



Selecting optimal conditioning parameters for landslide susceptibility: an experimental research on Aqabat Al-Sulbat, Saudi Arabia

Saeed Alqadhi¹ · Javed Mallick² · Swapan Talukdar³ · Ahmed Ali Bindajam⁴ · Nguyen Van Hong⁵ · Tamal Kanti Saha³

Received: 12 May 2021 / Accepted: 5 August 2021 / Published online: 14 August 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Landslides and other disastrous natural catastrophes jeopardise natural resources, assets, and people's lives. As a result, future resource management will necessitate landslide susceptibility mapping (LSM) using the best conditioning factors. In Aqabat Al-Sulbat, Asir province, Saudi Arabia, the goal of this study was to find optimal conditioning parameters dependent hybrid LSM. LSM was created using machine learning methods such as random forest (RF), logistic regression (LR), and artificial neural network (ANN). To build ensemble models, the LR was combined with RF and ANN models. The receiver operating characteristic (ROC) curve was used to validate the LSMs and determine which models were the best. Then, utilising random forest (RF), classification and regression tree (CART), and correlation feature selection, sensitivity analysis was carried out. Through sensitivity analysis, the most relevant conditioning factors were determined, and the best model was applied to the important parameters to build a highly robust LSM with fewer variables. The ROC curve was also used to evaluate the final model. The results show that two hybrid models (LR-ANN and LR-RF) were predicted the very high as 29.67–32.73 km² and high LS regions as 21.84–33.38 km², with LR predicting 22.34km² as very high and 45.15km² as high LS zones. The LR-RF appeared as best model (AUC: 0.941), followed by LR-ANN (AUC: 0.915) and LR (AUC: 0.872). Sensitivity analysis, on the other hand, allows for the exclusion of aspects, hillshade, drainage density, curvature, and TWI from LSM. The LSM was then predicted using the LR-RF model based on the remaining nine conditioning factors. With fewer variables, this model has achieved greater accuracy (AUC: 0.927). This comes very close to being the best hybrid model. As a result, it is strongly advised to choose conditioning parameters with caution, as redundant parameters would result in less resilient LSM. As a consequence, both time and resources would be saved, and precise LSM would indeed be possible.

Keywords Landslide susceptibility mapping; · Hybrid machine learning; · Statistical technique; · Sensitivity analysis; · Optimum conditioning factors

Responsible Editor: Philippe Garrigues

✉ Javed Mallick
jmallick@kku.edu.sa

Saeed Alqadhi
sdaqadi@kku.edu.sa

Swapan Talukdar
swapantalukdar65@gmail.com

Ahmed Ali Bindajam
abindajam@kku.edu.sa

Nguyen Van Hong
hongnv@ig.vast.vn

Tamal Kanti Saha
TAMALKANTISAHA999@gmail.com

¹ Department of Civil Engineering, College of Engineering, King Khalid University, Abha, Kingdom of Saudi Arabia

² Department of Civil Engineering, College of Engineering, King Khalid University, P.O. Box: 394, Abha 61411, Kingdom of Saudi Arabia

³ Department of Geography, University of Gour Banga, Malda, India

⁴ Department of Architecture and Planning, College of Engineering, King Khalid University, Abha, Kingdom of Saudi Arabia

⁵ Institute of Geography, Vietnam Academy of Science and Technology, Hanoi, Vietnam

Introduction

Large-scale movements of rocks, gravel, and dirt from the top to the surface of a mountain are recognized as landslides (Cruden 1991). Landslides accounted for 5% of all natural disasters worldwide between 1994 and 2013 (CRED 2016). Landslide responsible for 4914 deaths, 27 110 people were homeless and 2.1 billion losses from 2014 to 2018 (EM-DAT 2013). Landslides are dangerous to people's lives, the environment, property, and resources. Therefore, the assessment of landslide susceptibility becomes an emerging topic in order to reduce the risk posed by landslides (Li and Chen 2020).

Landslides are triggered by a variety of factors, for instance geological evolution (Agostini et al. 2014), landform (Kim et al. 2015; Liucci et al. 2017), land use type, groundwater (Peng et al. 2018), irrigation (Alvioli et al. 2018; Peres and Cancelliere 2018), precipitation (Chang et al. 2017; Segoni et al. 2018). It is important to establish a landslide management and prevention scheme for a country in order to avoid accidents caused by landslides and to ensure the safe growth of mountainous regions. Regional landslide vulnerability maps are generally useful for reducing the impact of landslide hazards (Chen et al. 2019). According to Guzzetti et al. (2005), landslide susceptibility mapping helps to communicate the expected outcomes of landslides on the basis of local geographical conditions and managers agree that they can be used for successful land-like assessment and response (Hong et al. 2016a, 2016b). LSM findings are dependent on accurate data and models that are applicable (Bui et al. 2011). There has been a variety of landslide models developed, including mechanically, heuristically, and statistically dependent approaches (Luo and Liu 2018). Physically-based methods assess landslide susceptibility by analysing slope stability using the laws of mechanics (Crosta et al. 2003; Stamatopoulos and Di 2015). Since they require very precise topography details, these methods are typically used for a small area. Depending on the experts, heuristic-based approaches determine the probability of landslide events. According to Regmi et al. (2014) such methods are strongly based on professional knowledge and in general achieve moderate precision. Statistically based models, such as the probability-frequency ratio (FR) (Lee and Dan 2005; Chen et al. 2017b), statistical index (SI) (Regmi et al. 2014; Mandal and Mandal 2018), weight of evidence (WoE) (Xu et al. 2012; Xie et al. 2017), certainty factors (CF) (Chen et al. 2016; Hong et al. 2018), index of entropy (IoE) (Jaafari et al. 2014; Bui et al. 2018), logistic regression (LR) (Mandal and Mandal 2018; Zhang et al. 2019) and evidential belief function (EBF) (Ding et al. 2017; Pradhan and Kim 2017; Chen et al. 2018b), have been used more extensively in landslide susceptibility assessment than the two categories listed above. These approaches are more objective and quantitative, since they are focused on previous landslides and contributing factors. Standard statistical methods, on the other

hand, have a limited capacity to predict the dynamic and non-linear interactions between landslides and the conditioning variables (He et al. 2012). However, since no one approach or methodology is best for all areas, machine learning is taken into consideration.

With the accelerating development in remote sensing technologies, a massive amount of landslide-related data is becoming more easily accessible. Most studies have used big data to model landslide susceptibility using machine learning techniques, as these methods can analyse the dynamic relationship between landslide susceptibility and influencing factors (Zhu et al. 2018). According to Jordan and Mitchell (2015) machine learning is an artificial intelligence branch that employs computational algorithms to analyse and predict data through learning from training data. As per a review of the literature, different machine learning algorithms were used to assess landslide susceptibility, like artificial neural network (Chen et al. 2017c), neuro-fuzzy (Tien Bui et al. 2012), decision trees (Tien Bui et al. 2014; Hong et al. 2015), support vector machines (Chen et al. 2018a). Tsangaratos and Iliá (2016a, 2016b), for example, updated a DT model to measure the landslide vulnerability of a specific study area in Greece and achieved good results (AUC = 0.803). Marjanović et al. (2011) assessed landslide susceptibility using three machine learning approaches (SVM, DT, and logistic regression (LR)); the results showed that the SVM (AUC = 0.71) outperformed the other models. Pourghasemi and Rahmati (2018) examined ten sophisticated machine learning approaches for landslide susceptibility modelling, such as SVM, ANN, and RF, and discovered that the RF approach generated the best model (AUC = 0.837). While it is obvious that machine learning algorithms enhance prediction accuracy of regional landslide occurrence, the generalisation efficiency of single classifiers also needs to be improved (Truong et al. 2018). However, until now, landslide researchers have been unable to agree on an appropriate model for assessing landslide susceptibility (Chen et al. 2018a). As a result, a number of ensemble methods have recently gained popularity in geohazard susceptibility mapping (Chen et al. 2017c; Hong et al. 2018; Mahato et al. 2021).

In recent years, experimental hybrid approaches for landslide studies have been considered due to the need for explorations of new landslide methods and techniques in order to gain additional research background for drawing fair conclusions (Tien Bui et al. 2016; Chen et al. 2017a). So many hybrid methods for landslide modelling have been successfully used, which were created by integrating statistical techniques with machine learning approaches, such as the stepwise weight assessment ratio analysis (SWARA) technique, adaptive neuro-fuzzy inference system (ANFIS) (Dehnavi et al. 2015), ANN-fuzzy logic (Kanungo et al. 2006), EBF-fuzzy logic (Bui et al. 2015), ANFIS combined with the frequency ratio (Chen et al. 2017b), and rough set-SVM (Peng

et al. 2014). (Pham et al. 2019) for example, used Reduced Error Pruning Trees (REPT) and various ensemble strategies to build four hybrid landslide susceptibility models, with the best one achieving an AUC of 0.872. The detailed evaluation of landslide-related independent variables in each class of independent layers is an essential capability of hybrid methods in landslide studies. This capability may increase the popularity of this technique and aid researchers in future landslide studies. Prediction of landslide-prone areas using modern hybrid approaches, on the other hand, is essential for landslide research due to the method's better accuracy in identifying and greater prediction in relation to machine learning models.

Furthermore, previous comparable research paid little attention to the sensitivity study of thematic layers. After constructing hybrid models, thematic layers in this analysis were subjected to sensitivity tests. To improve the model's prediction accuracy, the most influential thematic layers were calculated using multiple machine learning-based sensitivity analyses and those that were redundant were excluded. Following that, best hybrid models integrated influential thematic layers to obtain high precision models. This technique was also used by other RS/GIS-based models, such as soil erosion susceptibility (Abdulkadir et al. 2019), Flood susceptibility (Islam et al. 2021), landslide susceptibility (Chen et al. 2018b; Mind'je et al. 2020), and soil properties prediction (Forkuor et al. 2017; Pham et al. 2019), in order to decrease uncertainties. The CART, RF, and correlation sensitivity analyses have been used in the research to classify the significant thematic layers in the model production. Furthermore, the model output was tested using the receiver operating characteristic (ROC) cross-validation process (Hanley and McNeil 1982) and by comparing the proposed model with the real landslide positions determined by in situ measurements. In view of all this, this research will play an important role in the sustainable management of landslides in the field of study by correctly forecasting the vulnerability to landslides and determining the level of association between the vulnerable landslide zones and their corresponding sensitive parameters, making this study special and novel for the field of study.

The principal objectives of the present study are to (1) develop hybrid machine learning based LS models by integrating LR as statistical technique and the machine learning algorithms like RF and ANN; (2) perform sensitivity analysis using CART, RF, and correlation feature selection; and (3) generate LSM using most sensitive parameters with best representative model. This research would assist planners, regulators, lawmakers, and municipal governments in reducing landslide incident in this region by better use and management practices.

Materials and methodology

Materials

A variety of materials were gathered for the current analysis in order to produce landslide causing variables from different sources. The "ALOS PALSAR DEM" was given by the National Aeronautics and Space Administration's Earth Science Data Systems. The sentinel-2 satellite image was obtained from the US Geological Survey's earth explorer repository (<https://earthexplorer.usgs.gov/>). The geological map was constructed using a Saudi Geological Survey map at a scale of 1:100000. Soil texture data was obtained during the field survey. Google Earth imagery and ArcGIS tools were used to digitise and plan the drainage chart and path (version 10.5).

