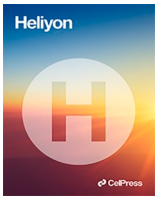
Heliyon 10 (2024) e23424

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh

Md Sharafat Chowdhury [a,b,*], Md Naimur Rahaman [a], Md Sujon Sheikh [a], Md Abu Sayeid [a], Khandakar Hasan Mahmud [a], Bibi Hafsa [a]

[a] Department of Geography and Environment, Jahangirnagar University, Savar, Dhaka, Bangladesh
[b] Information and Communication Technology Division, Dhaka, Bangladesh

ARTICLE INFO

Keywords:
Geographic information systems
Logistic regression
Random forest
Decision and regression tree
AUC of ROC
Landslide susceptibility map

ABSTRACT

The frequency of landslides and related economic and environmental damage has increased in recent decades across the hilly areas of the world, no exception is Bangladesh. Considering the first step in landslide disaster management, different methods have been applied but no methods found as best one. As a result, landslide assessment using different methods in different geographical regions has significant importance. The research aims to prepare and evaluate landslide susceptibility maps (LSMs) of the Chattogram district using three machine learning algorithms of Logistic Regression (LR), Random forest (RF) and Decision and Regression Tree (DRT). Sixteen landslide conditioning factors were determined considering topographic, hydro-climatic, geologic and anthropogenic influence. The landslide inventory database (255 locations) was randomly divided into training (80 %) and testing (20 %) sets. The LSMs showed that almost 9–12 % of areas of the Chattogram district are highly susceptible to landslides. The highly susceptible zones cover the Chattogram district's hill ranges where active morphological processes (erosion and denudation) are dominant. The ROC values for training data were 0.943, 0.917 and 0.947 and testing data were 0.963, 0.934 and 0.905 for LR, RF and DRT models, respectively. The accuracy is higher than the previous research in comparison to the extent of the study area and the size of the inventory. Among the models, LR showed the highest prediction rate and DRT showed the highest success rate. According to susceptibility zones, DRT is the more realistic model followed by LR. The maps can be applied at the local scale for landslide hazard management.

1. Introduction

The frequency, intensity and uncertainty of all types of natural disasters are significantly increasing across the world driven by the adverse impact of climate. According to the EM-DAT database, of 16472 natural disaster records, 40.2 % of them occurred in Asia and 12.75 % of all disasters occurred in Southern Asia. Landslide alone comprises 5.08 % of all the natural disasters occurred worldwide. Asia alone is affected by 53.88 % of all landslides that occur globally and the Asian landslide alone shares 2.74 % of all the natural

* Corresponding author. Department of Geography and Environment, Jahangirnagar University, Savar, Dhaka, Bangladesh.
E-mail address: sharafat.44@geography-juniv.edu.bd (M.S. Chowdhury).

https://doi.org/10.1016/j.heliyon.2023.e23424
Received 21 August 2023; Received in revised form 2 December 2023; Accepted 4 December 2023

disasters in the world [1]. Historically, Bangladesh experiences several natural disasters such as floods, droughts, cyclones, tidal surges, river bank erosion, salinity intrusion and earthquake due to its complex geography and climate. Over the last thirty years, hill-cutting problems become prominent in the hilly areas due to development activities and unplanned urbanization (including unplanned migration). Besides, climate change-induced short-time extreme rainfall increased in Bangladesh. As a result, landslides frequently occur in the fragile hilly landscape, causing huge casualties and significant economic loss [2,3]. From the period of 2000–2018, 204 landslides occurred in south-eastern Bangladesh resulting in 727 casualties and 1017 injuries [4]. Changes in land cover in different forms in hilly areas, especially in the Chattogram district, resulted in severe landslides and the frequency is continuously increasing. Hill cutting, expansion of brick kilns, hill soil collection for brick kilns, agricultural activities on hill slopes, unplanned urbanization, and human migration to hilly areas are the major drivers of land cover change in the hilly areas of Bangladesh. These types of land cover changes make the fragile hilly landscapes more vulnerable to landslides [2,3]. From 2000 to 2018, landslides increased at a rate of 4 % and around 19 landslides occurred each year (Sultana 2020). Though the landslide risk is particularly evident in the city corporation areas of Chattogram due to the presence huge population and resources, other areas are also becoming vulnerable to landslides causing serious damage to the environment, rural peoples and natural resources [5,6].

Landslide susceptibility mapping is considered the first step in landslide hazard assessment. Subsequently, it helps in landslide management and disaster loss reduction in a region [6–8]. The assumption is that proper monitoring, scientific assessment and detection of landslide-prone areas is the best approach to landslide risk reduction [9]. An accurate landslide susceptible map and the relevant spatial data have a significant value in decision-making, disaster policy formulation, proper land use plan implementation at the local scale and taking essential measures for disaster risk reduction and prevention to reduce larger loss during the disaster [10].

There are several types of landslide susceptibility mapping techniques such as physically based models, qualitative, semi-quantitative and quantitative. Physically based models extract the internal process of landslides. Semi-quantitative techniques combine qualitative (expert opinion) and quantitative techniques [11]. The quantitative analysis measures the bivariate, multivariate or inherent relationship between landslide incidents and the corresponding spatial arrangement of the conditioning factors in a given landslide zone using statistical, machine and deep learning techniques [11–13]. In a quantitative method, the numerical approximation of the likelihood of landslide occurrence in a given landslide zone is measured using a landslide inventory database. The presumption is that the actual landslides (landslides in the inventory database) and the factors related to landslide occurrences are homogeneously distributed over the study area. There are many quantitative methods have been popularly used in landslide susceptibility mapping such as frequency ratio, information value [14–20], logistic regression [13,21–33], random forest [19,34–40], support vector machine [13,41–45], and regression and decision tree [15,19,23,26,46–51]. Among the quantitative methods, machine learning models had been producing more reliable and better results compared to statistical models even in the data-scarce regions [10,15,39,52,53,54]. Sometimes, bivariate models can produce similar [16] or better result compared to logistic regression model [54] and logistic regression model can over perform to machine learning model [55]. Though a number of techniques have been applied to map landslide-susceptible zones across the world, no single method is developed as a suitable one [56]. The performance of a model changes from region to region and different methods produce different results in a given study area [7,11,57–62]. To overcome this limitation, the error rate of different models are compared for a single study area and the model that produces the highest accuracy is considered the best model for the given study area. This is the best and easiest strategy to choose the optimal model for landslide hazard mapping for a study area. The hypothesis is “the best model will produce the lowest error rate and it will be considered as the best predictive model” [56,63]. For this reason, researchers compared different computing techniques [57,60,61], spatial data sources [59], inventory mapping [62], the combination of spatial data [59], computing software [64] etc. to get the best landslide susceptibility map. Over time GIS-based techniques have become more popular among the scientific community for landslide-prone area prediction [10,65]. GIS and remote sensing techniques have been popularly applied to carry out many studies in landslide susceptibility mapping research across the world [61].

The rapid development of GIS (Geographic Information Systems) and the easy integration of other technology into the GIS environment enable users to easy application of several landslide susceptibility mapping models [10,66,63]. Machine learning models can be easily integrated into GIS that can simulate landslide susceptibility zones in an accurate and scientific manner. CRAN-R software, in this case, by analysing data, enables the prediction of landslides by different machine learning models and the result can be further integrated into GIS to predict probable landslide-susceptible zones.

Though some landslide susceptibility maps have been prepared for the Chattogram district but covering the whole district is limited and only application of bivariate statistical models are found. Also, machine learning methods were applied to produce landslide susceptibility maps of Chattogram Metropolitan Areas. So, the aim of the current research is to prepare landslide susceptibility maps of Chattogram District utilizing GIS-based machine learning models. The machine learning models chosen to compare are logistic regression, random forest and decision and regression tree which have been widely used in different study areas across the world with higher accuracy. The landslide susceptibility map of the whole Chattogram district was not produced using the selected machine learning models before. The landslide susceptibility map and the spatial databases will be helpful for land use planning, identifying vulnerable areas, and sustainable hill planning in the region. The scientific community, policy-makers and stakeholders will be beneficial from this research.

2. Methodology

2.1. Study area

Fig. 1 shows the location of the study area. The Chattogram District is situated between the latitudes 21°54′ and 22°59′ North and

the longitudes 91°17′ and 92°13′ East. The district is surrounded by Feni and Khagrachhari in the west, Rangamati in the north, Cox's Bazar in the southeast and the Bay of Bengal in the south. The proximity of the Bay of Bengal supports the region with higher moisture and excessive rainfall. Even the tropical sunshine also supports the hilly areas with a continuous denudation process. The river systems in Chattgogram district also support moisture and erosion, though rainfall is the main agent of the erosion process in Chattogram.

There are many anticlines and synclines found in the Chattogram Hill Tracts which are folded into a sequence of long sub-meridional (NNW-SSE). These folded anticlines and synclines are composed of Upper Tertiary sandy-argillaceous sediments which

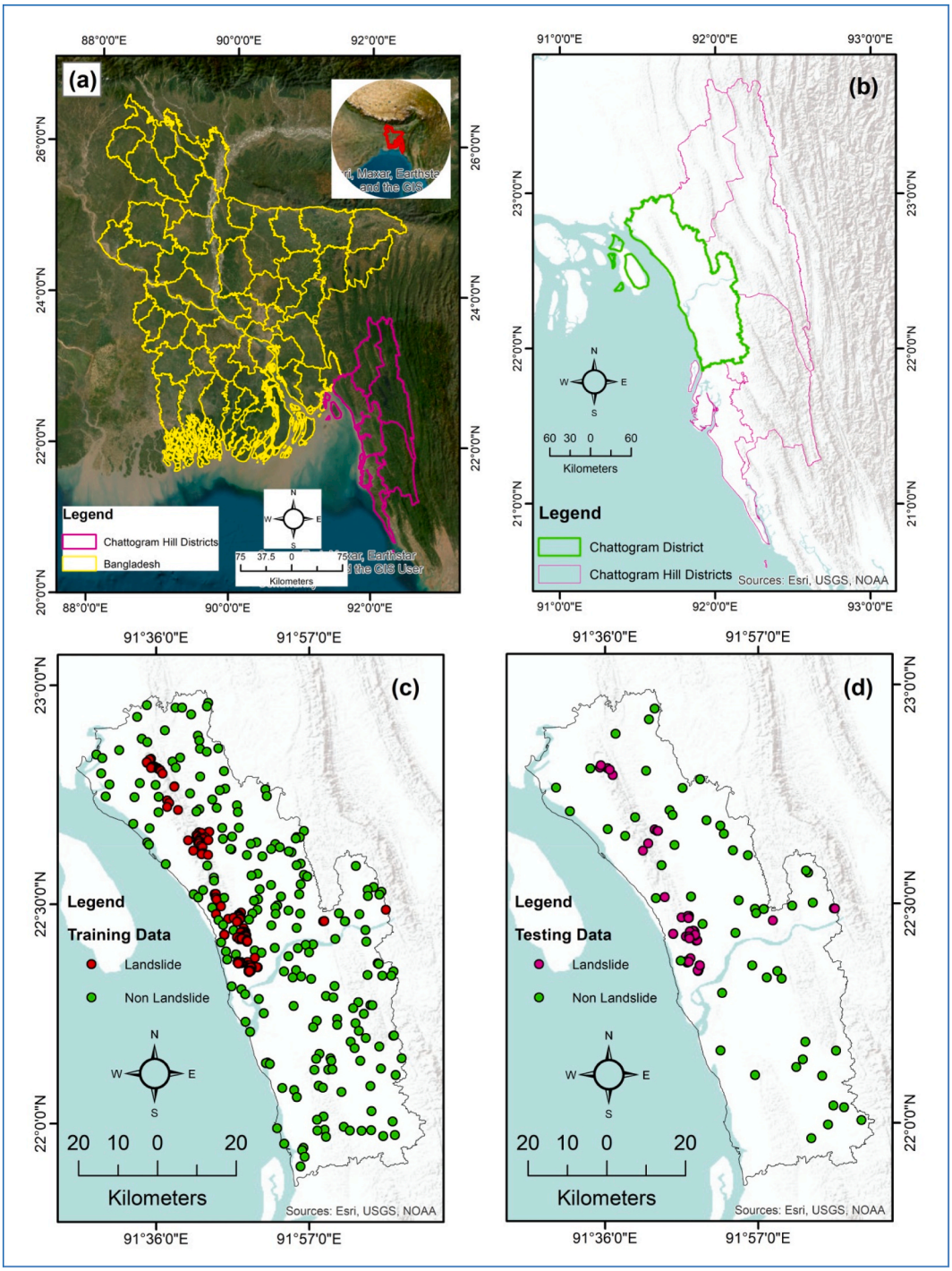


Fig. 1. Location of the study area and landslide points. a) Location of Chattogram District in Bangladesh. b) Chattogram District in Chattogram Hilly Areas. c) landslide training and d) testing data in the study area.

