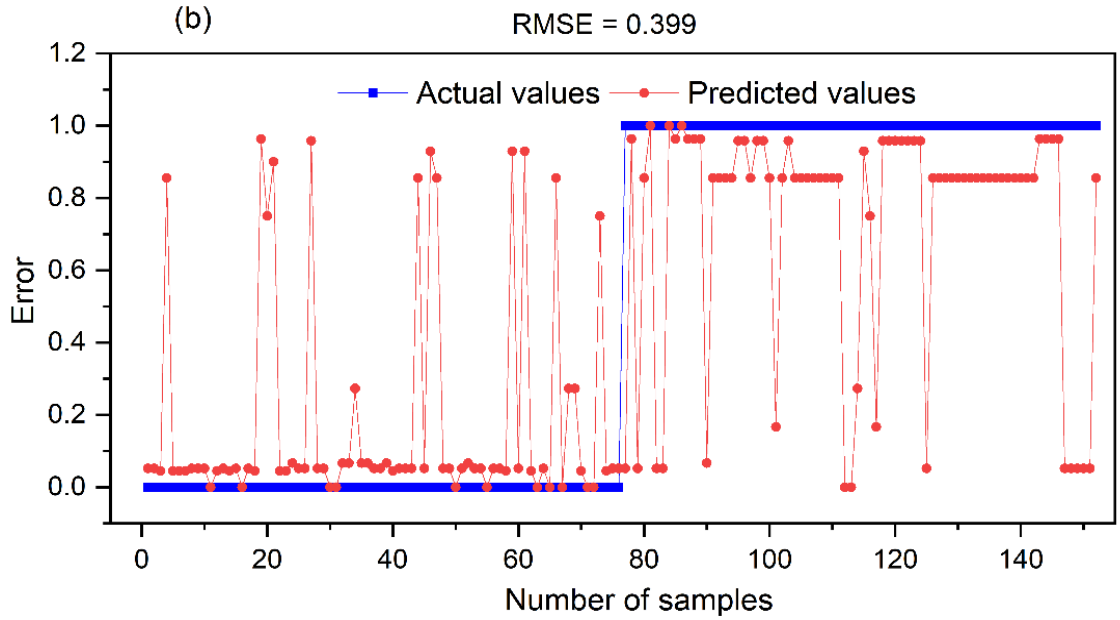

Fig. 5. (continued)

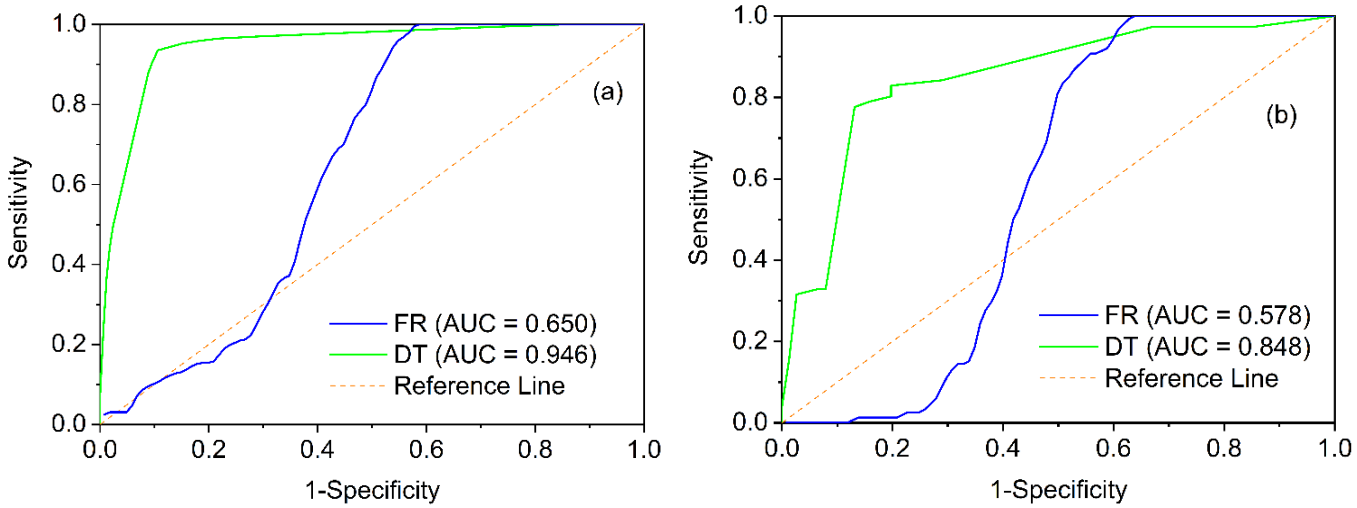

Fig. 6. AUC analysis of the models using (a) training dataset and (b) validation dataset

In this study, the original continuous values were directly utilized by the Decision Tree model, which inherently handles both continuous and categorical variables; therefore, additional reclassification is neither required nor feasible at this stage. Reclassification was applied only for the Frequency Ratio model and for map visualization purposes, where class boundaries were defined using natural break optimization to best preserve data distribution. Given these considerations, the current classification approach is sufficient and appropriate for the modeling framework adopted in this study.