Study area

Aqabat Al-Sulbat is a 199km² area in Saudi Arabia's Asir Region, located along the Abha-Bahah Road in the northern part of the Balqarn Area (Fig. 1). Between 19°45'4.407"N and 19°54'42.055"N, and 41°40'52.31"E and 41°53'6.169"E, respectively, is the study region. After a rainy storm swept over the Al-Baha Region in the winter season of January and February 2020, a landslide (lateral spreading) occurred in the Aqabat Al-Sulbat region of road 15, station 907+00. Rainfall and damp weather had a key part in the huge landslide's development (lateral spreading). The lateral spreading from Al-Baha City is projected to spread out across a 50-km area. Major A 400-500 m stretch of the current old/new road is undergoing unfavourable lateral motions towards the Wadi side, as shown by fractures in the pavement. The heavy downpour soaked the upper few metres of the earth on the road embankment, resulting in shallow rotational slides. The infiltration and flowdown began on the mountain side (toe of hills) and moved down the wadi side's south facing slope. There were heaps of debris/rock avalanches along the failing slope (deep Wadi deep to 100 m, roughly).

The height of the study area ranges from 989 to 2404 m above sea level, with an average elevation of 1768 m. The site is part of the Ablah Group, which is structurally shielded by the Farwah Shear Zone between the Shwas-Tayyah and Al Lith-Bidah structural belts, as seen on the geological map. This group includes the Jerub, Rafa, and Thurat formations, which contain a sequence of volcanic and epiclastic rocks. Fractures may be in various shapes and sizes, but they all divide the rock into cubic or quadrangular blocks. Both joints show signs of weathering and potential corrosion seats. Plane collapse is a form of slope failure caused by the movement of a volume of rocks over a single discontinuity. Slipping collapse of a long continuous rock slope often entails the formation of releasing surfaces that allow the mass to descend dip

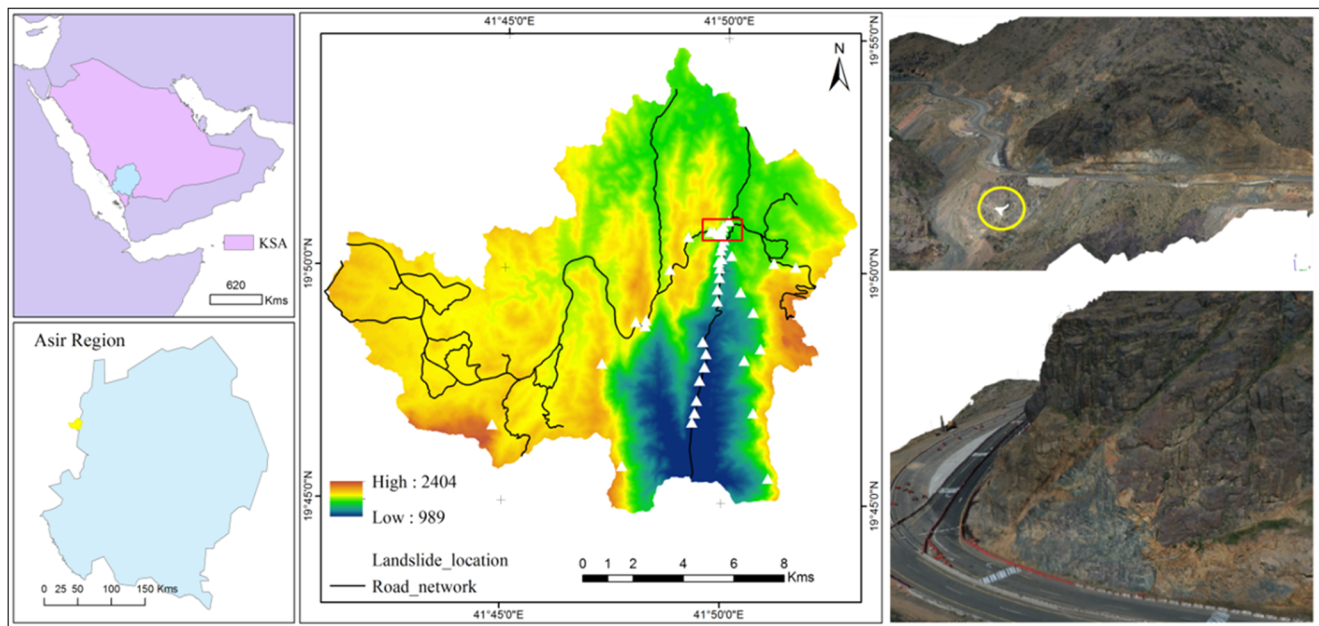


Fig. 1 Showing the study area, landslide locations and highlighted one of the prone landslide area with 3D model with “visual shadowing” of nadir-viewing UAV photos caused by overhanging rocks

over the basal plane. In the Aqabat Al-Sulbat area, wedge failure of a rock slope seems to be a very simple pattern of failure. The climatic conditions in the research area are cold and semi-arid. The long-term average rainfall is 218 mm, with February and June accounting for 75% of total rainfall. The annual high and low temperatures in the study area are 29.5 °C and 16.8 °C, respectively.

Landslide inventories

For training and evaluating the LS model, the method of developing landslide inventories is important. This study investigated and identified the locations of landslide events using data from a field survey and a government report. 50 landslide sites were gathered using the global positioning system (GPS) and Google Earth. Eighty percent (40 points) of the total data points were chosen at random to create the training datasets, while the remaining 20% (10 points) were chosen to create the research databases. The current study makes use of classification-dependent LS mapping, which incorporates binary data. As a result, we needed areas that were not vulnerable to landslides. Tang et al. (2020) suggest that with an excellent result, the same number of negative points be used. However, no specific literature on the set of negative point numbers has been discovered. As a result, 50 non-landslide sites were selected at random based on past landslide history, local people's views, and Google Earth. Non-landslide points or locations were those where a landslide had yet to be observed. Non-landslide data, including landslide data, is partitioned by an 80-20 ratio for teaching and processing datasets. A total of 80% of both landslide and non-landslide

regions is used to create the training dataset. By assigning a value of 0 to landslide sites and a value of 1 to non-landslide sites, we were able to generate binary training datasets. The research database has also been developed in binary format. The data was retrieved from the landslide-causing variables using the training dataset in the ArcGIS 10.5 software's spatial analyst' toolbox.

Landslide conditioning parameters

The complex relationship between landslide triggering factors and past landslides is the subject of the landslide susceptibility modelling (Jebur et al. 2015). The selection of suitable parameters is a difficult process. Expert experience of the research area and previous literature can help in the selection of landslide triggering factors (Wang et al. 2018). For modelling landslide vulnerability, elevation, aspect, slope, geology, topographic wetness index, curvature, soil texture, lineament, drainage density, distance to road, land use/ land cover, and NDVI were chosen as landslide triggering factors.

Elevation

Various geological, geomorphological, and climatic variables, such as lithological units, weathering, precipitation and wind action, influence altitude (Youssef and Pourghasemi 2021; Mallick et al. 2021). Altitude is one of the topographic variables that influences slope instability, according to Mallick et al. (2021). It's been utilised in nearly every landslide susceptibility analysis. A DEM was used to construct the

elevation map. The height of the study area varies from 2404 to 989 m (Fig. 2a).

Slope

The number of landslides increases as the slope angle becomes higher. The connection between shear forces and resistance to shear (safety factor) may be used to determine the slope's stability against failure, where the driving force of mass movement rises as the slope angle increases. Slope steepness is connected to both the shear forces acting on the hill slope and the displacement of the landslide mass, according to Pham et al. (2019). According to Nhu et al. (2020), slope gradient is important in subsurface flow and has an influence on soil moisture content which are closely linked to the incidence of landslides. The elevation map was obtained using a DEM. The research area's altitudes ranging from 2404 to 989 m (see Fig. 2a.)

Curvature

For landslide conditioning, curvature is one of the most critical topographical parameters. The primary function of curvature is to induce surface run-off. It has an effect on infiltration as well. Concave (negative curvature), flat (zero curvature), and convex (positive curvature) are the three forms of curvature (positive curvature). Convex slopes are much more capable of producing run-off than concave slopes (Lee and Min 2001). Drainage is also determined by the modes of curvature in mountainous areas (Fig. 2e).

Aspect

Wind directions, precipitation patterns, sunshine influence, discontinuity orientations, hydrological processes, evapotranspiration, soil moisture concentration, vegetation, and root growth are all factors that have direct and indirect effects on landslides (Nhu et al. 2020). Several research have revealed that there is a link between the aspect and other geo-environmental variables (landslide-related factors) (Youssef and Pourghasemi 2021). Flat (1), North (0–22.5°; 337.5°–360°), Northeast (22.5°–67.5°), East (67.5°–112.5°), Southeast (112.5°–157.5°), South (157.5°–202.5°), Southeast (202.5°–247.5°), West (247.5°–292.5°), and Northeast (292.5°–337.5°) were the nine categories used to classify the slope aspect in the research region (Fig. 2c).

Geology

The physical features of rocks and their types specifically control pitch instability or failure. The Saudi Geological Survey digitised the geology map of the sample area at a scale of 1:100000. The toolbox of ArcGIS “Space Analytics” has

been converted into raster format. The geological map of the study region, on the other hand, shows ten geological types, including “(1) QuarzSyeniteStock, Blades, and Dikes (GA), (2) Hornblende diorite, Mafic diorite, and Mafic tonalite, and 3) QuarzSyeniteStock, Blades, and Dikes (GA), (4) Andesitic and dacitic pyroclastic rocks with small flows (AJ), (5) Feldspathic and lithic greywacke (BHK), and (6) Andesitic, dacitic, and basaltic flood mud, lithic and crystal tuff, and volcanic breccia (JQ), (7) Monzogranite (GR), olivine gabbro, metagabbro, metadiorite, and anortho (GB), (8) amphibole and biotite schist (BA), (9) feldspathic greywacke, carbonaceous chert, argillite, and slate (BHR)”.

Soil texture

Soil samples were taken from a variety of sites within the test site. The research field yielded 32 soil samples, each containing approximately 1 kg of accumulated stability (0–30 cm deep). The GPS coordinates of the sampling sites were registered. It was decided to use a stratified composite sampling method. Elevation areas, LULC zones, and soil moisture zones were used to divide the study region. The site was then examined independently, with two replicates gathered two to three metres apart at each survey site. The soil texture and organic matter were isolated using a normal procedure, with each sample being carefully weighed and sieved at a depth of 2 mm. The hydrometer method was used to measure the volume of the soil grains (texture analysis) (Stokes' law). Figure 2f shows the distribution of soil types in the sample area. Soil texture is divided into four categories: loamy sand, sandy loam, loam, and sandy loam.

Lineament density

The lineament refers to any area where there are longitudinal tectonic faults, resulting in a lessening of rock potency. The lineament will take many different types, including fault, fracturing, and shearing. A lineament is a kind of discontinuity and weaker section in geological formations. Long considered to be the most important factor in landslides is the poorer component of the geology, or lineament (Nhu et al. 2020). Using ENVI software, the sentinel-2 satellite image was used to extract lineament. The line density function was used to generate the map of lineament density (Fig. 2h). Landslides are more likely to occur in areas with a higher density of lineaments.