have been folded over time. The low elevated elongated hill ranges and the intervening valleys which are visible in the surface topography are the output of these folded anticlines and synclines. The folded structures of the topography can be demarcated by their en-echelon orientation which becomes more intricate and intense following the east direction. Consequently, from west to east, the folded flanks are divided into three North-South trending zones: (a) the Western Zone (highly compressed structures, merely basic box-like folds, contiguously associated ridge-like asymmetric anticlines; (b) the Middle Zone (steep sides and gentle crests separated by gentle synclines) and (c) the Eastern Zone (narrow anticlines are severely disrupted and have steep cutting flanks, anticlines are mostly related to thrust faults) [67]. Unconsolidated sedimentary rocks and readily weatherable feldspar are found in the surface strata of this area. In the surface strata, yellowish-brown to reddish-brown loams are found. Geologically, the underlain rocks of these hill ranges are derived from the Tipam and Surma formations. The hill ranges are only a few hundred meters high, but they feature steep slopes that are connected by gullies and rills [68–70].

2.2. Data and data sources

The landslide inventory database was prepared by conducting extensive field surveys and google earth time series image interpretation. The geological map of the study area was acquired from the Geological Survey of Bangladesh (GSB) (server link: http://

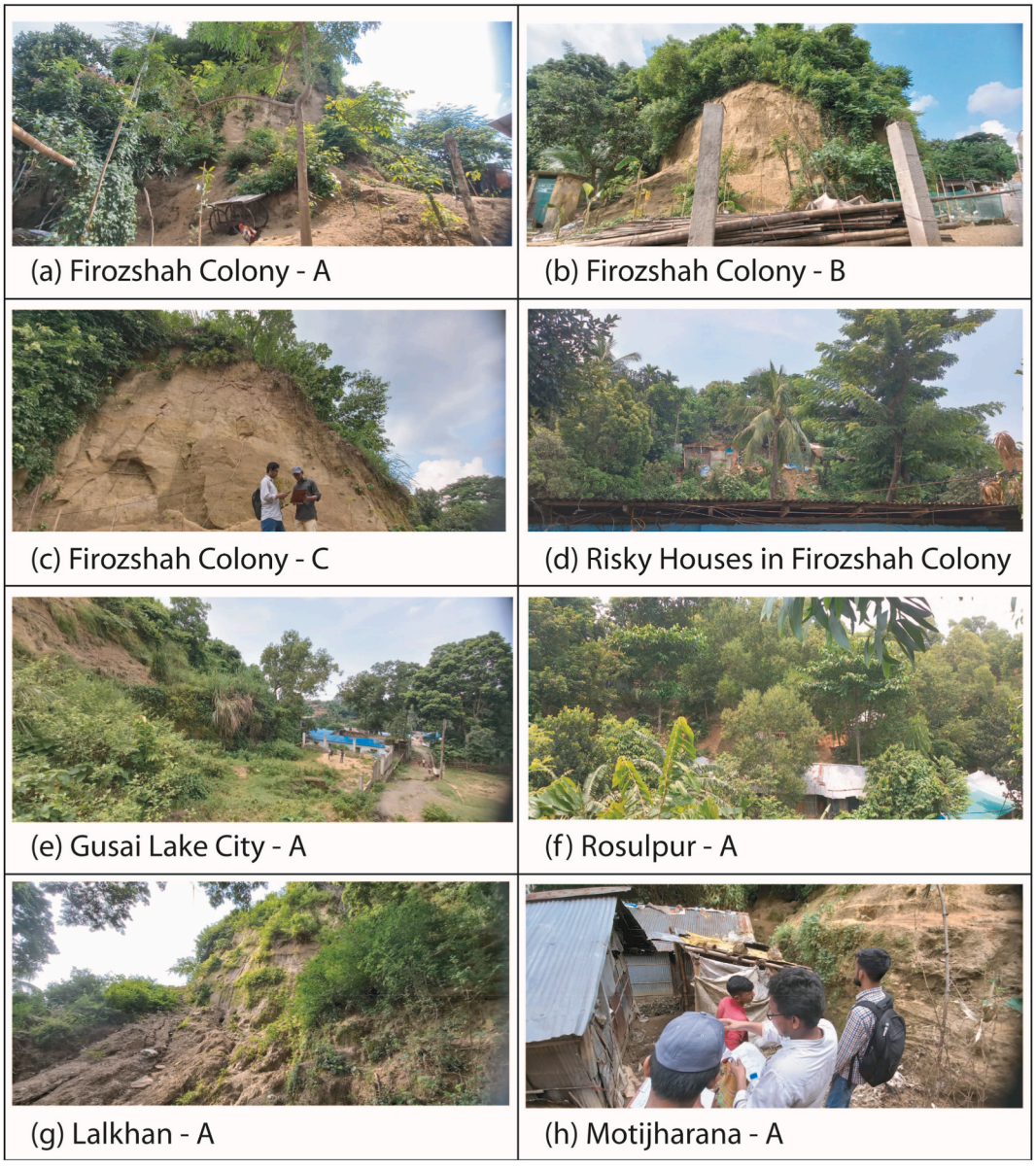


Fig. 2. Landslide site in Chattogram.

www.gsb.gov.bd/site/view/commondoc/Geo-scientifc%20Map/-?page=3&rows=20) Both Landsat satellite images and SRTM-DEM (Shuttle Radar Topography Mission Digital Elevation Model) were directly acquired from USGS geological survey website (https://earthexplorer.usgs.gov/). Data of the road network of the study area was acquired from the GEOFABRIK server (Link: https://download.geofabrik.de/).

2.3. Landslide inventory mapping

The landslide inventory database was prepared by conducting extensive field surveys during August 2022–November 2022 and by interpreting time series google earth image interpretation [71]. Records of landslides from local newspapers, existing literature, including theses, reports, and published and unpublished works, as well as government data and archives, were used to detect landslide locations. Finally, an inventory database of 255 landslide locations was prepared (Fig. 1). Among the identified landslides, the slide is the most dominant category in the study area followed by fall, flow and topples. Fig. 2(a - h) shows some photos of the field survey in different location of the study area.

2.4. Landslide conditioning factor preparation

The spatial database of landslide conditioning factors includes sixteen factors which were selected based on field experience and the review of previous literature (Table 1). Pixel size of the conditioning factor processed from SRTM-DEM and Landsat images are 30 m. The geology, distance to river, drainage density and distance to road are also converted into 30 m resolution for better comparison. Some factors are deducted after the multicollinearity test. The selected landslide conditioning factors are shown in Fig. 3(a-p).

Elevation (Fig. 3a), slope (Fig. 3c), aspect (Fig. 3b), general curvature (Fig. 3d), plain curvature (Fig. 3e) profile curvature (Fig. 3f) are derived from the arc toolbox of ArcMap 10.5 software. Geology (Fig. 3g) is used as a geological unit. Distance to road (Fig. 3n) and distance to stream (Fig. 3m) is measured by the Euclidian distance method in ArcMap 10.5 software. Drainage density (Fig. 3l) is measured using the line density tool of ArcMap 10.5 software. To measure the distance to stream and stream density, streams of the study area were extracted from DEM by morphometric analysis of the study area. STI (Fig. 3h), TRI (Fig. 3k), TWI (Fig. 3i) and TPI (Fig. 3j) are calculated using equations (1)–(4).

$$TWI = \ln\left(\frac{\alpha}{tan\,\beta}\right) \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$TRI = 1/\cos(\tan\beta * \pi\ /\ 180) \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$$STI = (m+1) \times (A_s/22.13)m \times \sin(\beta\ /\ 0.0896)n \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

$$TPI = \text{zero/near} - \text{zero} - \text{flat or a near continuous slope} \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Using equation (5), the NDVI (Normalized Difference Vegetation Index) (Fig. 3p) of the study area was calculated from Landsat 8 OLI image.

$$NDVI = (\text{NIR} - \text{RED})\ /\ (\text{NIR} + \text{RED}) \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

Table 1
Landslide conditioning factors selected for the research used in literatures in Bangladesh.

Conditioning Factor	Influence of the factor
Elevation	It affects the geological and geomoohological process and landslides tends to occur in high elevated areas.
Aspect	Slope direction indirectly affects the landslides by controlling vegetation growth, species distribution, sunlight etc.
Slope	Significantly affects the slope stability.
General Curvature Plan Curvature Profile Curvature	Generally concave slope retains more water than convex and flat slope resulted in instable slope condition
Geology	Different geological units have different impact on landslides.
Distance to river	Areas close to rivers are more prone to erosion as a result likelihood of landslides is higher.
Drainage Density	Higher stream density also affects landslides positively.
STI	STI indicates the power of overland flow to cause erosion. High STI values mean higher chance of erosion by overland flow.
TRI	It indicates the surface ratio between concave and convex upward slopes.
TWI	TWI indicates the impact of spatial scale on hydrological process. High TWI values indicates high infiltration rate of water as result areas with high TWI values are more prone to landslides.
TPI	It is the measure of topographic slope position which is the output of difference of a target cell and its surrounding cells.
Distance to Road	Road construction increases water infiltration and causes slope break resulted in higher landslides in close proximity to road.
LULC	Erosion and weathering process are significantly increased by anthropogenic activities such as forest removal, agricultural practice, and house construction.
NDVI	Low NDVI indicates low presence of vegetation. Theoretically low vegetated areas are more prone to landslides.