3.4. Construction of landslide susceptibility maps

Susceptibility maps were created using the results of the proposed landslide models (FR and DT). Landslide susceptibility zoning was developed by categorizing pixels in the study region based on their landslide probability estimates. In this work, the Natural Breaks approach was used to identify probability values with 70% of the landslide data in model building. In addition, 30% of landslide data were used to verify the reliability of the research area's landslide susceptibility zoning map by overlapping these real landslides on the landslide

susceptibility zone from susceptibility maps (Fig. 7). For landslide susceptibility mapping, five different levels were identified: very low, low, moderate, high, and very high.

The FR model was used to create a landslide susceptibility map with five levels: Very low susceptibility, Low susceptibility, moderate susceptibility, High susceptibility, Very high susceptibility (Fig. 7a). It can be seen that approximately 10.16% of the research region is classified as low-susceptibility, 8.98% as Moderate-susceptibility, and 11.32% as high-susceptibility. Results indicated that around 39.08% of landslides were found in very high-susceptibility locations, and 30.46% in very low-susceptibility areas.

With the DT model, a landslide susceptibility with 5 levels was built including, Very low susceptibility, Low susceptibility, moderate susceptibility, High susceptibility, Very high susceptibility (shown as Fig. 7b). It can be seen that about 6.7% of the study area is in the very low-susceptibility class and 20.6% in the low-susceptibility class. The results based on the testing data show that about 63% of landslides were located in high- and very high-susceptibility areas, indicating the good predictive capability of the DT model.