TWI

The TWI is a significant hydrologic parameter that has an effect on landslide prevention. The TWI is determined to be a significant landslide-controlled hydrological parameter. It measures the amount of water that has accumulated in the tank

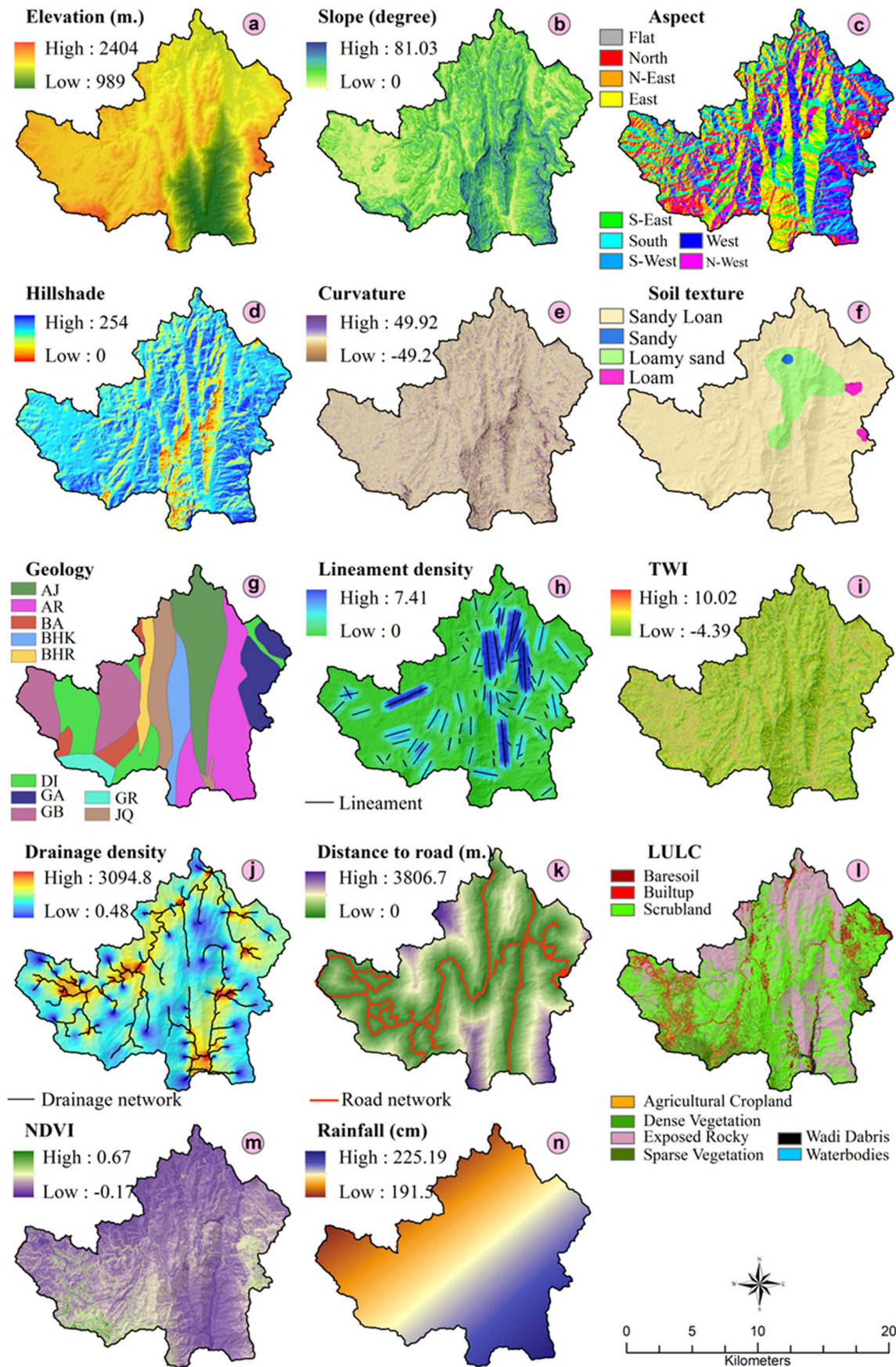


Fig. 2 Landslide conditioning factors: a. elevation, b. slope, c. aspect, d. Hill shade, e. Curvature, f. Soil texture, g. Geology, h. Lineament density, i. TWI, j. Drainage density, k. Distance to road, l. LULC, m. NDVI, n. Rainfall

or on the site. Its impact on soil moisture causes porous water pressure and lowers soil resistance, which controls slope failure in particular. The higher the TWI value, the more likely it is that landslides will occur. In the studied area, TWI ranges from 0–10.03 (Fig. 2i).

NDVI

The NDVI or woodland cover has a major impact on the occurrence of landslides. The increasing interception of drainage, evapotranspiration, and infiltration by the canopy of vegetation has a significant impact on soil hydrology. This may be critical for long-term rain. The stability of the soil improves as the root becomes stronger. A collapse is more likely to occur if there is more timber, and vice versa. This analysis used Sentinel 2 data from bands 8 and 4 to exclude vegetation. The NDVI is in the region of 0.67 to -0.17. In the western half of the sample area, there is more vegetation.

LULC mapping

By altering the LULC and disrupting the slope's longevity, LULC has an effect on the slope's resistance (Sidle and Ochiai 2000). In this paper, the sentinel-2 image was used to build a LULC map using a maximum likelihood classifier. Built-up areas, water sources, thick forest, sparse vegetation, cropland, scrubland, barren rocky, bare earth, and wadi debris are all examples of LULC (Fig. 2l).

Drainage density (Dd)

The Dd is the total length of all streams in a river system split by the basin's total surface area (Pal and Saha 2017). The landslide is inversely related to drainage density. The risk of landslide vulnerability increases as drainage density increases, and vice versa. The maps of drainage network were created using DEM data. The drainage density was then determined by integrating the digitised drainage map into ArcGIS software. The line density function in ArcGIS software was used to map drainage density. The drainage density was higher in the eastern half of the test area than in the western half (Fig. 2j).

Distance to road

Building highways in mountainous areas disturbs the existing ground cover and disrupts the slope balance by destroying the slope. As a result, transportation infrastructure contributes to slope failure. The stability of the slope transitions from smooth to unstable during road construction and vehicle travel, which can result in cracks. These cracks absorb a lot of water, causing slopes to loosen. As a result, a landslide is caused. Intense snow can accelerate the process of slope

collapse, resulting in landslides. The lower the distance or the closer the road is in a mountainous environment, the higher the risk of landslides. The road map was digitised using Google Earth, and the distance from the road map was derived using the Euclidean Distance Method in ArcGIS. The south-east portion of the sample area has a heavy concentration of roads (Fig. 2k).

Rainfall

However, one of the primary factors for LSM is landslide conditioning variables such as flooding, snowfall, and earthquake (Baeza et al. 2016). Since landslides in the current study area are caused by both serious and average rainfall, irregular or heavy rainfall is regarded as a major landslide triggering factor. The research district has four rain gauge stations that are evenly distributed. Both gauge stations in the study region receive almost the same amount of rainfall. As a result, in these cases, the inverse distance weighting (IDW) technique, which is similar to kriging, will provide reasonable results. As a result, in this analysis, the IDW was used to generate interpolated rainfall layers.

Multicollinearity statistic

To evaluate landslide vulnerability and risk, fourteen landslide-causing variables were selected for this study. Since mathematical and machine learning algorithms are susceptible to collinearities, it is important to carefully optimise and check the collinearity of the chosen variables when forecasting landslide susceptibility. Furthermore, collinearity will interrupt the modelling process, lowering the accuracy of landslide susceptibility prediction (Mukherjee and Singh 2020). In order to obtain adequate results, this analysis helps in the collection of appropriate parameters by excluding redundancy parameters (Talukdar et al. 2020). Methods for quantifying multicollinearity include variance inflation factors (VIF), resistance (TOL), linear support vector machine, and chi-square (Talukdar et al. 2021). The resistance and VIF were used in this study to quantify multicollinearity among the selected variables. Several previous studies using TOL and VIF obtained excellent results. It should be noted that, in order to achieve high precision results, the variables with collinearity problems should be removed at the end of the analysis. The higher the collinearity, the bigger the VIF. The coefficient of certainty was determined using a linear regression analysis with the landslide data as the response variables and the landslide causing measures as predictor factors to measure multicollinearity. Equations 1 and 2 were used to measure the VIF and TOL of the input variable using this value (Talukdar et al. 2021).

$$T = 1 - R^2 \quad (1)$$

$$VIF = \frac{1}{T} \quad (2)$$

Methods adopted for LSMs

In the realm of natural hazards, machine learning methods (MLTs) have recently become a cornerstone in solving spatial modelling issues (e.g., landslide susceptibility assessment) (Mallick et al. 2021; Zhang et al. 2017; Zhu et al. 2018; Youssef et al. 2016; Xie et al. 2017; Xi et al. 2019; Wang et al. 2018; Xu et al. 2012). MLTs, according to Youssef and Pourghasemi (2021), should be used as a supplement to human knowledge rather than as a replacement. MLTs should be combined with the processing capabilities of GIS and field datasets, according to Tang et al. (2020), in order to generate new accurate models that can be used by decision makers, Earth scientists, and planners (e.g., historical landslide inventory and geo-environmental factors). The advantages of machine learning approaches are (a) their capacity to modify their internal structure to the available landslide data, (b) their ability to retrieve information from large datasets in an automated manner, (c) They can build classification (predicting categorical predictor factors) and regression (predicting continuous dependent factors) models to furnish a precise landslide model; their models are quite cost effective and fast than traditional models, and they could even be widened to large area assessment (Felicesimo et al. 2013). In this work, a variety of sophisticated machine learning algorithms with varying degrees of complexity were used to assess their usefulness in landslide susceptibility mapping. LR, ANN, RF, and their ensembles are among them.

LR One form of regression analysis is logistic regression (LR), in which categorical outcomes can be predicted using a specific predictor (Saha and Pal 2019). Probabilities of potential events can be modelled using logistic functions. For two-class classification, the LR model is effective (Agarwal et al. 2016). “Considering n samples of the pairs, (x_i, y_i) , $i = 1, 2, \dots, n$, $y_i \in \{-1, +1\}$ is a binary category mark for each sample $i = 1, 2, \dots, n$ and weights (w, b) ” The frequency likelihood of the class has been modelled with the following equation in LR for binary classification:

$$P(y = \pm 1 | x, v) = \frac{1}{1 + \exp(-y(v^z x + i))} \quad (3)$$

“where I denotes the intercept, z indicates matrix transposition, and $v = (v_1, v_2, \dots, v_k)^z$ is the k -dimensional coefficient vector to be determined”

ANN Since it could address non-linear statistical environmental problems by optimising neurons from each hidden layer, the multilayer perceptron (MLP) model based on ANN has

been widely used for natural hazard modelling (Pradhan and Lee 2010; Xi et al. 2019; Pourghasemi et al. 2020; Mehrabi et al. 2020; Sun et al. 2020; Mahato et al. 2021). The MLP algorithm is a subset of ANN that is characterised by a parallel information system comprised of several neurons. These are linked by input, secret, and output layers (Fakiola et al. 2010; Pradhan and Lee 2010; Saha et al. 2021). During the training cycle, it is capable of learning tasks without prior knowledge of the problems and finding pattern similarities easily (Arnone et al. 2014). The modelling method includes several stages, including (a) transition function, which depicts a function applied to the weighted input of neurons to produce the output, (b) qualified network architecture, which determines the network conditions, and (c) common learning law, which requires complex algorithms applied to the neural network's relation weights on how to adjust correctly. In this study, the MLP architecture was used. The backpropagation algorithm was then used to practise the MLP architecture. The backpropagation algorithm has the advantage of reducing the global error between actual and forecasted data during the training cycle. The sigmoid and linear activation functions were used in the output and hidden layers. In a number of previous experiments, landslide susceptibility simulation has been frequently used (Fakiola et al. 2010; Pradhan and Lee 2010; Arnone et al. 2014; Zhang et al. 2019).

RF A RF is an ensemble ML algorithm that explores the spatial relationship between real events and landslide causes and then constructs a classification (Genuer et al. 2010). In order to predict or approximate efficiency, a set of features was selected and weighted based on the results of voting. Based on the results of the assessed decision trees, the majority of votes were pooled and a single decision tree was created for final classification (Beven and Kirkby 1979). A single decision tree is used to make highly precise predictions to minimize uncertainty. The RF model rectifies the decision tree's over-fitting dilemma during the training process. Obtaining high variance from different decision trees is crucial in RF classification. The establishment of most trees and factors used for sampling phase and spilt search are both core elements of the RF algorithm during analysis (Stumpf and Kerle 2011). In this analysis, 500 trees were used to train the model.