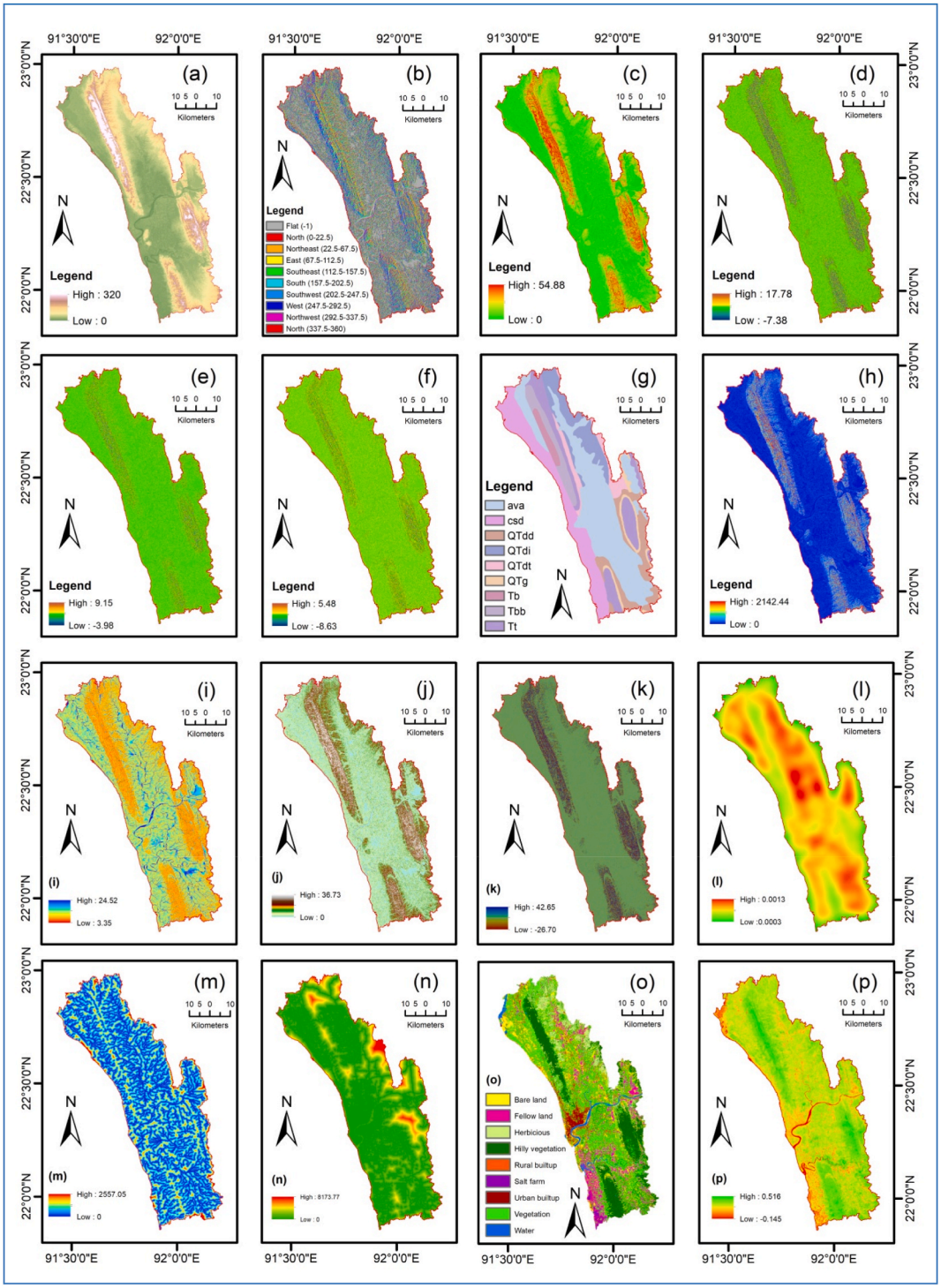


Fig. 3. Landslide conditioning factors: a = elevation, b = aspect, c = slope, d = general curvature, e = plan curvature, f = profile curvature, g = geology, h = STI, i = TWI, j = TPI, k = TRI, l = stream density, m = distance to stream, n = distance to road, o = LULC and p = NDVI).

2.5. Land use and land cover map

Land use and land cover map of the study area were classified by the machine learning classifier of Random Forest. Machine learning classifiers are more effective in demarcating different types of land use and land cover from satellite images [72,73]. Random

Forest classifier performs better and produces high overall accuracy compared to SVM and ANN in producing land cover classification maps [37,38,74]. Land use and land cover map of the study area were classified using the Random Forest classifier tool of QGIS 3.25 software. The landsat image of November 19, 2021 was acquired for LULC mapping. The path and row of Landsat image is 136 and 45. Nine land cover classes were identified and they are water body, urban built-up, rural built-up, bare land, fellow land, salt farm, mixed large trees, hilly vegetation and herbaceous vegetation (Fig. 3o). Descriptions of the selected land cover classes are given in Table 2.

2.6. Accuracy assessment of LULC map

Accuracy assessment of land use and land cover classification map from remote sensing images is very essential. The overall accuracy, producer accuracy, user accuracy, and kappa coefficient of the classified map are computed following Chowdhury and Hafsa (2022b) [75]. Table 3 shows the accuracy of the classified map measured using the input sample data of each land cover classes. The overall accuracy of the map is 0.99 and the Kappa is 0.99 indicating a higher accuracy of the land cover classification map.

2.7. Non-landslide location sampling

All the locations free of landslide locations are considered non-landslide locations. The "Random Sampling" tool of ArcMap 10.5 software was used to create a non-landslide location database. To remove the bias, the same number (255) of non-landslide locations was determined.

2.8. Training and testing data extraction

The inventory data was divided into training and testing using the "subset feature" tool of ArcMap 10.5 software [76]. From the dataset, 80 % (204 landslide locations) were used for training purposes and 20 % (51 landslide locations) were used for testing purposes [71,77,78]. The same method was also applied to non-landslide data. From the dataset, 80 % (204 non-landslide locations) were used for training purposes and 20 % (51 non-landslide locations) were used for testing purposes. Both landslide and non-landslide locations selected for training and testing purposes are merged together for the final application in the mathematical model.

2.9. Multicollinearity test

The presence of a strong correlation between dependent and independent variables in the dataset leads to multicollinearity. The presence of the collinearity effect leads to inaccurate prediction of the independent variable (overestimation or underestimation). So that before applying any multiple regression or machine learning models, testing multicollinearity is essential as these models are highly sensitive to multicollinearity effects. Especially in landslide susceptibility mapping, testing multicollinearity between landslides and landslide conditioning factors is mandatory [6,18,30,38,79]. Many techniques are used to quantify multicollinearity, but variance inflation factors (VIFs) are the most popular and frequently employed [30,79–82]. In this study, multicollinearity was evaluated using two statistical indexes of tolerance (TOL) (equation (6)) and VIFs (equation (7)).

$$Tolerance = 1 - C^2 \quad (6)$$

where C^2 represents the coefficient of determination for the regression of the explanatory variable on the remaining independent variables. VIF is demonstrated as the correspondence of tolerance and is calculated as follows:

$$VIF = \frac{1}{Tolerance} \quad (7)$$

2.10. Logistic regression

The logistic regression model is a very popular method extensively used in landslide susceptibility mapping research. It is successfully used in a number of case areas across the world [13,14,23,26,29,33,83,84]. The logistic regression model predicts the

Table 2
Description of the land cover types.

SN	Land cover type	Description
1	Water body	River, pond, lake
2	Urban built up	Densely built buildings in the urban area and related construction
3	Rural built up	Built up found outside city, sparsely built houses.
4	Bare land	Completely bare surface, sand cover
5	Fellow land	Land to be used for cultivation but seems partly bare
6	Salt farm	Salt farm along the coast
7	Mixed large trees	Large trees located in the rural or urban area in the low elevated land
8	Hilly vegetation	Dense vegetation over the hillslope
9	Herbaceous vegetation	Sparse vegetation over medium to low hills, especially the tea garden

Table 3
User and producer accuracy of individual land cover class.

	1	2	3	4	5	6	7	8	9
PA [%]	100	100	100	100	99.99	100	99.99	100	100
UA [%]	100	100	100	99.99	100	100	100	99.99	100
Kappa	1	1	1	0.99	1	1	1	0.99	1

dependent variable by using a group of independent variables. The dependent variable is always binary or dichotomous (0 and 1) [16]. Details of the logistic regression model is found in Chau and Chan (2005) [85].

The logistic regression model is considered a benchmark model and is frequently used to assess how the other created models function. This multivariate statistical technique is typically applied to solve binary classification (0, 1) problems. To predict binary dependent variables it calculates the weights of each of the dependent variables which is one of the major advantages of the model [32]. Besides, the model can handle non-normally distributed data and can use both continuous and discrete variables during calculation [15,16,18,30,32,82]. The logistic regression model was developed by the "stats" library in CRAN-R software. The logistic regression function can be expressed by a quantitative relationship as follows (equation (8)) [16,18,85]:

$$P = 1 / (1 + exp^{-z}) \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

where P: is the probability of landslide occurrence that estimated values vary from 0 to 1. Variable Z is landslide causal factors and assumed as a linear combination of the causal factors X_i (i = 1,2, …n) as (equation (9)):

$$Z = e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\ldots\ldots\ldots\ldots\beta_i x_i)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

where β_i is the optimal regression coefficient reflecting the contributions of each factor; β_0 is the constant coefficient; x_i is the input conditioning factors; and P is the occurrence probability of coseismic landslides.

The coefficients are calculated based on the inherent relationship between the dependent and independent variables where landslides are the dependent variables and the conditioning factors independent variables. The corresponding values of the independent variables are used to ascertain the probability of an independent variable (landslides). Z is an index that allows us to combine the independent variables affecting the dependent variables. The value of Z varies between - ∞ to + ∞ [30].

2.11. Random forest

Breiman (2001) [86] first developed the Random Forest by integrating the bagging sampling approach of Breiman (1996) [87], random split selection [88] and the random feature selection process of Ho (1995, 1998) [89,90] and Amit and Geman [91] (1997) [39,92]. Random Forest is a tree-based ensemble classification method [86,93–95] which predicts dependent variables using the independent variables [94]. Random Forest classifier creates many decision trees for a dataset [94] using different subsets of the data [39]. A collection of random variables were split from the main dataset on which the tree of the Random Forest depends [94]. The RF is a well-known method found in many literatures and it has high performance in landslide susceptibility mapping. It has a number of benefits, including (1) Its non-parametric nature base; (2) Its ability to assess the significance of variables used; (3) the algorithm can estimate the missing values in the dataset; and (4) It has the capability of regression, classification, and unsupervised learning [84]. Each tree in the RF is built using a subset of the predictor variables. The number of trees (ntree) and predictors (mtry) used to form each tree might vary depending on the dataset. In the RF model, the robust error is estimated with the testing dataset. For this purpose, each tree is built from the bootstrap sample of the training dataset as follows (equation (10)):

$$MSE = n^{-1} \sum_{i=1}^{n} (t_i - \bar{t}_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

where t_i is the average of all out-of-bag predictions, n is the number of out-of-bag observations in each tree, and MSE is the mean square error obtained during the construction of the classification trees. The explained variable's percentage is determined as follows (equation (11)):

$$V_{ex} = 1 - \frac{MSE}{V_z} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

where Vz denotes the response variable's overall variation. Finally, the RF produces a single prediction that represents the average of all aggregations. In this research, the "randomforest" library of CRAN-R software was used to build the Random Forest model.

2.12. Decision and regression tree

Decision and Regression Tree is a non-parametric supervised machine learning algorithm. It is prominently used for modelling large data sets using classification and regression methods. The dependent variable (target variable) is then predicted by using the independent variables (other data features). The model creates some simple decision rules using the independent variables to predict

the target variable [26]. The underlying idea behind a decision tree is to split the provided training dataset into multiple subsets. Each subset (split) consists of sets of more homogeneous or less homogeneous states of the dependent (target) variable. The decision tree model evaluates all the input attributes (independent variables) to calculate the impact of each independent variable on the dependent variable at each split in the tree. This process is recursively performed in the model and a decision tree is formed as an end result [46]. There are many types of algorithms available [15,47]. In this research "rpart" library of CRAN-R software was used to build the Decision and Regression Tree model.