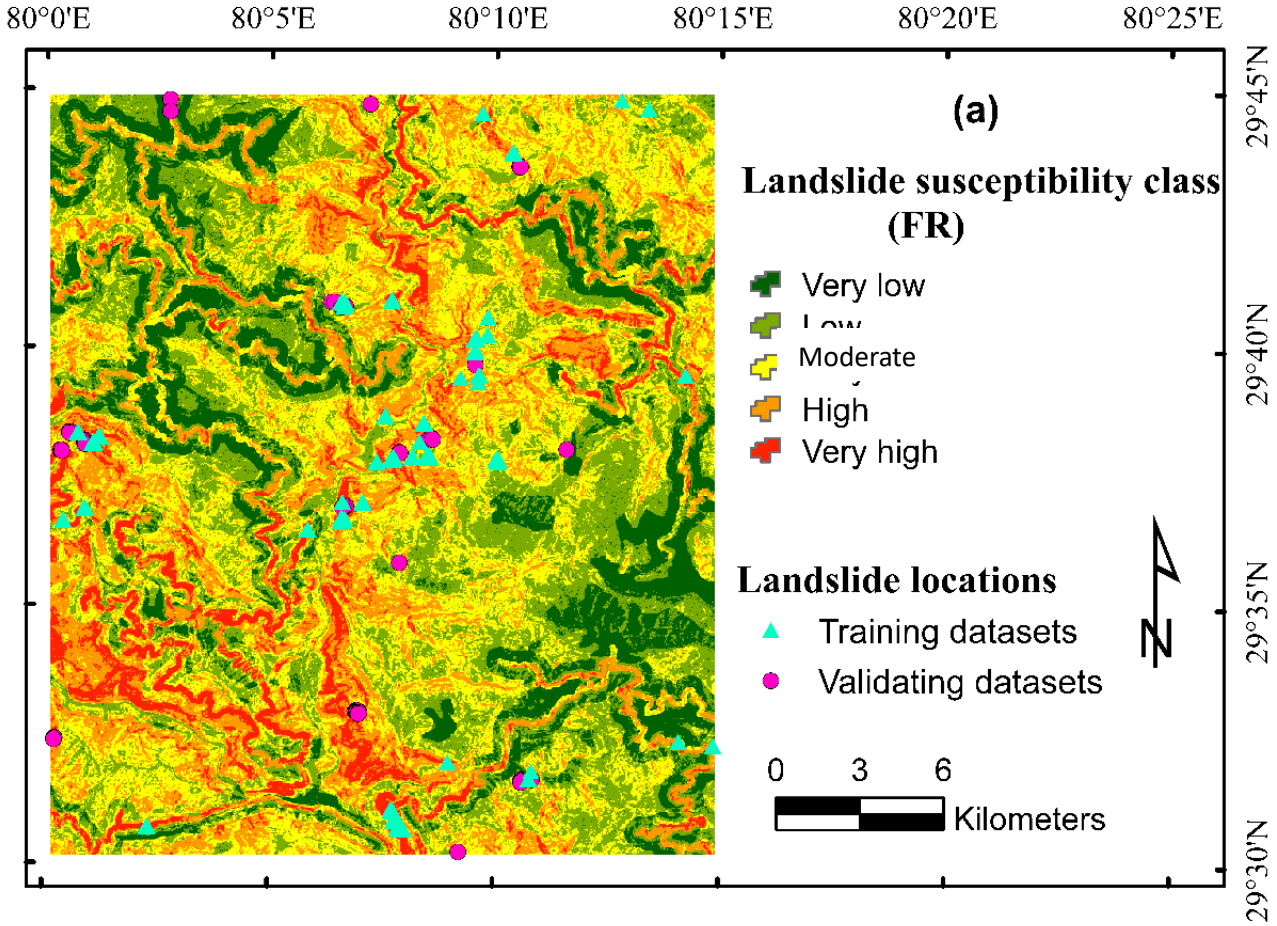

Fig. 7. Landslide susceptibility maps using the models: (a) FR, (b) DT

The landslide susceptibility map was developed on the basis of models that were validated by analysis of Frequency Ratio (FR) and percentage of landslide pixels of each susceptibility class for each model on the map (Fig. 8). The FR results obtained from the two models used were similar. However, in the high-susceptibility class, the DT model showed a higher FR value than the FR model, indicating a stronger association between this susceptibility class and the observed landslide distribution. In contrast, the percentage of landslide pixels results for each class of the two

models were similar at high-susceptibility level (nearly 40%) and significantly different at very low-susceptibility level in which FR-based validation further confirms that the DT model shows a stronger association between higher susceptibility classes and the observed landslide distribution than the FR model.

4. Discussion

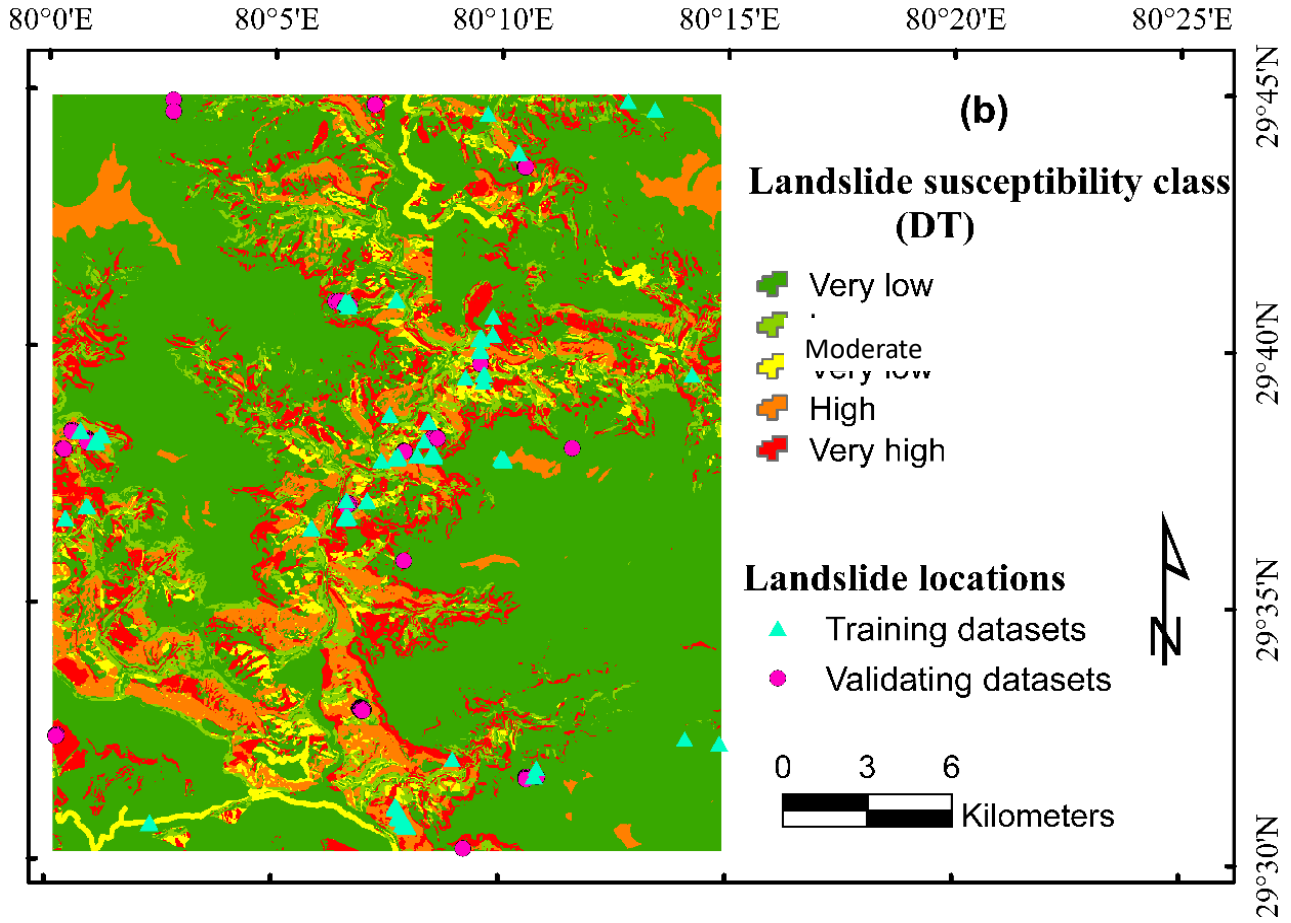

Fig. 7. (continued)