Procedure for hybrid machine learning model To begin, data on conditioning parameters was gathered using binary landslide inventory datasets. The LR model was implemented on the basis of the data sets and weights have been obtained for all conditioning parameters. Following that, the weights were combined with the conditioning parameters. As a result, the weighted conditioning parameters have been computed. However, according to the collected weights, the weights datasets is erratic and multi-directional (Some are positive and negative, some has high value). As a result, in order to

render the datasets unidirectional and standard, the weighted datasets were normalised using a fuzzy membership function with a scale of 0 to 1. A membership function that is close to one implies a high membership function, and is associated with high landslide cases, and vice versa. Again, landslide inventory was used to gather data from normalised weighted parameters. Then, machine learning algorithms such as ANN and RF were used to build a hybrid machine learning model.

Validation of LSMs

Model confirmation is the most important factor in determining the precision of results. Both the simulation datasets and the practice data can be used to monitor forecasting (Abedini et al. 2019; Khatun et al. 2021). The field under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to compare the model's performance to fact. The ROC curve is a standard method for analyzing landslide susceptibility maps all over the world (Pradhan and Lee 2010; Xi et al. 2019; Zhang et al. 2019; Mehrabi et al. 2020; Sun et al. 2020). The X and Y axes in ROC reflect the false positive and true positive rates, respectively (Singha et al. 2020; Mallick et al. 2021). It compares the trade-offs between two prices. In ROC, the area under the curve ranges from 0 to 1.0. The higher the AUC ranking, the more robust the forecast or the closer the predicted models and realities are to each other. An AUC value greater than 0.70, according to Talukdar et al. (2021), implies that the projected model and the ground reality are in strong agreement.

Sensitivity analysis of the models

CART-based sensitivity analysis CART stands for classification because regression trees, and it's a supervised machine learning method for classification and prediction that's non-parametric (Nefeslioglu et al. 2010). The CART has become a popular data mining technique due to its effectiveness and ease of use in solving a wide variety of problems in agriculture, economics, engineering, and remote sensing (Steinberg 1995; Choubin et al. 2018). Two types of CART have been used for modelling: classification trees and regression trees. To forecast a single component, classification trees were used, while regression trees were used to estimate a continuous parameter. It uses a step-by-step tool to define splitting laws (Steinberg 1995). A continuous parameter, such as a regression tree, that predicts the value of a dependent parameter based on multiple independent parameters (Fakiola et al. 2010) and a space demarcated by independent parameters depending on dependent parameters (Fakiola et al. 2010). Unlike the classification tree, the regression tree does not yield divisions with dependent parameters. The entire sample has

been partitioned into two or three homogeneous sets, focusing on the most significant splitter in input variables. CART, a decision tree algorithm key's advantage (Breiman et al. 1984) is the existence of cross-validation, which helps to recognise over-fitting issues that will result in poor potential predictions. CART has the added benefit of producing more precise forecasts than other statistical models (Choubin et al. 2018). Regression trees, on the other hand, do not have predefined categories; the end result is a return value for each of the dependent parameter's original values. In the current study, this method was used to measure the weight or force of influence of different parameters in explaining predicted landslide susceptibility models. This is the first research, to the best of the authors' knowledge, to use CART-based sensitivity analysis to classify the model's most sensitive parameters.

RF-based sensitivity analysis As with classification and regression, RF-based sensitivity analysis has recently gained popularity. Many scholars have used it to investigate the impact of variables on the model (Salam et al. 2020). The details of the RF algorithm have been discussed in the methods for LSMs section.

Correlation feature selection-based sensitivity analysis

Feature selection is a machine learning preprocessing phase that is useful in decreasing dimensionality, eliminating unnecessary results, increasing learning precision, and enhancing outcome comprehension (Gopika and ME 2018). The central idea is that an attribute of a subclass is fine if it is strongly associated with the class but not with other class attributes.

Landslide susceptibility mapping using refined parameters

Based on the sensitivity analysis, some of the least sensitive parameters were omitted from the datasets, and the best representative model with the same configuration was applied to the remaining parameters. As a result, LS mapping based on optimal conditioning parameters would be performed. Similarly, validation of the proposed model would be conducted to assess the accuracy of the chosen LS modelling parameters. The whole work was summarized in the methodological flow chart (Fig. 3).

Results

Landslide susceptibility modelling

Several measures have been followed to develop and apply the optimized and LR-based hybrid ensemble ML algorithms for landslide susceptibility mapping, like multicollinearity

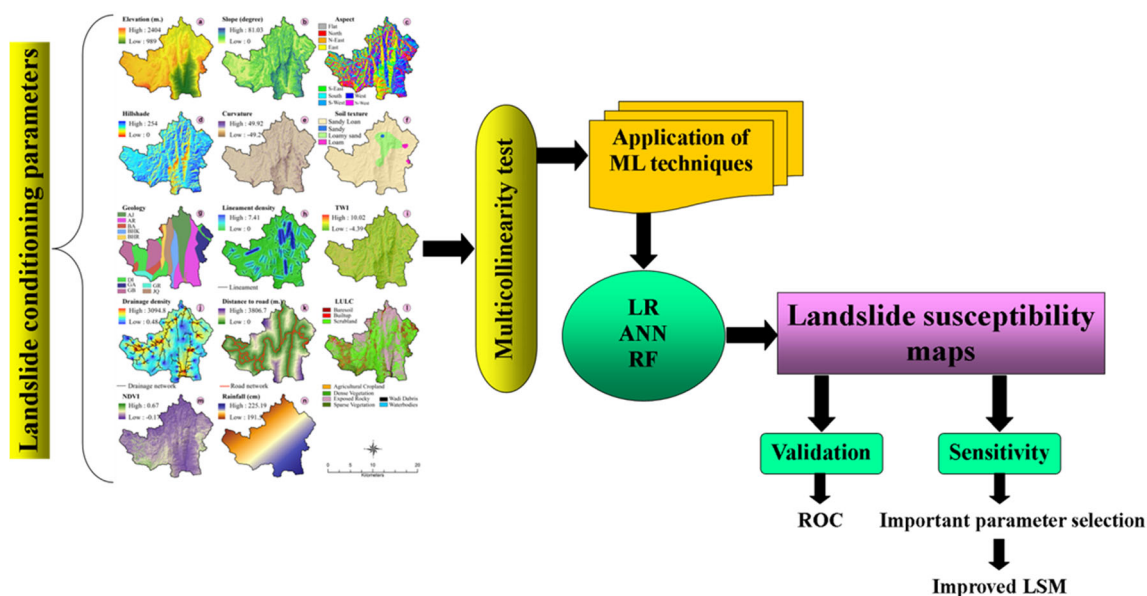


Fig. 3 Methodological flow chart of the whole work

analysis, LSM construction and validation, and finally sensitivity analysis.

Multicollinearity analysis Before the modelling of LS, multicollinearity was tested with the use of TOL and VIF techniques for the landslide causation variables. The LS model performs poorly due to the collinearity of the variables. Variables with TOL values of 0.677 and VIF values of 5.644 is considered to be used during landslide sensitivity modelling processes in this research (Table 1). NDVI has the lowest VIF (1.47), followed by Aspect (VIF=1.513), geology (VIF=1.534), and LULC (VIF=1.682). Elevation received the maximum VIF rating (5.644), led by rainfall (2.708), slope (5.644), and soil texture (2.708) (Table 1).

Application of Logistic regression for making weighted parameters and normalization of parameters Logistic regression model applied to the variables like slope, aspect, hillshade, curvature, TWI, lineament density, etc (Fig. 4) to generate ultimate landslide susceptibility mapping. Some of them have an inverse correlation with landslide susceptibility like aspect, hillshade, curvature and geology while others have a positive relationship. As a result, operating the model swiftly would be cumbersome. Fuzzy membership was used to render the variables unidirectional. A degree of certainty of membership is specified in Fuzzy set theory, with membership values ranging from 0 to 1. Values toward 1 in the fuzzy membership indicates higher susceptible zone and vice versa. In the case of LULC, slope, aspects, and geology, linear fuzzy logic was used, while for the rest of the parameters, Gaussian fuzzy logic was used. Figure 4 depicts very high and high vulnerability zone in south and south-eastern portion while western portion

displays less susceptible zone. The zones are correspondent with fuzzy membership values. The weights derived from LR model as follows:

Elevation * 0.206, Slope * 0.263, Aspect * -0.059, Hillshade * -0.292, Curvature * -1.24, Soil texture * 4.85, Geology * -1.3, Lineament density * 5.13, TWI * 0.006, Drainage density * 5.6, Distance to road * 0.14, LULC * 2.94, NDVI * 5.73, Rainfall * 6.5

Configuring machine learning algorithms The machine learning (ML) algorithms have been first configured by trial and error process. We have tested the models at the 400, 500, 800, and 1000 iterations, so only slight changes and updates were

Table 1 Analysis of collinearity between landslide-causing factors using the multicollinearity test

Parameters	Tolerance	VIF
DEM	0.177	5.644
Slope	0.194	5.163
Aspect	0.661	1.513
Hill shade	0.513	1.948
Curvature	0.478	2.090
Soil texture	0.369	2.708
Geology	0.652	1.534
Lineament density	0.442	2.262
TWI	0.528	1.895
Distance to road	0.443	2.259
Drainage density	0.511	1.958
LULC	0.594	1.682
NDVI	0.677	1.476
Rainfall	0.179	5.583

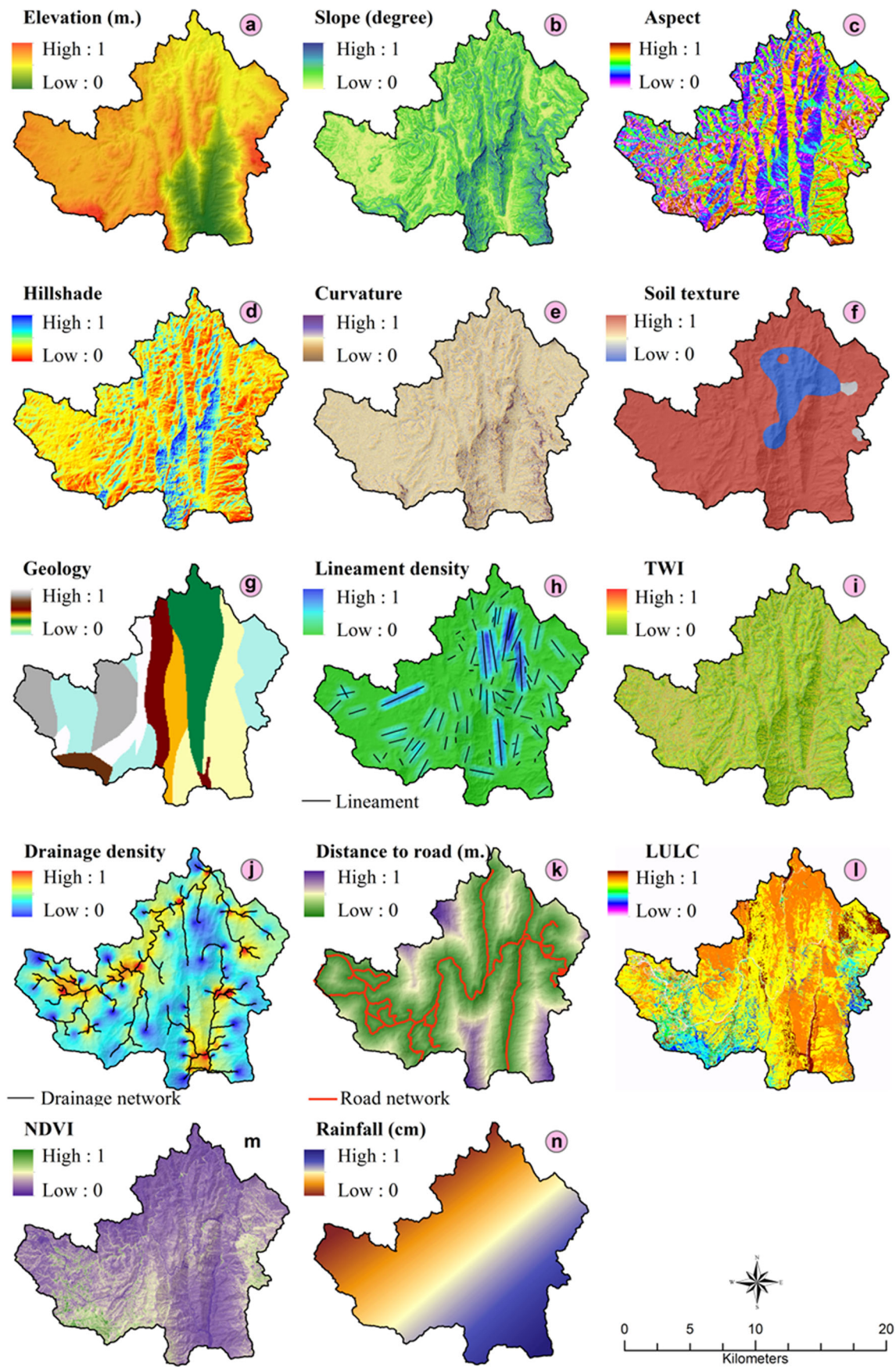


Fig. 4 Standardized landslide conditioning factors: a. elevation, b. slope, c. aspect, d. Hill shade, e. Curvature, f. Soil texture, g. Geology, h. Lineament density, i. TWI, j. Drainage density, k. Distance to road, l. LULC, m. NDVI, n. Rainfall

seen for 1000 iterations. On the other hand, best parameters by trial and error process have been selected for LS modeling. The details of best parameters for RF and ANN have been presented in Table 2.