2.13. Validation of the map

The receiver operating characteristics (ROC) graph is extensively used in landslide map validation [6,58,96,97]. The ROC curve consists of both x and y-axis in a diagonal plotting area. In the x-axis false positive rate is plotted and in the y-axis true positive rate is plotted. The x-axis displays 1-specificity and the y-axis displays sensitivity. The specificity and sensitivity are measured using equations (12) and (13). In the current research, the ROC curves success rate is measured using the training dataset and the ROC curves prediction rate is measured using the training dataset [6,58,97,98].

$$X = 1 - specificity = 1 - \left(\frac{TN}{TN + FP}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

$$Y = 1 - sensitivity = 1 - \left(\frac{TP}{TP + FN}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (13)$$

The result of the success rate is the comparison between the training data and the landslide susceptibility maps. Additionally, using the validation data, prediction rates were determined [58,97]. The success rate reflects how well the LSM divides the landslides among the susceptibility zones, or how well the model fits the data [21,98,99]. However, the prediction rate cannot accurately determine the likelihood of future landslides; therefore, the prediction rate is calculated [98].

3. Results and discussions

Sixteen landslide conditioning factors were selected after the collinearity effect test to measure the relationship between landslides and the spatial arrangement in the study area. After that, the models were applied to measure the weight of the conditioning factors to landslide occurrences. To prepare the final map, the weights of the conditioning factors were engaged using a simple overlay method of ArcMap 10.5 software. To engage the weights of the conditioning factors either the specific equation was followed such as for the logistic regression model or conditional maps were combined such as for the decision and regression tree model.

3.1. Testing multicollinearity

Among many other methods, variance inflation factors (VIFs) are widely accepted to check the collinearity effects. Multicollinearity is tested using VIFs and tolerance (TOL) and the result is shown in Table 4. The larger the VIF value and the lower the TOL value the higher the collinearity effect. The VIF value must be less than 10 and TOL must be more than 0.1 for the dataset to be acceptable. In Tables 4 and VIF values of all the conditioning factors are less than 10 and TOL values are greater than 0.1 indicating the selected conditioning factors are free of collinearity effects and any kind of bias. The result of the final analysis will be free of bias and

Table 4
Coefficients of multicollinearity test.

Unstandardized Coefficients			Standardized Coefficients			Collinearity Statistics	
			Beta	t	Sig.	Tolerance	VIF
(Constant)	0.122	0.153		0.798	0.426		
DEM	0.161	0.111	0.061	1.451	0.148	0.586	1.706
Aspect	0.710	0.237	0.111	2.997	0.003	0.749	1.336
Slope	0.199	0.244	0.056	0.817	0.414	0.222	4.514
Profile curvature	0.061	0.130	0.021	0.467	0.641	0.497	2.012
Plan curvature	−0.023	0.133	−0.009	−0.171	0.865	0.411	2.431
General curvature	−0.057	0.156	−0.016	−0.362	0.718	0.555	1.802
Geology	0.988	0.170	0.241	5.805	0.000	0.598	1.672
TWI	0.069	0.062	0.054	1.116	0.265	0.436	2.294
TPI	−1.501	0.479	−0.116	−3.133	0.002	0.754	1.326
TRI	0.308	0.238	0.091	1.293	0.197	0.207	4.824
Distance to stream	0.066	0.066	0.034	0.994	0.321	0.863	1.159
Stream density	0.355	0.074	0.188	4.800	0.000	0.672	1.488
STI	−0.332	0.192	−0.064	−1.729	0.085	0.761	1.315
LULC	0.292	0.073	0.161	4.028	0.000	0.642	1.557
Road	0.431	0.080	0.187	5.370	0.000	0.845	1.183
NDVI	0.079	0.195	0.015	0.407	0.684	0.722	1.384

there will be no collinearity effects of conditioning factors over landslide occurrence. TRI and slope showed the highest collinearity and the values are 4.824 and 4.514, respectively (Table 4). The lowest value of tolerance was also found for these two factors and they are 0.207 and 0.222, respectively. Distance to stream and distance to road showed the lowest VIF of 1.159 and 1.326 respectively (Table 4).

3.2. Landslide susceptibility maps

3.2.1. Logistic regression

The confusion matrix of the LR model indicates that 171 landslide absences and 189 landslide presences were correctly predicted (Table 5). The model incorrectly predicts 15 landslide absences and 33 landslide presences. The percentage of correctly predicted absences is 83.82 % and the presence is 92.65 % (Table 5).

The coefficient of the logistic regression model indicates the impact of the predictors on response variables. From Table 6, the importance and significance of the variables are found. The result showed that aspect has the largest impact on landslide occurrence with a coefficient value of 7.6384, followed by geology (7.2108), distance to road (5.1324), Stream density (3.5885), LULC (2.9591) and slope (2.0746). The negative value indicates a negative impact on the variables. But in the case of landslide conditioning factors, the negative value indicates that the factor is inversely affecting landslide occurrence in an area. A negative coefficient was found for only two conditioning factors, STI (−1.5394) and TPI (−11.0749). The final map is prepared by combining the weights of the factors using ArcMap software (Fig. 6a).

3.2.2. Random forest model

In the random forest model, data were split into training (80 %) and testing (20 %). The prediction of the model is shown in Fig. 4 and Table 6. Fig. 4 shows the aggregate OOB predictions indicating the error rate of the model. It is understood from Fig. 4 that the resulting model will produce a 25 % error rate for new observations. So, for a reasonably good model, 75 % of the results will be accurate. The rows of Table 6 correspond to the actual observations of the input data and the columns demonstrate the predictions for observation [34].

The RF model predicts 1 for 8 observations and predicts 0 for 19 observations. The findings of the model showed that 148 landslide absences are predicted correctly and the percentage is 89 %. 167 landslide's presence is predicted correctly and the percentage is 96.05 % (Table 7).

Table 8 shows the importance of the selected conditioning factors derived from the random forest model and ordered according to the mean decrease Gini. According to Table 8, the high the value of a conditioning factor the higher the impact of that conditioning factor on landslide occurrence. The random forest model demonstrated that geology is the most important variable (42.153 %), followed by TRI (20.143 %), elevation (19.036 %), stream density (15.670 %) and Slope (13.212 %) (Table 8). The final map is prepared by combining the weights of the factors using ArcMap software (Fig. 6b).

3.2.3. Decision and Regression Tree Model

The R programming language and the rpart package were used to construct the regression trees in the current study. Typically, this leads to a complicated decision tree model. Such a tree structure needs to be "pruned" in order to extract only the most important information (i.e., the nodes that account for the majority of deviation) required for further analysis. The confusion matrix of the DRT model (Table 9) indicates that 194 landslide presences were correctly predicted which is 95.10 % and 181 landslide absences were correctly predicted which is 88.72 %. The model incorrectly predicts 10 landslide absences and 23 landslide presences.

Fig. 5 shows the tree structure of the Decision and Regression Tree Models. The importance of the nodes and the percentage corresponding to landslides are also shown there. According to Fig. 5, TRI gets the highest importance and it is the root node.

The value for landslide occurrence for a single condition is found in the terminal node of each leaf. Such as, If TRI $<$1.76045 and Elevation $<$28.5 and Geology $<$0.2132355, then the likelihood of landslide occurrence will be 0.0132 and this condition reflects 37 % of the landslide events. Root and terminal nodes and their corresponding importance derived from CRAN-R software are shown in Fig. 5. From Fig. 5, landslide susceptibility map is prepared by following the below conditions using AcrMap 10.5 software (Fig. 6c):

If TRI$<$ 1.76045 and Elevation$<$ 28.5 and Geology$<$ 0.2132355, then assign value 0.0132; If TRI$<$ 1.76045 and Elevation$<$ 28.5 and Geology $\geq$ 0.2132355, then assign value 0.636; If TRI$<$ 1.76045 and Elevation $\geq$ 28.5 and Distance to road $\geq$ 63.541, then assign value 0.308 etc.

3.3. Landslide susceptibility map evaluation

Landslide susceptibility maps were produced for all three models LR, RF and DRT shown in Fig. 6. These models produced pixel

Table 5
Confusion matrix of LR model.

Number		Predicted		recall	precision	accuracy
		0	1	0.838	0.919	0.882
Actual	0	171	33			
	1	15	189			

Table 6
Coefficients of logistic regression model.

Coefficients:	Estimate	Std.	Error	z value	Pr(>|z|)
(Intercept)	−6.03510	1.85770	−3.24900	0.00116	**
DEM	1.53280	1.13930	1.34500	0.17847	
Aspect	7.63840	2.50370	3.05100	0.00228	**
Slope	2.07460	2.63950	0.78600	0.43186	
Profile curvature	0.27480	1.45260	0.18900	0.84998	
Plan curvature	0.24030	1.38940	0.17300	0.86266	
General curvature	1.03450	1.63360	0.63300	0.52655	
Geology	7.21080	1.63960	4.39800	0.00001	***
TWI	0.98430	0.64800	1.51900	0.12878	
TPI	−11.07490	4.79200	−2.31100	0.02083	*
TRI	0.67300	2.70940	0.24800	0.80383	
Distance to stream	0.25350	0.67500	0.37500	0.70729	
Stream density	3.58850	0.77920	4.60500	0.00000	***
STI	−1.53940	1.75910	−0.87500	0.38153	
LULC	2.95910	0.75110	3.94000	0.00008	***
Road	5.13240	1.02470	5.00900	0.00000	***
NDVI	1.22260	2.63000	0.46500	0.64201	
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

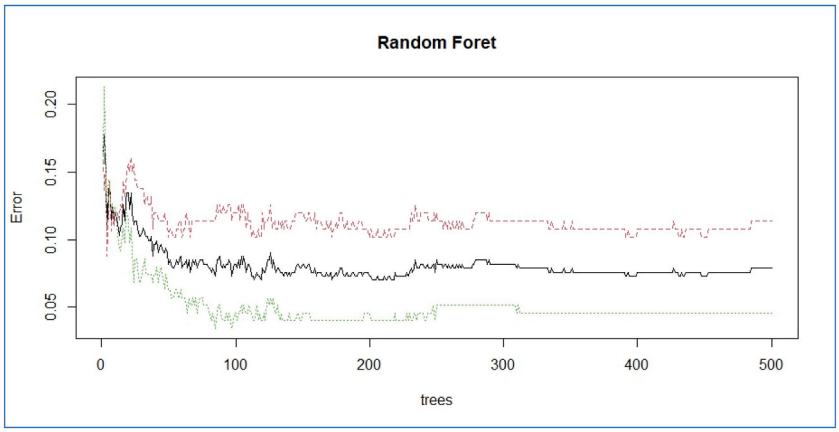


Fig. 4. The error rate of the overall RF model [OOB: out of bag (black line), 0: absent landslide (red line) and 1: present landslide (green line)]

Table 7
Confusion matrix of RF model.

Number		Predicted		recall	precision	accuracy
		0	1	0.919	0.944	0.924
Actual	0	148	19			
	1	8	167			

Table 8
Importance of the factors using training and testing data.

Factors	MeanDecreaseGini	Factors	MeanDecreaseGini
Geology	42.153	TPI	7.314
TRI	20.143	STI	5.677
Elevation	19.036	Plan curvature	5.231
Stream density	15.67	NDVI	4.435
Slope	13.212	Distance to stream	4.231
LULC	9.384	TWI	3.892
Road	7.827	General curvature	2.037
Aspect	7.516	Profile curvature	1.976

Table 9
Confusion matrix of DRT model.