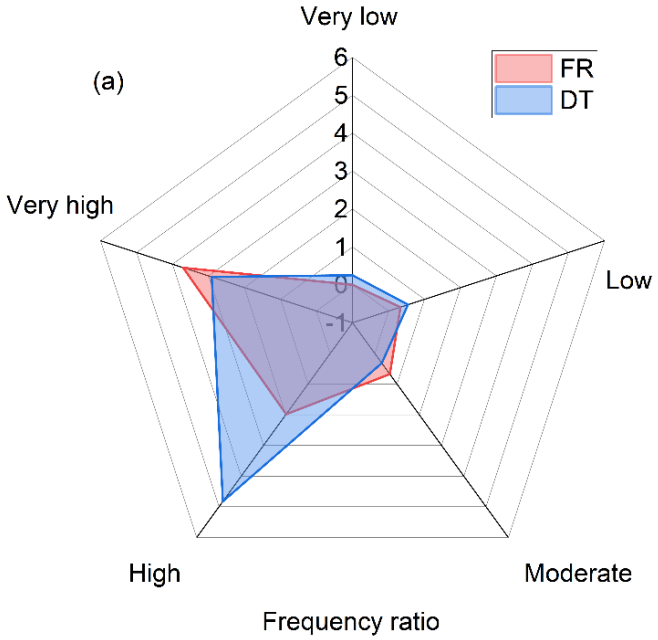

Fig. 8. Analysis of FR on the susceptibility maps using the models

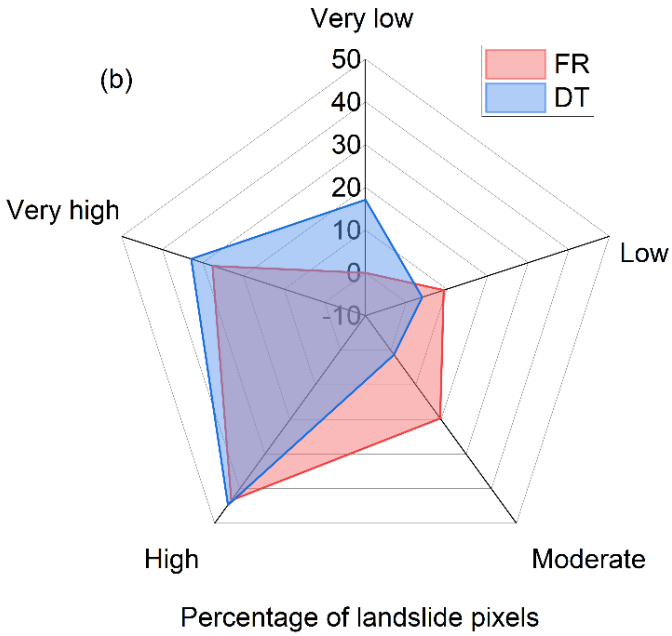

Fig. 8. (continued)

Development of the landslide susceptibility maps is often regarded as the first and most significant step in assessing landslide hazard and risk in order to improve landslide management of any area. Therefore, it is desirable to develop and apply a number of models capable of accurately forecasting landslide-prone locations. In this study, FR and DT models were used for landslide susceptibility mapping. Results showed that the accuracy of the DT model is significantly better than FR model in predicting landslide susceptibility at the study area. It is appropriate as the DT has several advantages such as (i) It presents a clear and simple structure that breaks down complex problems into smaller, more manageable components, (ii) It is able to handle missing values in the data which is a common problem in real-world problems including landslide susceptibility prediction, and it is capable of handling datasets with high errors and noisy values, (iii) it is capable of capturing complex non-linear relationships between input features and the target variables used in landslide susceptibility prediction, (iv) it requires little data preparation, such as normalization or scaling, and thus it can efficiently handle large datasets, and (v) It is considered as a nonparametric technique which requires no assumptions about the space distribution and on the structure of classifier. In contrast, FR model faces several disadvantages such as (i) it assumes independence between the predictor variables, which may not always be the case in real-world datasets including prediction of landslide susceptibility, (ii) it is susceptible to bias towards the majority class if the data is imbalanced, (iii) it only considers the frequency of occurrence of each class and does not take into account other factors that may influence the prediction.

In comparison, the modeling results of this study are comparable to those of previous landslide susceptibility studies. A decision tree (DT) model applied to landslide analysis in Istanbul, Turkey, achieved an AUC of 89.6% [42]. Another DT-based study conducted in Phidim, Nepal, reported an AUC of 95.9% [3]. In addition, a landslide susceptibility study in Turkey showed that the frequency ratio (FR) method had lower predictive performance (AUC = 0.826) than the artificial neural network (ANN) model (AUC = 0.852) [21]. It can be seen that when it comes to analyzing and mapping susceptible landslides, ML approaches (DT or ANN) outperform classical methods (FR). Predictability, on the other hand, varies from case to case, based on the input data

(number of variables, order, kind of variables, and so on) utilized in each model. A previous study conducted in the same area used three machine learning models, including multilayer perceptron (MLP), Naïve Bayes (NB), and alternating decision tree (ADT), for landslide spatial prediction [12]. Compared with that study, the DT model used in this work (AUC = 0.848) outperformed the ADT model (AUC = 0.840), but showed lower performance than the NB (AUC = 0.873) and MLP (AUC = 0.864) models [12]. A number of hybrid and ensemble ML models have been applied for the generation of landslide susceptibility maps but performance of simple DT model is comparable in performance to other complex models, thus it can be applied for landslide susceptibility mapping. As discussed above, it can be seen that DT based ML method is a powerful tool in creating high reliable landslide susceptibility maps in comparison to FR method.