Hybrid model validation and comparisons LR based hybrid machine learning model's efficiency was tested by comparing actual and expected outcomes of the landslide events. "The correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) were used to test two models." The findings also revealed that all models had higher correlations and lower error values. The LR-RF model was found to be the best (R : 0.971, MAE:0.0195; RMSE: 0.081; RAE: 2.351), followed by the LR-MLP (R : 0.968, MAE: 0.1034; RMSE: 0.1413; RAE: 4.2431) (Table 3).

Construction of the LSMs As seen in Fig. 5, LSMs were produced using two hybrid ML algorithms and individual LR-based model. Because of the continuous dimension or stretching format of the generated LS maps, these values should be divided into different categories. A variety of classifier systems are available for converting continuous maps into categorized maps, including natural break, equal interval, quantile, and standard deviation. Researchers propose the natural break classifier as an indicator that is robust, accurate, and consistent (Baeza et al. 2016). Jenks optimization is a data clustering approach used to identify the optimum value structure for the various classes, and is often referred to as the natural break classificatory. Landslide susceptibility maps were graded into five subclasses using Jenks' natural break algorithm: very high, high, moderate, low, and very low. The aim of this approach is to minimise the variance from the average of the other classes while reducing the mean deviation from the average class value. The method thus eliminates variance in intra-class while maximising variance in inter-classes. The LR-RF model calculated the very high and high LS zones to be 32.73 km² and 21.84 km², respectively, with moderate (47.53 km²), low (39.83 km²), and very low (55.15 km²) LS zones (Table 4).

Validation of LS maps Three landslide susceptibility models were validated using the testing datasets. The efficiency of

Table 3 Error measures for hybrid models

Error measures	LR-MLP	LR-RF
Correlation coefficient	0.968	0.971
Mean absolute error	0.1034	0.0195
Root mean squared error	0.1413	0.081
Relative absolute error	4.2431	2.351
Root relative squared error	5.2065	3.482

three LS models was assessed using the ROC curve. The AUC value of the ROC curve reflects the reliability or compatibility of the LS model with the real world. According to the AUC value of the ROC curve, the LR-RF model (AUC=0.941), led by LR-MLP (AUC=0.915), and LR (AUC=0.872), are the most representative model. Effective analysis of the landslide vulnerability was carried out by all models. Depending on AUC, the LR-RF>LR-MLP>LR could be reorganised in LS models.

Sensitivity analysis of LS model

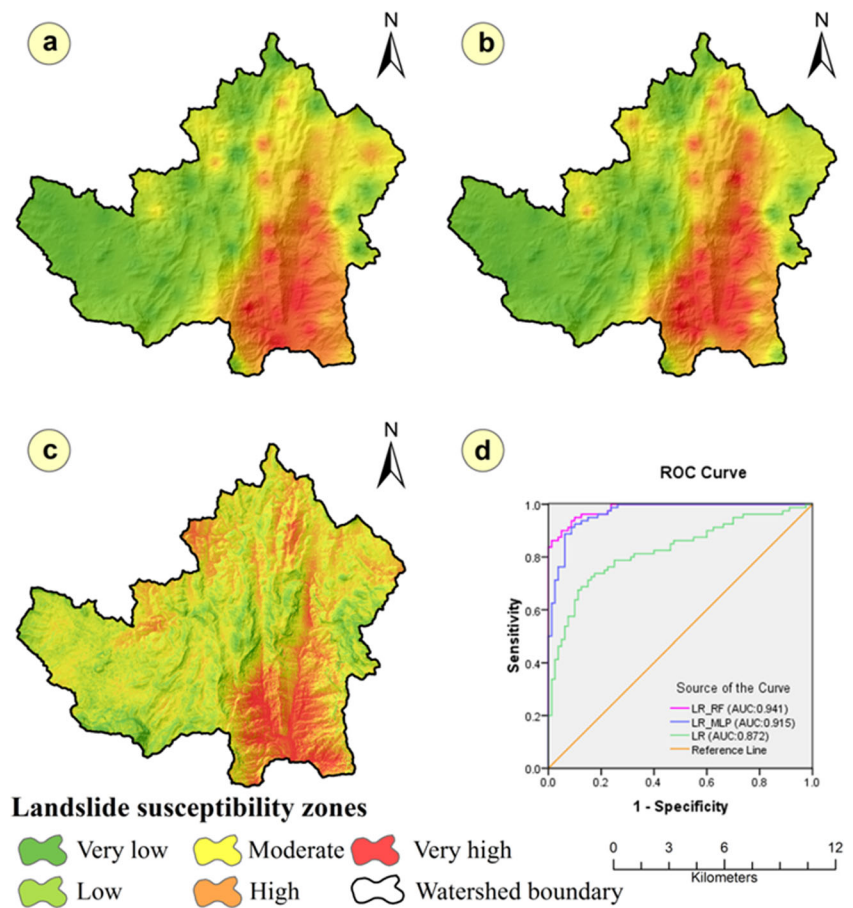
The implementation of improved ensemble machine learning algorithms for LS mapping will only demonstrate the probability of a landslide occurring in the future based on the dynamic statistical relationship between historical landslide patterns and their triggering variables. Neither of these models, however, addresses the role of some variables in the occurrence of a landslide. The concern arises as to how management policies can be planned and implemented if the effect of such factors on the occurrence of landslides cannot be determined. The discovery of landslide-influencing factors will aid in reducing the occurrence and danger of landslides. Three models, RF, CART and correlation feature selection were used to assess the parameters that are most sensitive for landslides (Fig. 6).

Rainfall, soil texture, and DEM are highly sensitive parameters, while aspect, curvature, and other low-sensitivity parameters as per RF model are designated (Fig. 6a). DEM and lineament have been deemed the most sensitive parameters in CART modelling, although aspect, and curvature have been deemed less sensitive (Fig. 6b). DEM, rainfall, and

Table 2 Machine learning algorithms by setting model parameters

ML algorithms	Optimized model parameters
ANN	Batch size-100, Hidden Layer: 14, Learning rate: 0.5, Momentum: 0.3, Seed: 5, validation set size: 10, Validation threshold: 30
RF	Seed: 5, Number of features: 1, Number of iterations: 200, $2mtry$ (the number of variables tested at each node), 300 trees

Fig. 5 Landslide susceptibility modeling using two hybrid models and one statistical models, such as a) LR-MLP, b) LR-RF, c) LR, and d) ROC



lineament are considered more resilient than other variables by using correlation feature selection method (Fig. 6c).

Landslide susceptibility mapping using optimum conditioning parameters

The efficiency of the RF model is the best of the three models, indicating that it is ideally fit for landslide susceptibility measurement, as the AUC under this model is 0.941. The selected model was used to determine landslide susceptibility zones

Table 4 Computation of areas under different landslide susceptible zones

Landslide susceptible zone	Area (km ²)		
	LR-MLP	LR-RF	LR
Very Low	55.15	43.75	13.54
Low	39.83	41.21	48.53
Moderate	47.53	49.07	67.55
High	21.84	33.38	45.15
Very High	32.73	29.67	22.34

over the study region by excluding all less sensitive parameters, such as aspect, hillshade, drainage density, curvature, and TWI, as shown in Fig. 6. The generated LSM using LR-RF and optimum conditioning factors has been classified into five classes like previous models. The natural break algorithm was used to classify the LSM model. The ROC curve was then used to measure the model's accuracy and compare it to three hybrid models to see how much the optimal parameters influenced the model's performance. This model's AUC is 0.927, which is not higher than the best model (LR-RF) (0.941), but it clearly produces reliable results and achieves a high degree of similarity without such variables. As with the best model (LR-RF), the most vulnerable zone was discovered in the south and south-eastern portions of the study region, while the least vulnerable zone was discovered in the western part (Fig. 7). So, even though five factors are omitted, the outcome does not alter dramatically. However, the optimum parameter based model predicted very high LS zone as 20.46km², followed by high (41.87km²), moderate (72.17km²), low (34.02km²), and very low (28.56km²)

However, the generated final LSM model showed very precise landslide susceptibility zones in the study area (Fig. 7). It also shows the spatial distribution of the road over the landslide susceptibility models. It indicates that past landslides