Number		Predicted		recall	precision	accuracy
		0	1	0.887	0.948	0.919
Actual	0	181	23			
	1	10	194			

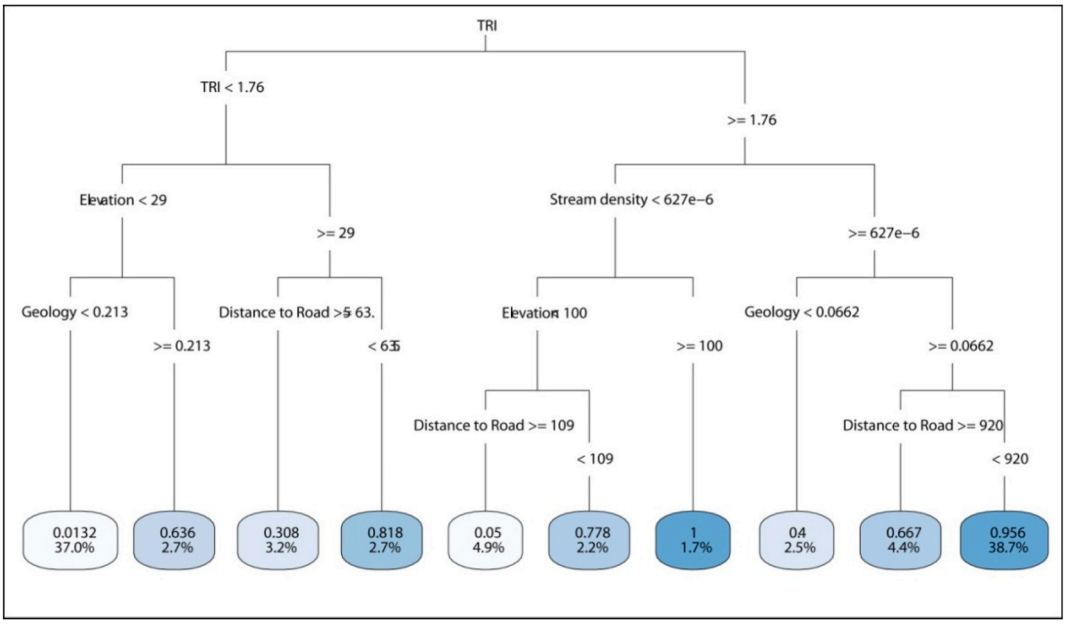


Fig. 5. Tree structure in DRT model.

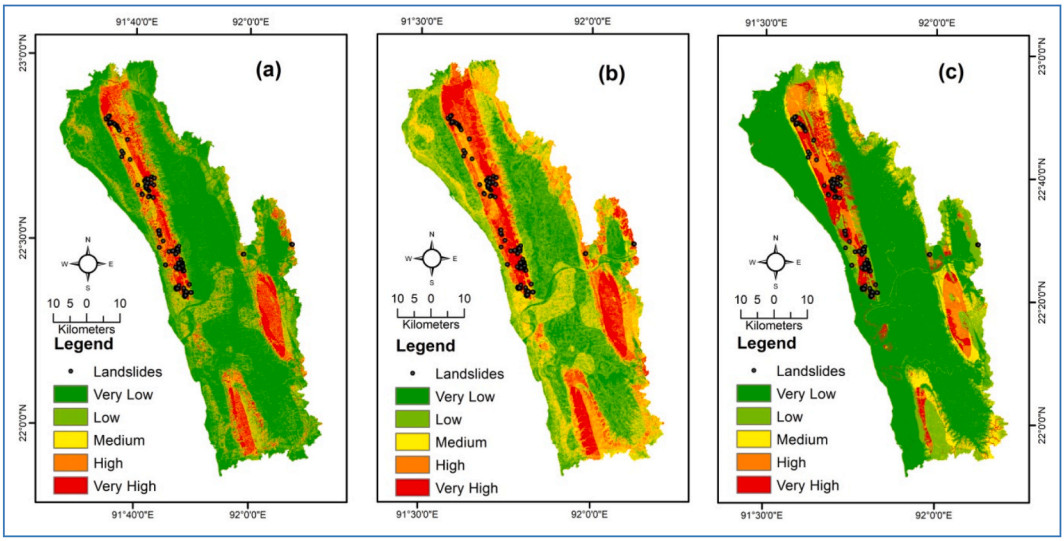


Fig. 6. Landslide susceptibility maps prepared by a) LR, b) RF, c) DRT models.

values, which were subsequently categorized using the natural break classification technique. The produced landslide susceptibility maps using the three models met two spatial effective rules: (1) the high-susceptibility class should only include tiny areas, and (2) the current landslide pixels should belong to that class (Bui et al., 2012; Guo et al., 2021). Fig. 6 indicates that the high-elevated areas are prone to landslides and the low-elevated areas which are actually river basins and coastal plain (flat areas with low elevation) are free of landslides. It can be compared from Fig. 3 a (elevation of the study area) that there are three distinct hilly zones in the study area and

from Fig. 1 it is seen that landslides are distributed in these hills. This type of distribution of landslides makes the hilly areas vulnerable to landslides which has been reflected in the susceptibility map.

Fig. 7 shows the percentage of area covered by each susceptible zone of each model. Results exhibited that according to the LR and DRT model above 60 % area is covered by a very low susceptible zone which is a very large portion of the total studied area. Both models also have a very low portion of the area covered by very high susceptible zones (LR = 9 % and DRT = 8 %). Very low to medium susceptible zones of the RF model is homogenously classified (Very low = 23 %, Low = 28 % and High = 20) but the very high susceptible zone covers a relatively small portion (12 %) of the total area. According to the hypothesis, LR and DRT will be the best model for landslide susceptibility mapping over the RF model. Because these two models have demarcated a very small area of high susceptible zone but have high accuracy. Theoretically, DRT is the best model according to success rate and LR is the best model according to prediction rate.

3.4. Rationality of the landslide susceptibility maps

The rationality of the LSM was evaluated by comparing the landslide density in each of the susceptibility classes [100] following the assumption of Guo et al. (2021) [101]. All the models showed a similar trend in landslide distribution in susceptibility classes, with an increased number of landslides in the high susceptible zones (Table 10) and low areas in high susceptible areas (Fig. 7). The highest number of landslides was distributed in the very high susceptible zone of the DRT model (207 landslides) followed by RF (182 landslides) and LR (181 landslides). There are no landslides in the very low and low susceptible class of the RF model and the very low susceptible class of LR models. In contrast, the very low and low susceptible classless of the DRT model has 10 and 17 landslides, respectively. Considering the very high and high susceptibility classes, the highest percentage of landslides was distributed in the RF model (93.73 %) followed by LR (88.24 %) and DRT (86.67 %). In addition, considering the three highest susceptible classes (medium to very high), the highest percentage of landslides was distributed in the RF model (100 %) followed by LR (95.29 %) and DRT (89.41 %) (Table 10). The assumption of landslide density infers RF as the best model followed by LR and DRT. Due to the highest number of misclassifications of landslides into very low and low susceptible classes DRT model is determined as the lowest performing model though in terms of susceptible areas, this model performed better than the other models [101,102].

3.5. Accuracy of the Map

The success and prediction rate of the model for landslide occurrence and non-occurrence was determined using the area under the ROC curve (AUC of ROC) [34,103]. The cut value of AUC is 0.5 indicating a model is fit, and rising AUC values improve model quality. Values under 0.5, however, reflect a random fit. In landslide susceptibility mapping research, model validation is commonly performed by both success rate and prediction rate curves. The method compares the current landslide and non-landslide locations with the prepared landslide susceptibility map to detect the acceptability of the map [24,34,104]. The training landslide pixels were employed to prepare the success rate curve and the testing landslide pixels were employed to prepare the prediction rate curve. The success rate can assist in evaluating how accurately the locations of existing landslides have been characterized in the generated landslide susceptibility maps. On the other hand, prediction rate describes the power of landslide susceptibility maps to predict future landslides in the study area. The success rate shows the area under the ROC for LR, RF and DRT models are 0.943, 0.917 and 0.947, respectively (Fig. 8a). The prediction rate shows the area under the ROC for LR, RF and DRT models are 0.963, 0.934 and 0.890, respectively (Fig. 8b). The result is above 0.7 indicating an excellent performance of the model [105]. Though machine learning models produce better results than other models [10,15,39,52,53,66], in this research, the LR model showed a better prediction rate compared to other models machine learning models [55]. DRT model misclassified some landslides and the RF model classified comparatively larger extent of very high and high susceptibility classes which ultimately affected AUC values leading to higher accuracy in the LR model.

3.6. Comparison of LSM with previous map

The findings of this study suggest that for large-scale landslide susceptibility mapping, the LR, RF, and DRT approaches provide satisfactory results. The methods showed excellent performance in both success and prediction rates according to the AUC classification scale [105]. There is difficulty in comparing results with previous research because most of the previous research conducted on the Chattogram Metropolitan area [68,106,107] and Chattogram City Corporation Area [5] except for two conducted on whole Chattogram district [6,108]. The size of these study areas is small as the size of the Chattogram Metropolitan area varies from 680 to 720 square kilometres in the papers and the size of the Chattogram City Corporation Area is 170 square kilometres. The highest and lowest success rate obtained in previous research are 0.951 and 0.839, respectively by Ahmed et al. (2015a) [68], 99.47 % and 93.35 %, respectively by Ahmed et al. (2018) [106], 98.0 and 96.0, respectively by Rahaman et al. (2017) [107], 0.939 and 0.873, respectively by Ahmed and Dewan (2017) [5]. Rahaman et al. (2017) obtained the highest prediction rate of 99 % and the lowest of 98 %. In the above-mentioned research, knowledge-driven or bivariate methods were used except for Rahaman et al. (2017) [107], which used the machine learning model of support vector machine. Rahaman et al. (2017) [107] obtained a success rate of 99.47 % and a prediction rate of 93.10 %. The models used performed excellently for a small study area in the Chattogram district. When the whole Chattogram district was considered, the size of which is 6116.13 square kilometres, the bivariate and machine learning models produced less accuracy compared to the accuracy of the small area. Mourin et al. (2018) [108] obtained 88.7 % accuracy using Support Vector Machine for the Chattogram district. Chowdhury and Hafsa (2022a) [6] obtained the highest success rate of 80.43 % and the lowest success rate of 70.11 % for the same area using bivariate statistical models. Some research [17,24,52,109,110] concluded that

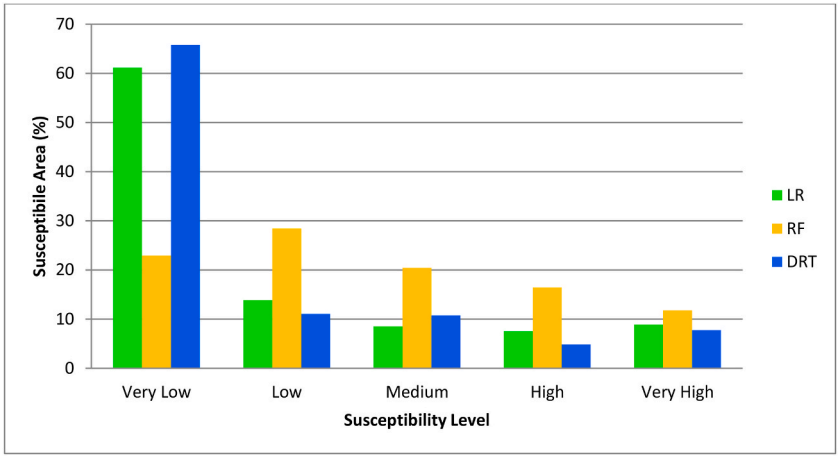


Fig. 7. Susceptible areas of landslide susceptibility maps.