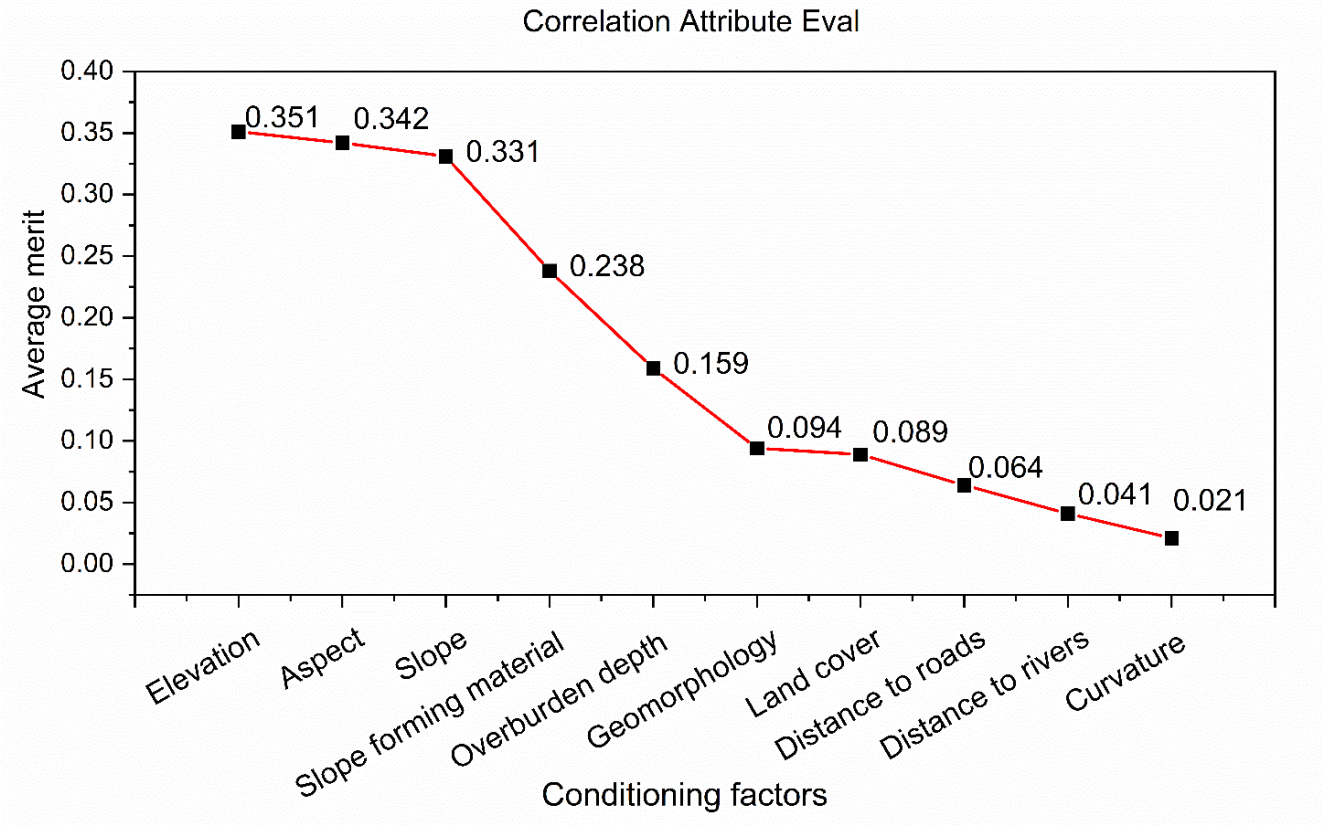

Fig. 9. Feature selection analysis of landslide conditioning factors

In order to evaluate influence of the conditioning factors in landslide modeling, the feature selection analysis was done using Correlation based feature selection method, which is one of the most popular feature selection methods for evaluation of importance of landslide conditioning factors used in landslide susceptibility modeling [43, 44]. Fig. 9 shows the importance of the landslide conditioning factors used in this study. It can be observed that the topographical factors such as elevation, aspect and slope are the most important factors influencing the landslide susceptibility modeling in this study, which is also in line with the published works [36]. Out of these, elevation is the most important factor for prediction of landslide susceptibility in this study because it can influence several other factors that contribute to landslides [45]. For example, higher elevations generally receive more rainfall, which can increase the weight of the soil and make it more susceptible to movement of the slope [46, 47]. Higher elevations also tend to have steeper slopes, which can increase the gravitational force acting on the slope materials. In addition, higher elevations are

often located in areas with less vegetation cover, which can reduce the stability of the slope materials [48].