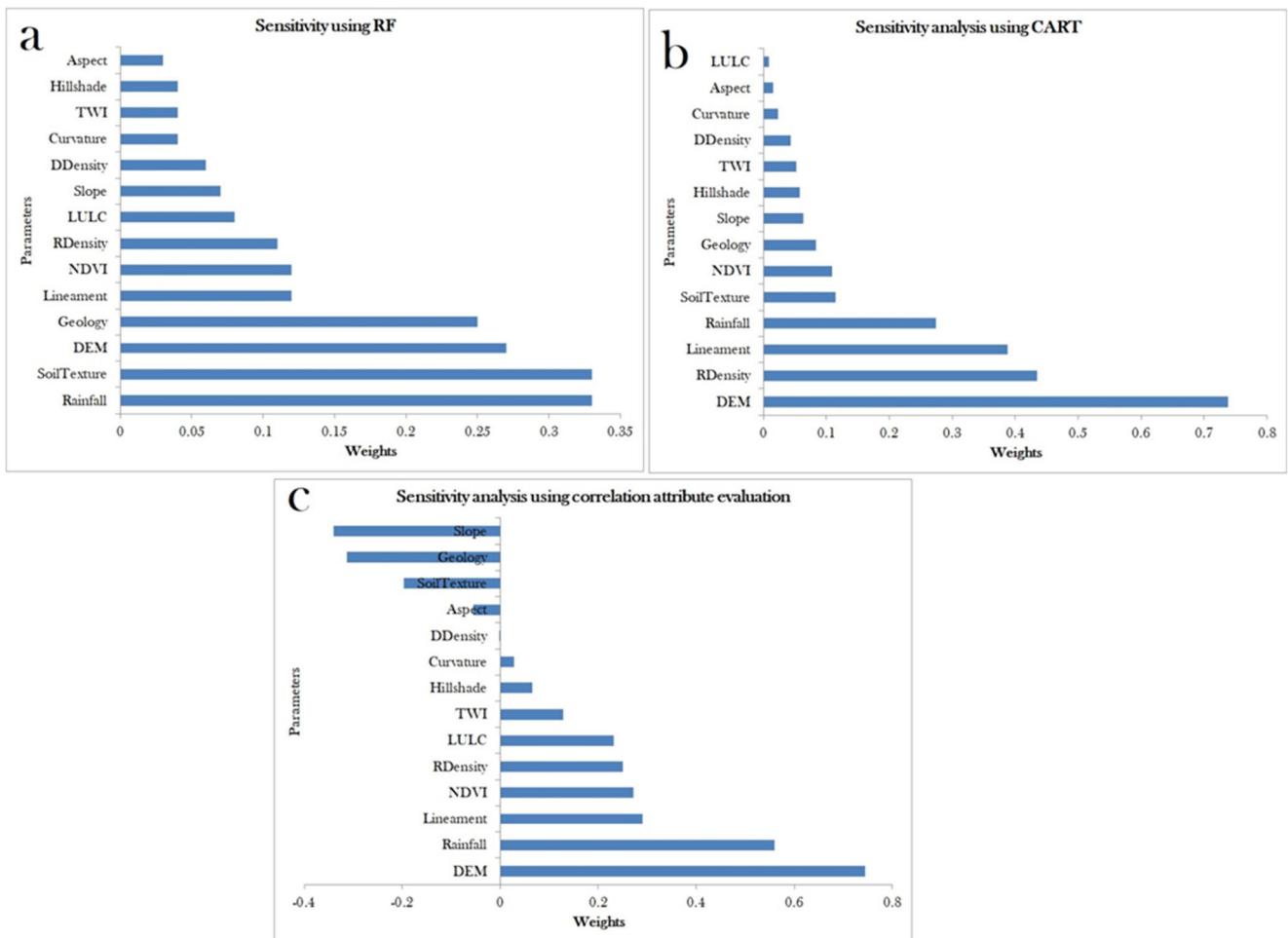


Fig. 6 Sensitivity analysis for best LS model (LR-RF) using (a) RF, (b) CART, and (c) correlation feature selection

have been observed along the road networks of study area’s southern part. While, the model also shows that the very high and high landslide susceptibility zones have been predicted

along the same regions, which indicates the construction of road along the higher elevation and high slope areas hampered the equilibrium conditions. As a result, the landslides have

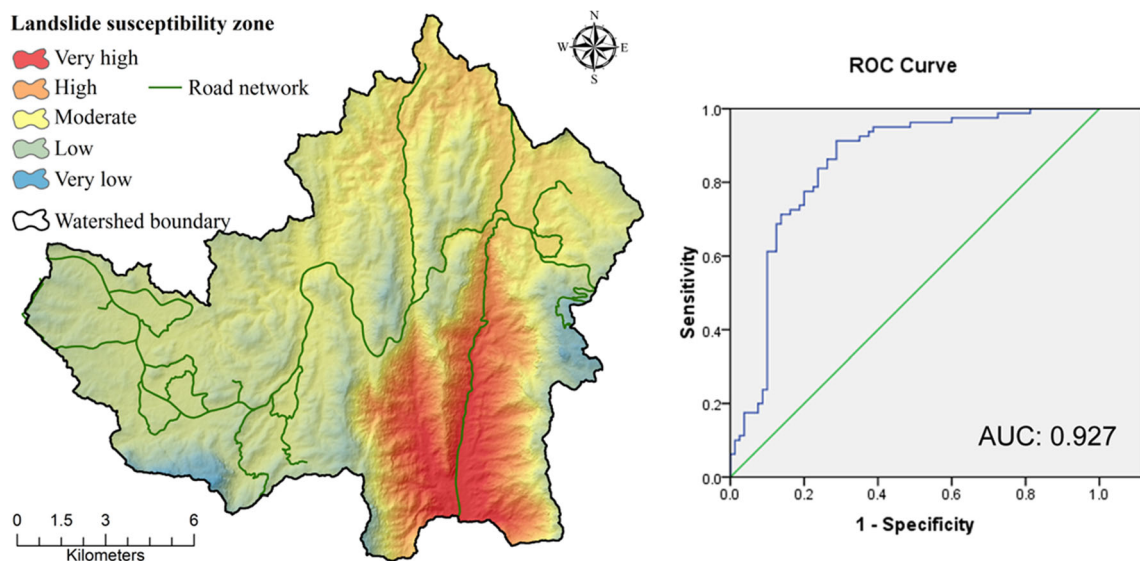


Fig. 7 Landslide susceptibility mapping based on optimum conditioning factors using LR-RF model with ROC curve for validation

been occurred and in future, these regions would have the higher chances to observe landslide. Therefore, it should be necessary to examine the geological stability of any regions before proceeding for construction, otherwise it will cause a natural hazard.

Discussion

Landslide vulnerability is a term that describes the likelihood of a landslide occurring in a given area, as well as the relationship between past landslide locations and potential conditioning factors (Romer and Ferentinou 2016). Before proceed for landslide susceptibility modelling, the preparation of the conditioning variables and landslide inventory is an essential task. A landslide inventory map is a first step toward evaluating landslide susceptibility, danger, and probability (Rosi et al. 2018). Landslide inventories are divided into two main categories: landslide-event inventories linked to a cause and historical landslide inventories (Rosi et al. 2018). We used the latter formation in this analysis, which was the product of several landslide events over a long period of time. Many smaller landslides, on the other hand, could have been abandoned due to different degrees of alteration by subsequent landslides, erosional cycles, vegetation formation, and human impacts (Rosi et al. 2018). As a result, using multi-temporal high-resolution satellite imagery to interpret smaller landslides can be a useful complement to the existing landslide inventory and efficient for enhancing the accuracies of landslide susceptibility charts. Sixteen conditioning variables were chosen based on current literature and multicollinearity analysis: slope angle, slope aspect, height, plan curvature, profile curvature, TWI, STI, SPI, distance to rivers, distance to bridges, distance to faults, NDVI, climate, land use, lithology, and rainfall. Furthermore, the relationships between landslide incidence and these variables were examined using the sensitivity analysis.

However, many approaches have been used in connection with the advancement of GIS technologies to forecast the spatial distributions of landslides in previous decades, including sophisticated machine learning models (Pourghasemi and Kerle 2016; Youssef et al. 2016) and classical regression analysis (Ding et al. 2017; Zhang et al. 2017). In general, statistical models take longer to analyse inputs, outputs, and spatial analysis, but MLTs have the benefit of automatically detecting connections between dependent and independent variables (Mallick et al. 2021). The choice of the optimal model among several machine learning approaches, according to Felicísimo et al. (2013), is the most significant element that determines the accuracy of the landslide susceptibility mapping. Many publications have used various machine learning models for landslide susceptibility mapping

(Mallick et al. 2021; Zhang et al. 2017; Zhu et al. 2018; Youssef et al. 2016; Xie et al. 2017; Xi et al. 2019; Wang et al. 2018; Xu et al. 2012), however the accuracy of some of these approaches is still dubious. In landslide susceptibility evaluation, choosing the optimal model among several MLTs is critical (Mallick et al. 2021; Felicísimo et al. 2013). MLTs have a number of advantages, including the fact that they are very simple to use and that their prediction accuracy generally outperforms some of the more traditional approaches, such as the analytical hierarchy process (Tien Bui et al. 2012; Pourghasemi and Rahmati 2018). Both bivariate and machine learning methods, on the other hand, have constraints that could be tackled by using ensemble models. As a result, new ensemble approaches and techniques in the context of landslide simulation must be explored and compared. Shirzadi et al. (2017) used a Naive Bayes trees (NBT) and random subspace (RS) ensemble system for landslide susceptibility mapping in the Bijar region of Kurdistan province (Iran), and their results showed that NBT-RS greatly boosted the accuracy of the NBT base classifier. Hong et al. (2018) discovered that the J48 Decision Tree with the Rotation Forest model has the best prediction capability (AUC =0.855), greatly improving the accuracy of the J48 Decision Tree base classifier. Pham et al. (2019) combined the MultiBoost (MB) ensemble and support vector machine (SVM) models to model the vulnerability of landslides in Uttarakhand State, Northern India, and found that the MBSVM outperforms the LR and single SVM models. Despite the fact that multiple ensemble methods have been used in landslide susceptibility mapping, there is still no consensus on which is the best ensemble approach. Furthermore, further experiments are needed to compare different areas in order to determine the difference between each system. In this research, we used a cutting-edge advanced machine learning hybrid technique called LR and its ensembles (RF and ANN) to forecast landslides in the Asir Region of Saudi Arabia.

Several authors stated that there are certain concerns about landslide susceptibility assessments, which are described follows: (1) Lee et al. (2017) assumed that these models cannot be utilised and were not valid for particular planning and assessment purposes, thus utilising these models for specific site characterization should be approached with caution and the size of the study should be carefully evaluated; (2) The variables utilised to build the landslide susceptibility models are important. It was discovered that some models, such as RF, performed well in some regions but poorly in others (Catani et al. 2013). (Hong et al. 2016a, 2016b).

Manual, specified interval, natural split, equivalent interval, quantile, standard deviation, geometrical interval, and landslide percentage are some of the classification techniques

for a landslide susceptibility map in GIS software (Baeza et al. 2016). In general, the reader would find it more difficult to understand and explain user-defined classification. As a result, instead of using a user-defined classification, existing automatic classification schemes can be used (Baeza et al. 2016). Furthermore, since landslide vulnerability indices have positive or negative skewness, quantile or normal split are the safest classification methods (Akgun et al. 2012). According to the histogram of data distribution, the natural break process, which is the most widely used form (Termeh et al. 2019), is the most appropriate method for modelling landslide susceptibility in the current research.

A comparative analysis of the LR, LR-ANN, and LR-RF models was carried out in this article. LR is a popular classification model, especially for binary classification problems (Kleinbaum and Klein 2010). As a result, we combined the ANN and RF models with the LR model to create a better classifier. The logistic regression and random forest approaches are combined in the LR-RF model. It has been shown that RF is one of the most widely used classification algorithms and that it can increase the efficiency of single classifiers (Genuer and Poggi 2020). Furthermore, RF will reduce the WoE model's reliance on freedom among the conditioning variables. As a result, the findings revealed that both the LR-RF model (AUC = 0.773 for training data) and the RF model (AUC = 0.802 for training data) would improve the efficiency of the standard LR model (AUC = 0.720 for training data), with the LR-RF model outscoring the others.

Landslide susceptibility maps can assist in the development of a critical guide for general planning and evaluation (Lee et al. 2017). For landslide control, hazard and risk assessments, these approaches are extremely valuable and practical (Zhu et al. 2018). As a result, understanding the distinctions between various MLTs is critical for selecting the best model for a certain study goal and/or location (Lee et al. 2017). Our findings, for example, suggested that these MLTs models might be critical as a preliminary approach for decision makers to identify risky regions in existing projects and potential new planning areas. So that potentially dangerous locations may be discovered and investigated further. Furthermore, the findings of this study revealed that all MLTs produce LSMs that perform considerably better.

However, optimum parameters selected by sensitivity analysis have further been used to implement for landslide susceptibility mapping using best model like LR-RF. The performance of the model is also quite close the best model (AUC: 0.927), indicating very high performance. Many researchers have proposed using many parameters for any type of potentiality mapping of natural hazards, but our research indicates that using many parameters will result in a less reliable model due to redundancy. As a result, when modelling some kind of predictive algorithm, it is important to examine the impact of the parameters.

Some critical parameters can produce a very accurate model, while redundant parameters can produce a less accurate model. When modeling, redundant parameters should be tested to save time and resources.