Table 10
Model result comparison according to landslide density.

Models	Susceptibility Level	Area (%)	Landslides	Landslides (%)	Frequency Ratio
LR	Very Low	61	0	0.00	0.00
	Low	14	12	4.71	0.34
	Medium	9	18	7.06	0.83
	High	8	44	17.25	2.29
	Very High	9	181	70.98	7.98
RF	Very Low	23	0	0.00	0.00
	Low	28	0	0.00	0.00
	Medium	20	16	6.27	0.31
	High	16	57	22.35	1.36
	Very High	12	182	71.37	6.06
DRT	Very Low	66	10	3.92	0.06
	Low	11	17	6.67	0.60
	Medium	11	7	2.75	0.25
	High	5	14	5.49	1.13
	Very High	8	207	81.18	10.45

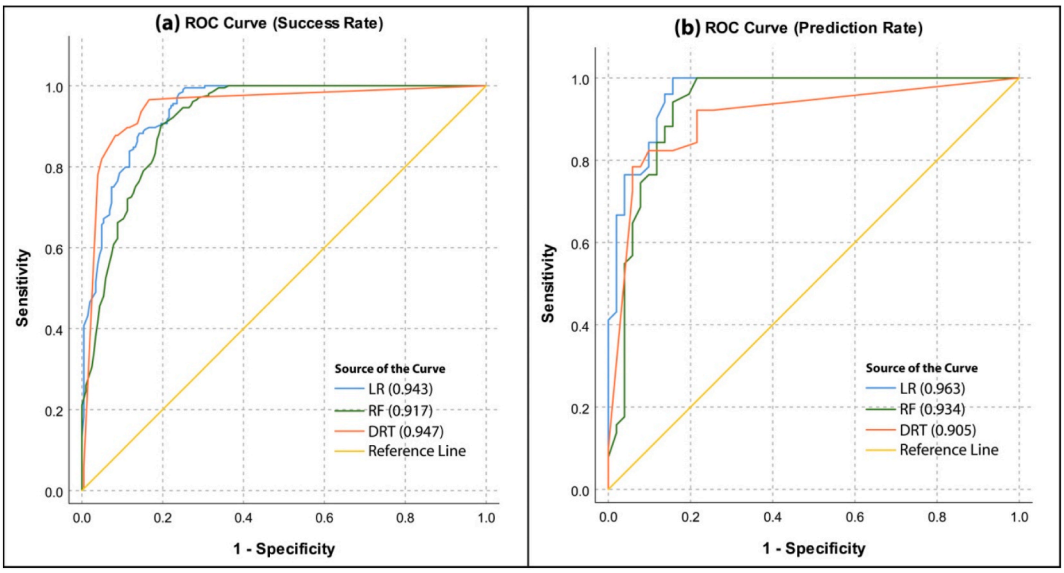


Fig. 8. Area under ROC curve a) success rate (training data) and b) prediction rate (testing data).

bivariate models are better for regional scale mapping but the current research indicates the capability of machine learning is higher than bivariate models for regional scale mapping. Additionally, it is apparent that the size of the landslide inventory database significantly affects the models' accuracy, especially for machine learning models. Mourin et al. (2018) [108] used 105 samples where 73 were used for training purposes and 32 were used for testing purposes. The training dataset of the current research was 204 and the accuracy is above 0.90 for all models.

The landslide susceptibility maps of the current research somewhat resemble the maps of the previous research. Such as maps produced by LR and FR models resemble the maps of bivariate statistical models (Chowdhury and Hafsa 2022a) [6] and the map produced by DRT resembles the map produced by the SVM method [108]. In the Chattogram district, the river basin areas or low elevated areas and the coastal areas are free of landslide risk but the high altitude areas are prone to landslides. Some hills located in the southeastern part of the study area never encountered landslides according to previous records. But all methods in all the research classified them as landslide-prone zones. This is the limitation of the mathematical models; they underestimate or overestimate landslides in the un-sampled areas.

In landslide susceptibility mapping, results always vary from model to model both in accuracy and in the importance of conditioning factors even for the different variants of the same models [15,19,26,108]. The variation in the importance of the variables from model to model is also evident in this research. The performance of the models is also satisfactory and higher compared to the results of other research [15,19,34]. The variation in the success rate and prediction rate of the models also resembled that of Youssef et al. (2016) [34]. Wang et al. (2021) [60] used different datasets and found LR outperformed over RF model. Youssef & Pourghasemi (2021) [61] found a 95.1 % success rate for the RF model in Asir Region, Saudi Arabia. Current research showed the overall highest success and prediction rate of the selected models compared to other regions [34]. Gui et al. (2023) [66] found best-first decision tree (BFT) outperformed against rotation forest and other ensemble models. But in this research, the success rate of DRT and prediction rate of LR is the highest compared to the other models. Characteristics of the study area affect the result of landslide susceptibility mapping models and in this study area, the LR model showed a better prediction rate than machine learning models which supports the result of Luo et al. (2019) [55] where the LR model performed better than Naive Bayes (NB) model.

The current map covers the whole Chattogram district with the highest success rate of 0.947 and lowest success rate of 0.916 and highest prediction rate of 0.963 and the lowest prediction rate of 0.905. Besides, three new machine learning algorithm has been used to prepare landslide susceptibility maps in this district. According to the extent of the study area, this research has produced the most accurate and reliable landslide susceptibility map for the Chattogram district.

3.7. Cause of landslide in the study area

According to the LR model, aspect, geology, road, stream density, LULC, slope and elevation are the most influential factors in landslide susceptibility mapping. According to the RF model, geology, elevation, TRI, stream density, slope and LULC have got the highest importance in landslide occurrence. According to the DRT model, TRI, elevation and geology account for 37 % of landslide occurrence and TRI, stream density, geology and distance to road account for 38.7 % of landslides in the study area. In all cases, geology gets the highest importance in landslide occurrence in the Chattogram district. Geology alone is the most crucial factor for landslide occurrence and some topographic and hydrologic factors are also accompanied by landslide occurrence. The combined effect of the mentioned factors is mainly responsible for the natural denudation and erosion process where human footprint in various forms makes the process faster. All the maps account for the hilly areas highly prone to landslides which were also evident during field observation. The river basin areas, the Halda basin, are prone to erosion, but due to low elevation and because of the destructive force they cannot be declared as landslides. The whole area is morphologically active and the morphologic activity is driven by a combination of multiple factors. The hilly areas are at high risk to the environment and people as the landfall, flow, slide or topple from high elevated areas can cause more damage to them.

4. Conclusion

GIS-based machine learning algorithms of logistic regression, random forest and decision and regression tree were used to prepare landslide susceptibility maps for a highly landslide-prone area, Chattogram district of Bangladesh. These three models were introduced for the first time to prepare landslide susceptibility maps for the Chattogram district as the previous study was mainly confined to Chattogram City. A landslide inventory database of 261 locations and sixteen landslide conditioning factors was used to account for the relationship between landslides and the conditioning factors. Geology alone is the most crucial factor for landslide occurrence in the Chattogram District. Besides, some topographic (elevation, slope and aspect) and hydrologic (TRI and stream density) factors also cause landslides accompanied by geology. Human interventions such as LULC and road construction have a minor impact compared to other factors. The accuracy of LR, RF and DRT models were 0.943, 0.917 and 0.947 respectively for success rate and 0.963, 0.934 and 0.905, respectively for prediction rate. According to the results, DRT produces the most realistic landslide susceptibility map for Chattogram District. According to the models, almost 9–12 % of areas of the Chattogram district are highly susceptible to landslides mainly covering the hilly areas. Chattogram City is located in very high susceptible zones. The resources and population in the entire hilly area are at high risk of landslides. Active morphological and denudation process is responsible for the high landslide potentiality in the hills. The research findings will be very supportive to the land use policy makers and landslide disaster planners and to the researchers. The machine learning models used in this research can also be replicated in other regions, but the result may affected by the study area and the number of conditioning factors used.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Md Sharafat Chowdhury: Writing - review & editing, Writing - original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Md Naimur Rahaman:** Software, Investigation, Data curation. **Md Sujon Sheikh:** Software, Investigation, Data curation. **Md Abu Sayeid:** Investigation, Data curation. **Khandakar Hasan Mahmud:** Supervision. **Bibi Hafsa:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was funded by Information and Communication Technology Division, Government of Bangladesh. Grant No: 56.00.0000.053.20.001.22-306.

References

[1] EM-DAT, The international disasters database. https://www.emdat.be/, 2022. (Accessed 22 August 2022).
[2] E. Alam, Landslide hazard knowledge, risk perception and preparedness in southeast Bangladesh, Sustainability 12 (16) (2020) 6305, https://doi.org/10.3390/su12166305.
[3] B. Ahmed, The root causes of landslide vulnerability in Bangladesh, Landslides 18 (5) (2021) 1707–1720, https://doi.org/10.1007/s10346-020-01606-0.
[4] S. Sultana, Analysis of landslide-induced fatalities and injuries in Bangladesh: 2000–2018, Cogent Social Sciences 6 (1) (2020), 1737402, https://doi.org/10.1080/23311886.2020.1737402.
[5] B. Ahmed, A. Dewan, Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in Chittagong City Corporation, Bangladesh, Rem. Sens. 9 (4) (2017) 304, https://doi.org/10.3390/rs9040304.
[6] M. Chowdhury, B. Hafsa, Landslide susceptibility mapping using bivariate statistical models and GIS in chattagram district, Bangladesh, Geotech. Geol. Eng. 40 (2022) 3687–3710, https://doi.org/10.1007/s10706-022-02111-y.
[7] E. Nohani, M. Moharrami, S. Sharaf, K. Khosravi, B. Pradhan, B.T. Pham, S. Lee, A.M. Melesse, Landslide susceptibility mapping using diferent GIS-based bivariate models, Water 11 (1402) (2019) 1–22, https://doi.org/10.3390/w11071402.
[8] S. Segoni, G. Pappafco, T. Luti, F. Catani, Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization, Landslides 17 (2020) 2443–2453, https://doi.org/10.1007/s10346-019-01340-2.
[9] M. Azarafza, M. Azarafza, H. Akgün, P.M. Atkinson, R. Derakhshani, Deep learning-based landslide susceptibility mapping, Sci. Rep. 11 (1) (2021), 24112, https://doi.org/10.1038/s41598-021-03585-1.
[10] X. Zhou, W. Wu, Y. Qin, X. Fu, Geoinformation-based landslide susceptibility mapping in subtropical area, Sci. Rep. 11 (1) (2021), 24325, https://doi.org/10.1038/s41598-021-03743-5.
[11] A. Tyagi, R.K. Tiwari, N. James, A review on spatial, temporal and magnitude prediction of landslide hazard, J. Asian Earth Sci. X 7 (2022), 100099, https://doi.org/10.1016/j.jaesx.2022.100099.
[12] Y. Huang, L. Zhao, Review on landslide susceptibility mapping using support vector machines, Catena 165 (2018) 520–529, https://doi.org/10.1016/j.catena.2018.03.003.
[13] G.F. Lin, M.J. Chang, Y.C. Huang, J.Y. Ho, Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression, Environ. Geol. 224 (2017) 62–74, https://doi.org/10.1016/j.enggeo.2017.05.009.
[14] Z. Umar, B. Pradhan, A. Ahmad, M.N. Jebur, M.S. Tehrany, Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia, Catena 118 (2014) 124–135, https://doi.org/10.1016/j.catena.2014.02.005.
[15] L.J. Wang, M. Guo, K. Sawada, J. Lin, J. Zhang, A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network, Geosci. J. 20 (1) (2016) 117–136, https://doi.org/10.1007/s12303-015-0026-1.
[16] A.R. Rasyid, N.P. Bhandary, R. Yatabe, Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia, Geoenviron Disasters 3 (2016) 19, https://doi.org/10.1186/s40677-016-0053-x.
[17] B.K. Maheshwari, Earthquake-induced landslide hazard assessment of chamoli district, uttarakhand using relative frequency ratio method, Indian Geotech. J. 49 (1) (2019) 108–123, https://doi.org/10.1007/s40098-018-0334-2.
[18] A. Wubalem, M. Meten, Landslide susceptibility mapping using information value and logistic regression models in GonchaSisoEneses area, northwestern Ethiopia, SN Appl. Sci. 2 (2020) 807, https://doi.org/10.1007/s42452-020-2563-0.
[19] T. Chen, L. Zhu, R.Q. Niu, C.J. Trinder, L. Peng, T. Lei, Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models, J. Mt. Sci. 17 (2020) 670–685, https://doi.org/10.1007/s11629-019-5839-3.
[20] P. Saha, M. Islam, J.T. Oyshi, R. Khanum, A. Nishat, A sustainability analysis on the trends and frequency of the channel flow of a carp breeding river against human interventions and governing public–private partnership (PPP) as adaptation, SN Appl. Sci. 2 (2021) 969, https://doi.org/10.1007/s42452-020-2766-4.
[21] S. Lee, J.H. Ryu, I.S. Kim, Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea, Landslides 4 (4) (2007) 327–338, https://doi.org/10.1007/s10346-007-0088-x.
[22] Á.M. Felicísimo, A. Cuartero, J. Remondo, E. Quirós, Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides 10 (2) (2013) 175–189, https://doi.org/10.1007/s10346-012-0320-1.
[23] O.F. Althuwaynee, B. Pradhan, H.J. Park, J.H. Lee, A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping, Landslides 11 (2014) 1063–1078, https://doi.org/10.1007/s10346-014-0466-0.
[24] M. Meten, N.P. Bhandary, R. Yatabe, GIS-based frequency ratio and logistic regression modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia, J. Mt. Sci. 12 (6) (2015) 1355–1372, https://doi.org/10.1007/s11629-015-3464-3.
[25] A.X. Zhu, Y. Miao, R. Wang, T. Zhu, Y. Deng, J. Liu, L. Yang, C.Z. Qin, H. Hong, A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping, Catena 166 (2018) 317–327, https://doi.org/10.1016/j.catena.2018.04.003.