It is true rainfall is important factor in the occurrence of landslides. Most of the landslides in the study area occur during monsoon period where other topographical, geological and geo-environmental factors are favorable. Therefore, in this study we have considered rainfall as triggering factor and not contributing factor considering local conditions.

Plate tectonic movements in Himalayan terrain is certainly playing role in the stress accumulation and release in different sections of Himalayan arch which are measured by instrumental data and in the field and by remote sensing data. Thrusting of Indian plate and uplifting of the Himalayas are slow process. Earthquakes generated by tectonic forces. These tectonic forces are very slow in rock movements within short period unless they are due to sudden release of stresses in form of earthquakes causing landslides. It requires time series real time studies even then prediction of earthquakes accurately in any area is not possible. Therefore, tectonic forces have not been considered in the modeling.

Moreover, geological structural features such as bedding planes play roles in landslides. However, these features can be incorporated in local individual landslide studies. Our study is for identifying the landslide susceptible areas on regional scale. Therefore, general geology, topography and other geo-environmental factors have been considered in the model studies. Impact of faults, folds, bedding planes, shears, infilling material along discontinuities etc. may be considered locally based on the developed susceptibility maps by the proposed model studies for the treatment of individual landslides. In this study we have considered shallow landslide problems. In future studies we will consider deeper geological features in the model studies. Detailed subsurface geology can be considered in future study on large scale mapping.

5. Concluding remarks

This study demonstrates that simple, interpretable approaches such as the Frequency Ratio method and the Decision Tree model can effectively delineate landslide susceptibility zones in hilly and mountainous regions, using the Uttarakhand Himalaya as a representative example. The comparative evaluation reveals that the DT model offers substantially better predictive performance than the FR method, as reflected by its higher AUC value (0.848), indicating its suitability for susceptibility mapping not only in the study area but also in other Himalayan terrains and similar mountainous environments. The findings reinforce the growing evidence that machine-learning methods, even relatively simple ones like DT, provide improved predictive capability compared with traditional statistical techniques, particularly in regions characterised by complex terrain and heterogeneous geo-environmental conditions. While the present study provides a robust susceptibility assessment based on the available regional-scale datasets and ten representative conditioning factors, it is constrained by the absence of more detailed structural, hydrological, and subsurface information that could further refine model outputs. Future studies may incorporate larger and more detailed inventories, additional conditioning parameters, and advanced single or hybrid machine-learning approaches to enhance model generalisability and support broader landslide hazard management efforts across diverse mountainous regions.

Funding

The University of Transport Technology funds this research under the project entitled "Assessment of land subsidence susceptibility using artificial intelligence technology integrated with geospatial techniques", grant number ĐTTĐ2023-19.

References

[1]. D.H. Cornforth. (2005). Landslides in practice: investigation, analysis, and remedial/preventative options in soils. J. Wiley, Hoboken, N.J. https://lccn.loc.gov/2004007921

[2]. I. Alcántara-Ayala. (2002). Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries, *Geomorphology*, 47(2-4), 107-124. https://doi.org/10.1016/S0169-555X(02)00083-1

[3]. C.P. Poudyal. (2012). Landslide susceptibility analysis using decision tree method, Phidim, Eastern Nepal. *Bulletin of the Department of Geology*, 15, 69-76. https://doi.org/10.3126/bdg.v15i0.7419

[4]. E.M. Lee. (2009). Landslide risk assessment: the challenge of estimating the probability of landsliding. *Quarterly Journal of Engineering Geology and Hydrogeology,* 42(4), 445-458. https://doi.org/10.1144/1470-9236/08-007

[5]. M.S. Alkhasawneh, U.K. Ngah, L.T. Tay, N.A.M. Isa, M.S. Al-Batah. (2014). Modeling and testing landslide hazard using decision tree. *Journal of Applied Mathematics*, 2014, 929768.https://doi.org/10.1155/2014/929768

[6]. M. Pal, P.M. Mather. (2003). An assessment of the effectiveness of decision tree methods for land cover classification, Remote sensing of environment, 86(4), 554-565. https://doi.org/10.1016/S0034-4257(03)00132-9

[7]. H.S.B. Düzgün, A. Özdemir. (2006). Landslide risk assessment and management by decision analytical procedure for Dereköy, Konya, Turkey. *Natural Hazards*, 39, 245-263. https://doi.org/10.1007/s11069-006-0026-6

[8]. S. Wan, T.C. Lei, T.Y. Chou. (2010). A novel data mining technique of analysis and classification for landslide problems. *Natural Hazards*, 52, 211-230. https://doi.org/10.1007/s11069-009-9366-3

[9]. X. Wang, R. Niu. (2009). Spatial forecast of landslides in three gorges based on spatial data mining. *Sensors*, 9(3), 2035-2061. https://doi.org/10.3390/s90302035

[10]. F.E.S. Silalahi, Pamela, Y. Arifianti, F. Hidayat. (2019). Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. *Geoscience Letters*, 6, 10. https://doi.org/10.1186/s40562-019-0140-4