Conclusion

The current study provides a comprehensive understanding of the development of hybrid machine learning algorithms for LS mapping. According to the two hybrid and one statistical LS models, the very large LS zone occupies an area of 20–32.62km². The ROC curve was used to measure the LS models. The best representation model for LS modelling was LR-RF (AUC=0.958), followed by LR-ANN and LR. Five least sensitive parameters were defined and omitted from datasets based on sensitivity analysis. Again, the LR-RF model was applied to the remaining parameters to produce optimum conditioning parameter-based LS models with very high accuracy (AUC: 0.927). It indicates that there is no need for many parameters when modelling LSM, but essential parameters are needed for modelling in order to generate very high quality LSM maps. As a result, identifying very important parameters should be necessary and relevant.

The current study lays the comprehensive foundation for the introduction of two hybrids ML and one LR model for producing LSM, which could be used to predict various naturally occurring hazards such as flooding, deforestation, and fire susceptibility. Other preferred algorithms, such as the recurrent neural network, convolution neural network, recursive neural network, and other future popular algorithms, can be used to expand the applications of the ANN method. Over and above the conditioning variables, deep neural networks can provide extra insights. Though higher-level machine learning techniques take longer to compute, the results obtained from these new generation models are well worth the effort.

The current research has some drawbacks, such as the use of moderate resolution satellite images rather than hyperspectral images, the use of a semi-quantitative and conventional approach, the use of less rainfall gauge stations in the rainfall layer, and the use of fewer landslide sites for modelling rather than more evidence.

Nonetheless, our research has shed new light on the development of hybrid machine learning algorithms for LS mapping with selecting optimal conditioning variables. The findings will be beneficial to decision-makers and policymakers in the study area for the protection of natural ecosystems to reduce damage, live losses, and maintain habitat quality for biodiversity conservation.

Acknowledgements The authors thankfully acknowledge the Deanship of Scientific Research, King Khalid University for providing administrative and financial supports.

Author contribution Conceptualization, Saeed Alqadhi, Javed Mallick, Swapan Talukdar; Data curation, Ahmed Ali Bindajam, TamalKantiSaha; Formal analysis, Saeed Alqadhi, Javed Mallick, Swapan Talukdar and Nguyen Van Hong; Funding acquisition, Saeed Alqadhi; Methodology, Javed Mallick and Swapan Talukdar, Saeed Alqadhi, Nguyen Van Hong; Project administration, Saeed Alqadhi, Javed Mallick; Resources, Saeed Alqadhi, TamalKantiSaha; Software, Javed Mallick and Swapan Talukdar; Supervision, Javed Mallick; Validation: Swapan Talukdar and Javed Mallick; Writing – original draft, Javed Mallick Swapan Talukdar and Saeed Alqadhi; Writing – review & editing, Javed Mallick, Nguyen Van Hong.

Funding Funding for this research was given under award numbers R.G.P2/75/41 by the Deanship of Scientific Research; King Khalid University, Ministry of Education, Kingdom of Saudi Arabia.

Data availability The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

- Abdulkadir TS, Muhammad RUM, Wan Yusof K, Ahmad MH, Aremu SA, Gohari A, Abdurrahman AS (2019) Quantitative analysis of soil erosion causative factors for susceptibility assessment in a complex watershed. *Cogent Eng* 6. <https://doi.org/10.1080/23311916.2019.1594506>
- Abedini M, Ghasemian B, Shirzadi A, Bui DT (2019) A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling. *Environ Earth Sci* 78: 560. <https://doi.org/10.1007/s12665-019-8562-z>
- Agarwal S, Kachroo P, Regentova E (2016) A hybrid model using logistic regression and wavelet transformation to detect traffic incidents. *IATSS Res* 40:56–63. <https://doi.org/10.1016/j.iatssr.2016.06.001>
- Agostini A, Tofani V, Nolesini T, Gigli G, Tanteri L, Rosi A, Cardellini S, Casagli N (2014) A new appraisal of the Ancona landslide based on geotechnical investigations and stability modelling. *Q J Eng Geol Hydrogeol* 47:29–44. <https://doi.org/10.1144/qjegh2013-028>
- Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci* 38:23–34. <https://doi.org/10.1016/j.cageo.2011.04.012>
- Alvioli M, Melillo M, Guzzetti F, Rossi M, Palazzi E, von Hardenberg J, Brunetti MT, Peruccacci S (2018) Implications of climate change on landslide hazard in Central Italy. *Sci Total Environ* 630:1528–1543. <https://doi.org/10.1016/j.scitotenv.2018.02.315>
- Amone E, Francipane A, Noto LV, Scarbaci A, la Loggia G (2014) Strategies investigation in using artificial neural network for landslide susceptibility mapping: Application to a Sicilian catchment. *J Hydroinf* 16:502–515. <https://doi.org/10.2166/hydro.2013.191>
- Baeza C, Lantada N, Amorim S (2016) Statistical and spatial analysis of landslide susceptibility maps with different classification systems. *Environ Earth Sci* 75:1–17. <https://doi.org/10.1007/s12665-016-6124-1>
- Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. *Hydrol Sci Bull* 24:43–69. <https://doi.org/10.1080/02626667909491834>
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and Regression Trees - 1st Edition. In: CRC Press. Taylor Fr. Gr. <https://www.routledge.com/Classification-and-Regression-Trees/Breiman-Friedman-Stone-Olshen/p/book/9780412048418>. Accessed 11 May 2021
- Bui DT, Lofman O, Revhaug I, Dick O (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat Hazards* 59:1413–1444. <https://doi.org/10.1007/s11069-011-9844-2>
- Bui DT, Pradhan B, Revhaug I, Nguyen DB, Pham HV, Bui QN (2015) A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). *Geomatics, Nat Hazards Risk* 6:243–271. <https://doi.org/10.1080/19475705.2013.843206>
- Bui DT, Shahabi H, Shirzadi A et al (2018) Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron Highlands, Malaysia. *Remote Sens* 10:1527. <https://doi.org/10.3390/rs10101527>
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazards Earth Syst Sci* 13(11):2815–2831
- Chang JM, Chen H, Jou BJD, Tsou NC, Lin GW (2017) Characteristics of rainfall intensity, duration, and kinetic energy for landslide triggering in Taiwan. *Eng Geol* 231:81–87. <https://doi.org/10.1016/j.enggeo.2017.10.006>
- Chen W, Li W, Chai H, Hou E, Li X, Ding X (2016) GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji City, China. *Environ Earth Sci* 75:1–14. <https://doi.org/10.1007/s12665-015-4795-7>
- Chen W, Panahi M, Pourghasemi HR (2017a) Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *Catena* 157:310–324. <https://doi.org/10.1016/j.catena.2017.05.034>
- Chen W, Pourghasemi HR, Naghibi SA (2018a) 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* 77:647–664. <https://doi.org/10.1007/s10064-017-1010-y>
- Chen W, Pourghasemi HR, Panahi M, Kornejady A, Wang J, Xie X, Cao S (2017b) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* 297:69–85. <https://doi.org/10.1016/j.geomorph.2017.09.007>
- Chen W, Pourghasemi HR, Zhao Z (2017c) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int* 32:367–385. <https://doi.org/10.1080/10106049.2016.1140824>
- Chen W, Sun Z, Han J (2019) Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl Sci* 9:171. <https://doi.org/10.3390/app9010171>
- Chen W, Xie X, Peng J, Shahabi H, Hong H, Bui DT, Duan Z, Li S, Zhu AX (2018b) GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based

- random forest method. *Catena* 164:135–149. <https://doi.org/10.1016/j.catena.2018.01.012>
- Choubin B, Darabi H, Rahmati O, Sajedi-Hosseini F, Kløve B (2018) River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Sci Total Environ* 615:272–281. <https://doi.org/10.1016/j.scitotenv.2017.09.293>
- CRED (2016) Annual Disaster Statistical Review 2015 | Centre for Research on the Epidemiology of Disasters. CRED, In <https://www.cred.be/annual-disaster-statistical-review-2015-0>.
- Crosta GB, Imposimato S, Roddeman DG (2003) Numerical modelling of large landslides stability and runout. *Nat Hazards Earth Syst Sci* 3:523–538. <https://doi.org/10.5194/nhess-3-523-2003>
- Cruden DM (1991) A simple definition of a landslide. *Bull Int Assoc Eng Geol - Bull l'Association Int Géologie l'Ingénieur* 43:27–29. <https://doi.org/10.1007/BF02590167>
- Dehnavi A, Aghdam IN, Pradhan B, Morshed Varzandeh MH (2015) A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena* 135:122–148. <https://doi.org/10.1016/j.catena.2015.07.020>
- Ding Q, Chen W, Hong H (2017) Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. *Geocarto Int* 32:619–639. <https://doi.org/10.1080/10106049.2016.1165294>
- EM-DAT (2013) EM-DAT | The international disasters database. <https://www.emdat.be/>. Accessed 9 May 2021
- Fakiola M, Mishra A, Rai M, Singh SP, O'Leary RA, Ball S, Francis RW, Firth MJ, Radford BT, Miller EN, Sundar S, Blackwell JM (2010) Classification and regression tree and spatial analyses reveal geographic heterogeneity in genome wide linkage study of Indian visceral Leishmaniasis. *PLoS One* 5:e15807. <https://doi.org/10.1371/journal.pone.0015807>
- Forkuor G, Hounkpatin OKL, Welp G, Thiel M (2017) High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. *PLoS One* 12:e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- Genuer R, Poggi J-M (2020) Introduction to Random Forests with R. pp 1–8
- Genuer R, Poggi JM, Tuleau-Malot C (2010) Variable selection using random forests. *Pattern Recogn Lett* 31:2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Gopika N, ME AMK (2018) Correlation based feature selection algorithm for machine learning, 3rd International Conference on Communication and Electronics Systems (ICCES). In: IEEE. pp 692–695
- Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72:272–299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- He S, Pan P, Dai L, Wang H, Liu J (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>
- Hong H, Naghibi SA, Pourghasemi HR, Pradhan B (2016a) GIS-based landslide spatial modeling in Ganzhou City, China. *Arab J Geosci* 9(2):112
- Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB (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>
- Hong H, Pourghasemi HR, Pourtaghi ZS (2016b) Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 259:105–118. <https://doi.org/10.1016/j.geomorph.2016.02.012>
- Hong H, Pradhan B, Xu C, Tien Bui D (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* 133:266–281. <https://doi.org/10.1016/j.catena.2015.05.019>
- Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front* 12(3): 101075
- Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int J Environ Sci Technol* 11:909–926. <https://doi.org/10.1007/s13762-013-0464-0>
- Jebur MN, Pradhan B, Shafri HZM, Yusoff ZM, Tehrani MS (2015) An integrated user-friendly ArcMAP tool for bivariate statistical modelling in geoscience applications. *Geosci Model Dev* 8:881–891. <https://doi.org/10.5194/gmd-8-881-2015>
- Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* (80-.). 349:255–260
- Kanungo DP, Arora MK, Sarkar S, Gupta RP (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng Geol* 85:347–366. <https://doi.org/10.1016/j.enggeo.2006.03.004>
- Khatun R, Talukdar S, Pal S, Saha TK, Mahato S, Debanshi S and Mandal I (2021) Integrating remote sensing with swarm intelligence and artificial intelligence for modelling wetland habitat vulnerability in pursuance of damming. *Ecological Informatics*, p.101349.
- Kim MS, Onda Y, Kim JK, Kim SW (2015) Effect of topography and soil parameterisation representing soil thicknesses on shallow landslide modelling. *Quat Int* 384:91–106. <https://doi.org/10.1016/j.quaint.2015.03.057>
- Kleinbaum DG, Klein M (2010) Ordinal Logistic Regression. pp 463–488
- Lee S, Dan NT (2005) Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. *Environ Geol* 48:778–787. <https://doi.org/10.1007/s00254-005-0019-x>
- Lee S, Hong SM, Jung HS (2017) A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability* 9(1):48
- Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ Geol* 40:1095–1113. <https://doi.org/10.1007/s002540100310>
- Li Y, Chen W (2020) Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. *Water (Switzerland)* 12:113. <https://doi.org/10.3390/w12010113>
- Liucci L, Melelli L, Suteanu C, Ponziani F (2017) The role of topography in the scaling distribution of landslide areas: A cellular automata modeling approach. *Geomorphology* 290:236–249. <https://doi.org/10.1016/j.geomorph.2017.04.017>
- Luo W, Liu CC (2018) Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* 15:465–474. <https://doi.org/10.1007/s10346-017-0893-9>
- Mandal S, Mandal K (2018) Bivariate statistical index for landslide susceptibility mapping in the Rorachu river basin of eastern Sikkim Himalaya, India. *Spat Inf Res* 26:59–75. <https://doi.org/10.1007/s41324-017-0156-9>