[26] P.R. Kadavi, C.W. Lee, S. Lee, Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models, Environ. Earth Sci. 78 (2019) 116, https://doi.org/10.1007/s12665-019-8119-1.

[27] Y. Zhao, R. Wang, Y. Jiang, H. Liu, Z. Wei, GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing. Southeastern China, Eng. Geol. 259 (2019), 105147, https://doi.org/10.1016/j.enggeo.2019.105147.

[28] J. Du, T. Glade, T. Woldai, B. Chai, B. Zeng, Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet, Chinese Himalayas, Eng. Geol. 270 (2020), 105572, https://doi.org/10.1016/j.enggeo.2020.105572.

[29] P. Goyes-Penafel, A. Hernandez-Rojas, Landslide susceptibility index based on the integration of logistic regression and weights of evidence: a case study in Popayan, Colombia, Eng. Geol. 280 (2021), 105958, https://doi.org/10.1016/j.enggeo.2020.105958.

[30] E.R. Sujatha, V. Sridhar, Landslide susceptibility analysis: a logistic regression model case study in coonoor, India, Hydrology 8 (2021) 41, https://doi.org/10.3390/hydrology8010041.

[31] V.E. Nwazelibe, C.O. Unigwe, J.C. Egbueri, Integration and comparison of algorithmic weight of evidence and logistic regression in landslide susceptibility mapping of the Orumba North erosion-prone region, Nigeria, Model. Earth Syst. Environ. 9 (2023) 967–986, https://doi.org/10.1007/s40808-022-01549-6.

[32] C. Xi, M. Han, X. Hu, B. Liu, K. He, G. Luo, X. Cao, Effectiveness of Newmark-based sampling strategy for coseismic landslide susceptibility mapping using deep learning, support vector machine, and logistic regression, Bull. Eng. Geol. Environ. 81 (5) (2022) 1–21, https://doi.org/10.1007/s10064-022-02664-5.

[33] H.D. Abeysiriwardana, P.I. Gomes, Integrating vegetation indices and geo-environmental factors in GIS-based landslide-susceptibility mapping: using logistic regression, J. Mt. Sci. 19 (2) (2022) 477–492, https://doi.org/10.1007/s11629-021-6988-8.

[34] A.M. Youssef, H.R. Pourghasemi, Z.S. Pourtaghi, M.M. Al-Katheeri, Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia, Landslides 13 (2016) 839–856, https://doi.org/10.1007/s10346-015-0614-1.

[35] J.C. Kim, S. Lee, H.S. Jung, S. Lee, Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea, Geocarto Int. 33 (9) (2018) 1000–1015, https://doi.org/10.1080/10106049.2017.1323964.

[36] Y. Liu, W. Gong, X. Hu, J. Gong, Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data, Rem. Sens. 10 (6) (2018) 946, https://doi.org/10.3390/rs10060946.

[37] H. Hong, Y. Miao, J. Liu, A.X. Zhu, Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping, Catena 176 (2019) 45–64, https://doi.org/10.1016/j.catena.2018.12.035.

[38] L.H. Nguyen, D.R. Joshi, D.E. Clay, G.M. Henebry, Characterizing land cover/land use from multiple years of Landsat and MODIS time series: a novel approach using land surface phenology modeling and random forest classifier, Rem. Sens. Environ. 238 (2020), 111017, https://doi.org/10.1016/j.rse.2018.12.016.

[39] Y. Wang, D. Sun, H. Wen, H. Zhang, F. Zhang, Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China), Int. J. Environ. Res. Publ. Health 17 (12) (2020) 4206, https://doi.org/10.3390/ijerph17124206.

[40] H. Deng, X. Wu, W. Zhang, Y. Liu, W. Li, X. Li, P. Zhou, W. Zhuo, Slope-Unit scale landslide susceptibility mapping based on the random forest model in deep valley areas, Rem. Sens. 14 (17) (2022) 4245, https://doi.org/10.3390/rs14174245.

[41] M. Marjanović, M. Kovačević, B. Bajat, V. Voženílek, Landslide susceptibility assessment using SVM machine learning algorithm, Eng. Geol. 123 (3) (2011) 225–234, https://doi.org/10.1016/j.enggeo.2011.09.006.

[42] H. Akinci, M. Zeybek, Comparing classical statistic and machine learning models in landslide susceptibility mapping in Ardanuc (Artvin), Turkey, Nat. Hazards 108 (2021) 1515–1543, https://doi.org/10.1007/s11069-021-04743-4.

[43] K.V. Kamran, B. Feizizadeh, B. Khorrami, Y. Ebadi, A comparative approach of support vector machine kernel functions for GIS-based landslide susceptibility mapping, Applied Geomatics 13 (4) (2021) 837–851, https://doi.org/10.1007/s12518-021-00393-0.

[44] Z. Zhao, Z.Y. Liu, C. Xu, Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random forest, CF-SVM and CF-RF models, Front. Earth Sci. 9 (2021), 589630.

[45] M. Ado, K. Amitab, Landslide susceptibility mapping using support vector machine for Meghalaya, India, in: 2023 4th International Conference on Computing and Communication Systems (I3CS), IEEE, 2023, pp. 1–6, https://doi.org/10.1109/I3CS58314.2023.10127361.

[46] H.A. Nefeslioglu, E. Sezer, C. Gokceoglu, A.S. Bozkir, T.Y. Duman, Assessment of landslide susceptibility by decision trees in the metropolitan area of istanbul, Turkey, Math. Probl Eng. 15 (2010), https://doi.org/10.1155/2010/901095, 2010.

[47] B. Pradhan, A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, Comput. Geosci. 51 (2013) 350–365, https://doi.org/10.1016/j.cageo.2012.08.023.

[48] S. Park, J. Kim, Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance, Appl. Sci. 9 (5) (2019) 942, https://doi.org/10.3390/app9050942.

[49] B. Ghasemain, D.T. Asl, B.T. Pham, M. Avand, H.D. Nguyen, S.J. Janizadeh, Shallow landslide susceptibility mapping: a comparison between classification and regression tree and reduced error pruning tree algorithms, Vietnam Journal of Earth Sciences 42 (3) (2020) 208–227, https://doi.org/10.15625/0866-7187/42/3/14952.

[50] U. Khalil, I. Imtiaz, B. Aslam, I. Ullah, A. Tariq, S. Qin, Comparative analysis of machine learning and multi-criteriadecision making techniques for landslide susceptibility mapping of Muzaffarabad district, Front. Environ. Sci. 10 (2022) 1–19, https://doi.org/10.3389/fenvs.2022.1028373.

[51] J. Gui, L.R. Alejano, M. Yao, F. Zhao, W. Chen, GIS-based landslide susceptibility modeling: a comparison between best-first decision tree and its two ensembles (BagBFT and rfbft), Rem. Sens. 15 (4) (2023) 1007, https://doi.org/10.3390/rs15041007.

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

[53] D. Myronidis, C. Papageorgiou, S. Theophanous, Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP), Nat. Hazards 81 (1) (2016) 245–263, https://doi.org/10.1007/s11069-015-2075-1.

[54] G. Das, K. Lepcha, Application of logistic regression (LR) and frequency ratio (FR) models for landslide susceptibility mapping in Relli Khola river basin of Darjeeling Himalaya, India, SN Appl. Sci. 1 (2019) 1453, https://doi.org/10.1007/s42452-019-1499-8.

[55] X. Luo, F. Lin, Y. Chen, S. Zhu, Z. Xu, Z. Huo, M. Yu, J. Peng, Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features, Sci. Rep. 9 (1) (2019), 15369, https://doi.org/10.1038/s41598-019-51941-z.

[56] J.N. Goetz, A. Brenning, H. Petschko, P. Leopold, Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, Comput. Geosci. 81 (2015) 1–11, https://doi.org/10.1016/j.cageo.2015.04.007.

[57] B.T. Pham, B. Pradhan, D.T. Bui, I. Prakash, M.B. Dholakia, A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India), Environ. Model. Software 84 (2016) 240–250, https://doi.org/10.1016/j.envsoft.2016.07.005.

[58] A.M.S. Pradhan, Y.T. Kim, Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea, Bull. Eng. Geol. Environ. 76 (4) (2017) 1263–1279, https://doi.org/10.1007/s10064-016-0919-x.

[59] Y.W. Rabby, A. Ishtiaque, M.S. Rahman, Evaluating the effects of digital elevation models in landslide susceptibility mapping in Rangamati district, Bangladesh, Rem. Sens. 12 (17) (2020) 2718, https://doi.org/10.3390/rs12172718.

[60] H. Wang, L. Zhang, K. Yin, H. Luo, J. Li, Landslide identification using machine learning, Geosci. Front. 12 (1) (2021) 351–364, https://doi.org/10.1016/j.gsf.2020.02.012.

[61] A.M. Youssef, H.R. Pourghasemi, Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia, Geosci. Front. 12 (2) (2021) 639–655, https://doi.org/10.1016/j.gsf.2020.05.010.