[11]. H. Saito, D. Nakayama, H. Matsuyama. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. *Geomorphology*, 109(3-4), 108-121. https://doi.org/10.1016/j.geomorph.2009.02.026

[12]. T.-H. Tran, N.D. Dam, F.E. Jalal, N. Al-Ansari, L.S. Ho, T.V. Phong, M. Iqbal, H.V. Le, H.B.T. Nguyen, I. Prakash, B.T. Pham. (2021). GIS-Based Soft Computing Models for Landslide Susceptibility Mapping: A Case Study of Pithoragarh District, Uttarakhand State, India. *Mathematical Problems in Engineering*, 2021, 9914650. https://doi.org/10.1155/2021/9914650

[13]. M. Gholami, E.N. Ghachkanlu, K. Khosravi, S. Pirasteh. (2019). Landslide prediction capability by comparison of frequency ratio, fuzzy gamma and landslide index method. *Journal of Earth System Science*, 128, 42. https://doi.org/10.1007/s12040-018-1047-8

[14]. B.T. Pham, A. Jaafari, I. Prakash, D.T. Bui. (2019). A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bulletin of Engineering Geology and the Environment*, 78, 2865-2886. https://doi.org/10.1007/s10064-018-1281-y

[15]. B.T. Pham, T.V. Phong, T. Nguyen-Thoi, P.T. Trinh, Q.C. Tran, L.S. Ho, S.K. Singh, T.T.T. Duyen, L.T. Nguyen, H.Q. Le, H.V. Le, N.T.B. Hanh, N.K. Quoc, I. Prakash. (2020). GIS-based ensemble soft computing models for

landslide susceptibility mapping. *Advances in Space Research*, 66(6), 1303-1320. https://doi.org/10.1016/j.asr.2020.05.016

[16]. K.-T. Chang, A. Merghadi, A.P. Yunus, B.T. Pham, J. Dou. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, 9, 12296. https://doi.org/10.1038/s41598-019-48773-2

[17]. E. Nohani, M. Moharrami, S. Sharafi, K. Khosravi, B. Pradhan, B.T. Pham, S. Lee, A.M. Melesse. (2019). Landslide susceptibility mapping using different GIS-based bivariate models. *Water*, 11(7), 1402. https://doi.org/10.3390/w11071402

[18]. A. Basofi, A. Fariza, A.S. Ahsan, I.M. Kamal. (2015). A comparison between natural and Head/tail breaks in LSI (Landslide Susceptibility Index) classification for landslide susceptibility mapping: A case study in Ponorogo, East Java, Indonesia. *2015 International Conference on Science in Information Technology (ICSITech), IEEE, pp 337-342*.doi: 10.1109/ICSITech.2015.7407828

[19]. H.-J. Oh, S. Lee. (2011). Landslide susceptibility mapping on Panaon Island, Philippines using a geographic information system. *Environmental Earth Sciences*, 62, 935-951. https://doi.org/10.1007/s12665-010-0579-2

[20]. P.V. Gorsevski, P. Jankowski. (2010). An optimized solution of multi-criteria evaluation analysis of landslide susceptibility using fuzzy sets and Kalman filter. *Computers & Geosciences*, 36(8), 1005-1020. https://doi.org/10.1016/j.cageo.2010.03.001

[21]. I. Yilmaz. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences*, 35(6), 1125-1138. https://doi.org/10.1016/j.cageo.2008.08.007

[22]. J. Barlow, Y. Martin, S.E. Franklin. (2003). Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Canadian Journal of Remote Sensing*, 29(4), 510-517. https://doi.org/10.5589/m03-018

[23]. H. Moayedi, M. Mehrabi, M. Mosallanezhad, A.S.A. Rashid, B. Pradhan. (2019). Modification of landslide susceptibility mapping using optimized PSO-ANN technique. *Engineering with Computers*, 35, 967-984. https://doi.org/10.1007/s00366-018-0644-0

[24]. O.F. Althuwaynee, B. Pradhan, S. Lee. (2012). Application of an evidential belief function model in landslide susceptibility mapping. *Computers & Geosciences*, 44, 120-135. https://doi.org/10.1016/j.cageo.2012.03.003

[25]. H.-J. Oh, B. Pradhan. (2011). Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Computers & Geosciences*, 37(9), 1264-1276. https://doi.org/10.1016/j.cageo.2010.10.012

[26]. S. He, P. Pan, L. Dai, H. Wang, J. Liu. (2012). Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology*, 171-172, 30-41.https://doi.org/10.1016/j.geomorph.2012.04.024

[27]. N.D. Dam, M. Amiri, N. Al-Ansari, I. Prakash, H.V. Le, H.B.T. Nguyen, B.T. Pham. (2022). Evaluation of Shannon entropy and weights of evidence models in landslide susceptibility mapping for the Pithoragarh district of Uttarakhand state, India. *Advances in Civil Engineering*, 2022, 6645007. https://doi.org/10.1155/2022/6645007

[28]. A. Shirzadi, K. Chapi, H. Shahabi, K. Solaimani, A. Kavian, B.B. Ahmad. (2017).