- Marjanović M, Kovačević M, Bajat B, Voženilek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. *Eng Geol* 123:225–234. <https://doi.org/10.1016/j.enggeo.2011.09.006>
- Mahato, S., Pal, S., Talukdar, S., Saha, T.K. and Mandal, P., 2021. Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models. *Geoscience Frontiers*, 12(5), p.101175.
- Mallick, J., Alqadhi, S., Talukdar, S., AlSubih, M., Ahmed, M., Khan, R.A., Kahla, N.B. and Abutayeh, S.M., 2021. Risk assessment of resources exposed to rainfall induced landslide with the development of GIS and RS based ensemble metaheuristic machine learning algorithms. *Sustainability*, 13(2), p.457.
- Mehrabi M, Pradhan B, Moayedi H, Alamri A (2020) Optimizing an adaptive neuro-fuzzy inference system for spatial prediction of landslide susceptibility using four state-of-the-art metaheuristic techniques. *Sensors (Switzerland)* 20:1723. <https://doi.org/10.3390/s20061723>
- Mind'je R, Li L, Nsengiyumva JB, Mupenzi C, Nyeshaja EM, Kayumba PM, Gasirabo A, Hakorimana E (2020) Landslide susceptibility and influencing factors analysis in Rwanda. *Environ Dev Sustain* 22: 7985–8012. <https://doi.org/10.1007/s10668-019-00557-4>
- Mukherjee I, Singh UK (2020) Delineation of groundwater potential zones in a drought-prone semi-arid region of east India using GIS and analytical hierarchical process techniques. *Catena* 194:104681. <https://doi.org/10.1016/j.catena.2020.104681>
- Nefeslioglu HA, Gokceoglu C, Sezer E et al (2010) Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul. *Turkey Math Probl Eng* 2010:1–15. <https://doi.org/10.1155/2010/901095>
- Nhu VH, Mohammadi A, Shahabi H, Ahmad BB, al-Ansari N, Shirzadi A, Clague JJ, Jaafari A, Chen W, Nguyen H (2020) Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. *Int J Environ Res Public Health* 17:1–23. <https://doi.org/10.3390/ijerph17144933>
- Pal S, Saha TK (2017) Exploring drainage/relief-scape sub-units in Atreyee river basin of India and Bangladesh. *Spat Inf Res* 25(5): 685–692
- Peng D, Xu Q, Liu F, He Y, Zhang S, Qi X, Zhao K, Zhang X (2018) Distribution and failure modes of the landslides in Heitai terrace, China. *Eng Geol* 236:97–110. <https://doi.org/10.1016/j.enggeo.2017.09.016>
- Peng L, Niu R, Huang B, Wu X, Zhao Y, Ye R (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology* 204:287–301. <https://doi.org/10.1016/j.geomorph.2013.08.013>
- Peres DJ, Cancelliere A (2018) Modeling impacts of climate change on return period of landslide triggering. *J Hydrol* 567:420–434. <https://doi.org/10.1016/j.jhydrol.2018.10.036>
- Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Tran TTT, Bui DT (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 175:203–218. <https://doi.org/10.1016/j.catena.2018.12.018>
- Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ Earth Sci* 75:1–17. <https://doi.org/10.1007/s12665-015-4950-1>
- Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* 162:177–192. <https://doi.org/10.1016/j.catena.2017.11.022>
- Pourghasemi HR, Razavi-Termeh SV, Kariminejad N, Hong H, Chen W (2020) An assessment of metaheuristic approaches for flood assessment. *J Hydrol* 582:124536. <https://doi.org/10.1016/j.jhydrol.2019.124536>
- Pradhan AMS, Kim YT (2017) Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea. *Bull Eng Geol Environ* 76: 1263–1279. <https://doi.org/10.1007/s10064-016-0919-x>
- Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Softw* 25:747–759. <https://doi.org/10.1016/j.envsoft.2009.10.016>
- Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A (2014) Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab J Geosci* 7:725–742. <https://doi.org/10.1007/s12517-012-0807-z>
- Romer C, Ferentinou M (2016) Shallow landslide susceptibility assessment in a semi-arid environment - A Quaternary catchment of KwaZulu-Natal, South Africa. *Eng Geol* 201:29–44. <https://doi.org/10.1016/j.enggeo.2015.12.013>
- Rosi A, Tofani V, Tanteri L, Tacconi Stefanelli C, Agostini A, Catani F, Casagli N (2018) The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. *Landslides* 15:5–19. <https://doi.org/10.1007/s10346-017-0861-4>
- Saha TK, Pal S (2019) Exploring physical wetland vulnerability of Atreyee river basin in India and Bangladesh using logistic regression and fuzzy logic approaches. *Ecol Indic* 98:251–265
- Saha TK, Pal S, Sarkar R (2021) Prediction of wetland area and depth using linear regression model and artificial neural network based cellular automata. *Ecological Informatics* 62:101272
- Salam R, Islam ARMT, Pham QB, Dehghani M, al-Ansari N, Linh NTT (2020) The optimal alternative for quantifying reference evapotranspiration in climatic sub-regions of Bangladesh. *Sci Rep* 10:20171. <https://doi.org/10.1038/s41598-020-77183-y>
- Segoni S, Rosi A, Lagomarsino D, Fanti R, Casagli N (2018) Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Nat Hazards Earth Syst Sci* 18:807–812. <https://doi.org/10.5194/nhess-18-807-2018>
- Shirzadi A, Bui DT, Pham BT, Solaimani K, Chapi K, Kavian A, Shahabi H, Revhaug I (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci* 76: 1–18. <https://doi.org/10.1007/s12665-016-6374-y>
- Sidle RC, Ochiai H (2000) *Landslides: Processes, Prediction, and Land Use*. Adger, W. N
- Singha P, Das P, Talukdar S, Pal S (2020) Modeling livelihood vulnerability in erosion and flooding induced river island in Ganges riparian corridor, India. *Ecol Indic* 119:106825. <https://doi.org/10.1016/j.ecolind.2020.106825>
- Stamatopoulos CA, Di B (2015) Analytical and approximate expressions predicting post-failure landslide displacement using the multi-block model and energy methods. *Landslides* 12:1207–1213. <https://doi.org/10.1007/s10346-015-0638-6>
- Steinberg D and PC (1995) *Tree-Structured Non-Parametric Data Analysis Classification and Regression Trees* by Salford Systems. San Diego, CA Salford Syst
- Stumpf A, Kerle N (2011) Object-oriented mapping of landslides using Random Forests. *Remote Sens Environ* 115:2564–2577. <https://doi.org/10.1016/j.rse.2011.05.013>
- Sun D, Wen H, Wang D, Xu J (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* 362:107201. <https://doi.org/10.1016/j.geomorph.2020.107201>
- Tang Y, Feng F, Guo Z, Feng W, Li Z, Wang J, Sun Q, Ma H, Li Y (2020) Integrating principal component analysis with statistically-based models for analysis of causal factors and landslide

- susceptibility mapping: A comparative study from the loess plateau area in Shanxi (China). *J Clean Prod* 277:124159
- Talukdar S, Ghose B, Shahfahad et al (2020) Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stoch Env Res Risk A* 34:2277–2300. <https://doi.org/10.1007/s00477-020-01862-5>
- Talukdar S, Pal S, Singha P (2021) Proposing artificial intelligence based livelihood vulnerability index in river islands. *J Clean Prod* 284:124707. <https://doi.org/10.1016/j.jclepro.2020.124707>
- Termeh SVR, Khosravi K, Sartaj M, Keesstra SD, Tsai FTC, Dijkstra R, Pham BT (2019) Optimization of an adaptive neuro-fuzzy inference system for groundwater potential mapping. *Hydrogeol J* 27:2511–2534. <https://doi.org/10.1007/s10040-019-02017-9>
- Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput Geosci* 45:199–211. <https://doi.org/10.1016/j.cageo.2011.10.031>
- Tien Bui D, Pradhan B, Nampak H, Bui QT, Tran QA, Nguyen QP (2016) Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *J Hydrol* 540:317–330. <https://doi.org/10.1016/j.jhydrol.2016.06.027>
- Tien Bui D, Pradhan B, Revhaug I, Trung Tran C (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. In: *Society of Earth Scientists Series*. Springer, Cham, pp 87–111
- Truong XL, Mitamura M, Kono Y, Raghavan V, Yonezawa G, Truong X, Do T, Tien Bui D, Lee S (2018) Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl Sci* 8:1046. <https://doi.org/10.3390/app8071046>
- Tsangaratos P, Ilia I (2016a) Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* 13:305–320. <https://doi.org/10.1007/s10346-015-0565-6>
- Tsangaratos P, Ilia I (2016b) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* 145:164–179. <https://doi.org/10.1016/j.catena.2016.06.004>
- Wang T, Wu SR, Shi JS, Xin P, Wu LZ (2018) Assessment of the effects of historical strong earthquakes on large-scale landslide groupings in the Wei River midstream. *Eng Geol* 235:11–19. <https://doi.org/10.1016/j.enggeo.2018.01.020>
- Xi W, Li G, Moayed H, Nguyen H (2019) A particle-based optimization of artificial neural network for earthquake-induced landslide assessment in Ludian county, China. *Geomatics, Nat Hazards Risk* 10:1750–1771. <https://doi.org/10.1080/19475705.2019.1615005>
- Xie Z, Chen G, Meng X, Zhang Y, Qiao L, Tan L (2017) A comparative study of landslide susceptibility mapping using weight of evidence, logistic regression and support vector machine and evaluated by SBAS-InSAR monitoring: Zhouqu to Wudu segment in Bailong River Basin, China. *Environ Earth Sci* 76:1–19. <https://doi.org/10.1007/s12665-017-6640-7>
- Xu C, Xu X, Lee YH, Tan X, Yu G, Dai F (2012) The 2010 Yushu earthquake triggered landslide hazard mapping using GIS and weight of evidence modeling. *Environ Earth Sci* 66:1603–1616. <https://doi.org/10.1007/s12665-012-1624-0>
- Youssef, A.M. and Pourghasemi, H.R., (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geoscience Frontiers*, 12(2), pp.639–655
- Youssef AM, Pourghasemi HR, El-Haddad BA, Dhahry BK (2016) Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. *Bull Eng Geol Environ* 75:63–87. <https://doi.org/10.1007/s10064-015-0734-9>
- Zhang Y, Guo L, Chen Y, Shi T, Luo M, Ju QL, Zhang H, Wang S (2019) Prediction of soil organic carbon based on Landsat 8 monthly NDVI data for the Jiangnan Plain in Hubei Province, China. *Remote Sens* 11:1683. <https://doi.org/10.3390/rs11141683>
- Zhang Y s, Yang Z h, Guo C b et al (2017) Predicting landslide scenes under potential earthquake scenarios in the Xianshuihe fault zone, Southwest China. *J Mt Sci* 14:1262–1278. <https://doi.org/10.1007/s11629-017-4363-6>
- Zhu AX, Miao Y, Wang R, Zhu T, Deng Y, Liu J, Yang L, Qin CZ, Hong H (2018) A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *Catena* 166:317–327. <https://doi.org/10.1016/j.catena.2018.04.003>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.