[62] V.K. Pandey, A.K. Tripathi, K.K. Sharma, Implications of landslide inventory in susceptibility modeling along a Himalayan highway corridor, India, Phys. Geogr. 43 (4) (2022) 440–462, https://doi.org/10.1080/02723646.2021.1872857.

[63] Z. Bai, Q. Liu, Y. Liu, Landslide susceptibility mapping using GIS-based machine learning algorithms for the Northeast Chongqing Area, China, Arabian J. Geosci. 14 (2021) 1–16, https://doi.org/10.1007/s12517-021-08871-w.

[64] L. Bragagnolo, R.V. da Silva, J.M.V. Grzybowski, Landslide susceptibility mapping with r.landslide: a free open-source GIS-integrated tool based on Artificial Neural Networks, Environ. Model. Software 123 (2020), 104565, https://doi.org/10.1016/j.envsoft.2019.104565.
[65] B. Aslam, A. Maqsoom, U. Khalil, O. Ghorbanzadeh, T. Blaschke, D. Farooq, R.F. Tufail, S.A. Suhail, P. Ghamisi, Evaluation of different landslide susceptibility models for a local scale in the Chitral District, Northern Pakistan, Sensors 22 (9) (2022) 3107, https://doi.org/10.3390/s22093107.
[66] S. Zhou, Y. Zhang, X. Tan, S.M. Abbas, A comparative study of the bivariate, multivariate and machine-learning-based statistical models for landslide susceptibility mapping in a seismic-prone region in China, Arabian J. Geosci. 14 (6) (2021) 440, https://doi.org/10.1007/s12517-021-06630-5.
[67] S.Q. Chowdhury, Chittagong Hill Tracts. *Banglapedia—The National Encyclopedia Of Bangladesh*, 2014.
[68] B. Ahmed, Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh, Landslides 12 (6) (2015) 1077–1095, https://doi.org/10.1007/s10346-014-0521-x.
[69] B. Ahmed, Landslide susceptibility modelling applying user-defined weighting and data-driven statistical techniques in Cox's Bazar Municipality, Bangladesh, Nat. Hazards 79 (3) (2015) 1707–1737, https://doi.org/10.1007/s11069-015-1922-4.
[70] M.A. Islam, S. Murshed, S.M. Kabir, A.H. Farazi, M.Y. Gazi, I. Jahan, S.H. Akhter, Utilization of opensource spatial data for landslide susceptibility mapping at Chittagong District of Bangladesh—an appraisal for disaster risk reduction and mitigation approach, Int. J. Geosci. 8 (4) (2017) 577–598, https://doi.org/10.4236/ijg.2017.84031.
[71] Y.W. Rabby, Y. Li, Landslide susceptibility mapping using integrated methods: a case study in the Chittagong hilly areas, Bangladesh, Geosciences 10 (12) (2020) 483, https://doi.org/10.3390/geosciences10120483.
[72] M. Pal, P.M. Mather, Support vector machines for classification in remote sensing, Int. J. Rem. Sens. 26 (5) (2005) 1007–1011, https://doi.org/10.1080/01431160512331314083.
[73] A. Schneider, Monitoring land-cover change in urban and peri-urban areas using dense time stacks of landsat satellite data and a data mining approach, Rem. Sens. Environ. 124 (2012) 689–704, https://doi.org/10.1016/j.rse.2012.06.006.
[74] H. Shih, D.A. Stow, Y.H. Tsai, Guidance on and comparison of machine learning classifiers for landsat-based land-cover and land-use mapping, Int. J. Rem. Sens. 40 (4) (2019) 1248–1274, https://doi.org/10.1080/01431161.2018.1524179.
[75] M.S. Chowdhury, B. Hafsa, Multi-decadal land cover change analysis over sundarbans mangrove forest of Bangladesh: a GIS and remote sensing based approach, Global Ecology and Conservation 37 (2022), e02151, https://doi.org/10.1016/j.gecco.2022.e02151.
[76] ESRI, How subset features works-ArcMap | documentation. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/how-subset-features-works.htm, 2010.
[77] N. Ahmed, A. Firoze, R.M. Rahman, Machine learning for predicting landslide risk of Rohingya refugee camp infrastructure, Journal of Information and Telecommunication 4 (2) (2020) 175–198, https://doi.org/10.1080/24751839.2019.1704114.
[78] Y.W. Rabby, Y. Li, J. Abedin, S. Sabrina, Impact of land use/land cover change on landslide susceptibility in Rangamati municipality of Rangamati District, Bangladesh, ISPRS Int. J. Geo-Inf. 11 (2) (2022) 89, https://doi.org/10.3390/ijgi11020089.
[79] X. Sun, J. Chen, Y. Bao, X. Han, J. Zhan, W. Peng, Landslide susceptibility mapping using logistic regression analysis along the Jinsha river and its tributaries close to Derong and Deqin County, southwestern China, ISPRS Int. J. Geo-Inf. 7 (11) (2018) 438, https://doi.org/10.3390/ijgi7110438.
[80] S.B. Bai, J. Wang, G.N. Lü, P.G. Zhou, S.S. Hou, S.N. Xu, GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China, Geomorphology 115 (1–2) (2010) 23–31, https://doi.org/10.1016/j.geomorph.2009.09.025.
[81] F. Djeddaoui, M. Chadli, R. Gloaguen, Desertification susceptibility mapping using logistic regression analysis in the Djelfa area, Algeria, Rem. Sens. 9 (10) (2017) 1031, https://doi.org/10.3390/rs9101031.
[82] H.R. Pourghasemi, O. Rahmati, Prediction of the landslide susceptibility: which algorithm, which precision? Catena 162 (2018) 177–192, https://doi.org/10.1016/j.catena.2017.11.022.
[83] M.L. Süzen, B.Ş. Kaya, Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping, International Journal of Digital Earth 5 (4) (2012) 338–355, https://doi.org/10.1080/17538947.2011.586443.
[84] K.T. Chang, A. Merghadi, A.P. Yunus, B.T. Pham, J. Dou, Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques, Sci. Rep. 9 (1) (2019) 1–21, https://doi.org/10.1038/s41598-019-48773-2.
[85] K.T. Chau, J.E. Chan, Regional bias of landslide data in generating susceptibility maps; Case of Hong Kong Island, Landslides 2 (2005) 280–290, https://doi.org/10.1007/s10346-005-0024-x.
[86] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[87] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, https://doi.org/10.1007/BF00058655.
[88] T.G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization, Mach. Learn. 32 (1998) 1–22.
[89] T.K. Ho, Random decision forests, in: Document Analysis and Recognition, Proceedings of the Third International Conference, Montreal, Quebec, Canada vol. 1, IEEE, New York City, NY, 1995, pp. 278–282.
[90] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 832–844, https://doi.org/10.1109/34.709601.
[91] Y. Amit, D. Geman, Shape quantization and recognition with randomized trees, Neural Comput. 9 (7) (1997) 1545–1588, https://doi.org/10.1162/neco.1997.9.7.1545.
[92] K. Fawagreh, M.M. Gaber, E. Elyan, Random forests: from early developments to recent advancements, Systems Science & Control Engineering: An Open Access Journal 2 (1) (2014) 602–609, https://doi.org/10.1080/21642583.2014.956265.
[93] D.R. Cutler, T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J.J. Lawler, Random forests for classification in ecology, Ecology 88 (11) (2007) 2783–2792, https://doi.org/10.1890/07-0539.1.
[94] A. Cutler, D.R. Cutler, J.R. Stevens, Random forests, in: C. Zhang, Y. Ma (Eds.), Ensemble Machine Learning, Springer, Boston, MA, 2012, https://doi.org/10.1007/978-1-4419-9326-7_5.
[95] P. Sajadi, Y.F. Sang, M. Gholamnia, S. Bonafoni, S. Mukherjee, Evaluation of the landslide susceptibility and its spatial difference in the whole Qinghai-Tibetan Plateau region by five learning algorithms, Geoscience Letters 9 (2022) 9, https://doi.org/10.1186/s40562-022-00218-x.
[96] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[97] J. Roy, S. Saha, A. Arabameri, T. Blaschke, D.T. Bui, A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India, Rem. Sens. 11 (23) (2019) 2866, https://doi.org/10.3390/rs11232866.
[98] V. Vakhshoori, M. Zare, Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? Geomatics, Nat. Hazards Risk 9 (1) (2018) 249–266, https://doi.org/10.1080/19475705.2018.1424043.
[99] H.R. Pourghasemi, B. Pradhan, C. Gokceoglu, Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran, Nat. Hazards 63 (2) (2012) 965–996, https://doi.org/10.1007/s11069-012-0217-2.
[100] A. Yavuz Ozalp, H. Akinci, M. Zeybek, Comparative analysis of tree-based ensemble learning algorithms for landslide susceptibility mapping: a case study in Rize, Turkey, Water 15 (14) (2023) 2661, https://doi.org/10.3390/w15142661.
[101] X. Guo, B. Fu, J. Du, P. Shi, Q. Chen, W. Zhang, Applicability of susceptibility model for rock and loess earthquake landslides in the eastern Tibetan plateau, Rem. Sens. 13 (13) (2021) 2546, https://doi.org/10.3390/rs13132546.
[102] D.T. Bui, B. Pradhan, O. Lofman, I. Revhaug, O.B. Dick, Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks, Geomorphology 171 (2012) 12–29, https://doi.org/10.1016/j.geomorph.2012.04.023.
[103] M. Ado, K. Amitab, A.K. Maji, E. Jasińska, R. Gono, Z. Leonowicz, M. Jasiński, Landslide susceptibility mapping using machine learning: a literature survey, Rem. Sens. 14 (13) (2022) 3029, https://doi.org/10.3390/rs14133029.
[104] C.J.F. Chung, A.G. Fabbri, Validation of spatial prediction models for landslide hazard mapping, Nat. Hazards 30 (3) (2003) 451–472, https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b.

[105] W. Chen, H. Pourghasemi, S. Naghibi, A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China, Bull. Eng. Geol. Environ. 77 (2) (2017) 647–664, https://doi.org/10.1007/s10064-017-1010-y.
[106] B. Ahmed, M.S. Rahman, R. Islam, P. Sammonds, C. Zhou, K. Uddin, T.M. Al-Hussaini, Developing a dynamic Web-GIS based landslide early warning system for the Chittagong Metropolitan Area, Bangladesh, ISPRS Int. J. Geo-Inf. 7 (12) (2018) 485, https://doi.org/10.3390/ijgi7120485.
[107] M.S. Rahman, B. Ahmed, L. Di, Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria, J. Mt. Sci. 14 (10) (2017) 1919–1937, https://doi.org/10.1007/s11629-016-4220-z.
[108] M.M. Mourin, A.A. Ferdaus, M.J. Hossain, Landslide Susceptibility Mapping in Chittagong District of Bangladesh Using Support Vector Machine Integrated with GIS,International Conference on Innovation in Engineering and Technology (ICIET), Dhaka,Bangladesh, IEEE, New York City, NY, 2018, pp. 1–5, https://doi.org/10.1109/CIET.2018.8660782.
[109] S. Lee, T. Sambath, Landslide susceptibility mapping in the DamreiRomel area, Cambodia using frequency ratio and logistic regression models, Environ. Geol. 50 (6) (2006) 847–855, https://doi.org/10.1007/s00254-006-0256-7.
[110] S. Park, C. Choi, B. Kim, J. Kim, Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea, Environ. Earth Sci. 68 (2013) 1443–1464, https://doi.org/10.1007/s12665-012-1842-5.