Rock fall susceptibility assessment along a mountainous road: an evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. *Environmental Earth Sciences*, 76, 152. https://doi.org/10.1007/s12665-017-6471-6

[29]. B.T. Pham, D.T. Bui, I. Prakash, M. Dholakia. (2015). Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS–based statistical approach of frequency ratio method. *International Journal of Engineering Research & Technology (IJERT)*, 4(11), 338-344.

[30]. J.R. Quinlan. (1986). Induction of decision trees. *Machine Learning*, 1, 81-106. *1986 Kluwer Academic Publishers, Boston*. https://doi.org/10.1007/BF00116251

[31]. J.R. Quinlan. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3), 221-234. https://doi.org/10.1016/S0020-7373(87)80053-6

[32]. J.R. Quinlan. (2014). C4. 5: Programs for Machine Learning. *Elsevier*.

[33]. D.T. Bui, B. Pradhan, I. Revhaug, C.T. Tran. (2014). 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. *Remote Sensing Applications in Environmental Research. Society of Earth Scientists Series. Springer, pp. 87-111*. https://doi.org/10.1007/978-3-319-05906-8_6

[34]. M. Ahmadlou, M. Karimi, S. Alizadeh, A. Shirzadi, D. Parvinnejhad, H. Shahabi, M. Panahi. (2019). Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto International*, 34(11), 1252-1272. https://doi.org/10.1080/10106049.2018.14742 76

[35]. H. Hong, J. Liu, D.T. Bui, B. Pradhan, T.D. Acharya, B.T. Pham, A.-X. Zhu, W. Chen, B.B. Ahmad. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399-413. https://doi.org/10.1016/j.catena.2018.01.005

[36]. D.T. Bui, T.-C. Ho, B. Pradhan, B.T. Pham, V.-H. Nhu, I. Revhaug. (2016). GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences*, 75, 1101. https://doi.org/10.1007/s12665-016-5919-4

[37]. S. Beguería. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*, 37, 315-329. https://doi.org/10.1007/s11069-005-5182-6

[38]. D.-H. Lee, Y.-T. Kim, S.-R. Lee. (2020). Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. *Remote Sensing*, 12(7), 1194. https://doi.org/10.3390/rs12071194

[39]. V.-H. Nhu, A. Mohammadi, H. Shahabi, B.B. Ahmad, N. Al-Ansari, A. Shirzadi, M. Geertsema, V.R. Kress, S. Karimzadeh, K.V. Kamran, W. Chen, H. Nguyen. (2020). Landslide detection and susceptibility modeling on cameron highlands (Malaysia): A comparison between random forest, logistic regression and logistic model tree algorithms. *Forests*, 11(8), 830. https://doi.org/10.3390/f11080830

[40]. D.T. Bui, B. Pradhan, O. Lofman, I. Revhaug. (2012). Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical Problems in*

Engineering, 2012, 974638. https://doi.org/10.1155/2012/974638

[41]. P. Yariyan, S. Janizadeh, T.V. Phong, H.D. Nguyen, R. Costache, H.V. Le, B.T. Pham, B. Pradhan, J.P. Tiefenbacher. (2020). Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. *Water Resources Management*, 34, 3037-3053. https://doi.org/10.1007/s11269-020-02603-7

[42]. H.A. Nefeslioglu, E. Sezer, C. Gokceoglu, A.S. Bozkir, T.Y. Duman. (2010). Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering*, 2010, 901095. https://doi.org/10.1155/2010/901095

[43]. M.R. Mezaal, B. Pradhan. (2018). An improved algorithm for identifying shallow and deep-seated landslides in dense tropical forest from airborne laser scanning data. *Catena*, 167, 147-159. https://doi.org/10.1016/j.catena.2018.04.038

[44]. Vasu, N.N., Lee, S.-R. (2016). A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea. *Geomorphology*, 263, 50-70. https://doi.org/10.1016/j.geomorph.2016.03.023

[45]. D.J. Varnes. (1978). Slope movement types and processes. *Landslides, Analysis and Control, Transportation Research Board, Special Report No. 176, National Academy of Sciences, pp. 11-33*.

[46]. R.K. Dahal. (2012). Rainfall-induced landslides in Nepal. *International Journal of Erosion Control Engineering*, 5(1), 1-8. https://doi.org/10.13101/ijece.5.1

[47]. P.T. Nguyen, T.T. Tuyen, A. Shirzadi, B.T. Pham, H. Shahabi, E. Omidvar, A. Amini, H. Entezami, I. Prakash, T.V. Phong, T.B. Vu, T. Thanh, L. Saro, D.T. Bui. (2019). Development of a novel hybrid intelligence approach for landslide spatial prediction. *Applied Sciences*, 9(14), 2824. https://doi.org/10.3390/app9142824

[48]. F.C. Dai, C.F. Lee, S.J. Wang. (2003). Characterization of rainfall-induced landslides. *International Journal of Remote Sensing*, 24(23), 4817-4834. https://doi.org/10.1080/01431160113100008242 4