Geosystems and Geoenvironment 2 (2023) 100158

Contents lists available at ScienceDirect

Geosystems and Geoenvironment

journal homepage: www.elsevier.com/locate/geogeo

Modelling and predicting of landslide in Western Arunachal Himalaya, India

Soumik Saha[a], Biswajit Bera[a,*], Pravat Kumar Shit[b], Debashish Sengupta[c], Sumana Bhattacharjee[d], Nairita Sengupta[e], Paromita Majumdar[f], Partha Pratim Adhikary[g]

[a] Department of Geography, Sidho-Kanho-Birsha University, Ranchi Road, P.O. Purulia Sainik School, Purulia 723104, India
[b] PG Department of Geography, Raja Narendralal Khan Women's College (Autonomous), Vidyasagar University, Midnapore 721102, India
[c] Department of Geology and Geophysics, Indian Institute of Technology (IIT), Kharagpur, West Bengal 721302, India
[d] Department of Geography, Jogesh Chandra Chaudhuri College (University of Calcutta), 30, Prince Anwar Shah Road, Kolkata 700033, India
[e] Department of Geography, Diamond Harbour Women's University, Sarisha 743368, India
[f] Department of Geography, Vidyasagar College for Women, 39 Sankar Ghosh Lane, Kolkata 700006, India
[g] ICAR Indian Institute Water Management, Bhubaneswar, Odisha 751023, India

ARTICLE INFO

Article history:
Received 6 July 2022
Revised 14 September 2022
Accepted 3 December 2022

Handling Editor: Vinod Samuel

Keywords:
Landslide susceptibility
Machine learning models
Neo-tectonic activities
Himalayan terrains

ABSTRACT

Landslides are the indicator of slope instability particularly in mountain terrain and causing different types of reimbursements and threats of life and property. The Himalayan terrains are highly susceptible to different natural hazards as well as disasters particularly land failure activities mainly due to inherent tectonic activities which further enhanced by various Neo-tectonic and Neolithic activities. This scientific study provides an enhanced framework for the assessment of proper and precise landslide susceptibility in the two districts of Arunachal Pradesh (Tawang and West Kameng) considering both physical and anthropogenic factors and various machine learning models (SVM, AdaBoost and XGBoost). At first, landslide inventory maps were developed based on previous landslide events. Here, 70% of the data were randomly selected for training and remaining was used for validation and optimization of the models using statistical implications and validation assessment methods. The result showed that the high and very high landslide susceptible areas are mainly concentrated in the middle portion along the Bhalukpong-Bomdia road section. Based on the AUC value and other statistical indicators it has been observed that AdaBoost is the most efficient model here (AUC = 0.92). AUC values of SVM and XGBoost are 0.85 and 0.89 respectively. AdaBoost model identifies that very low susceptibility class occupies 60.22% area and very high landslide susceptibility class occupies 15.51% area and it will be considered as more encouraging method for landslide susceptibility determination in this kind of cases for better accurateness. This high accuracy susceptibility map positively helps during the execution of various developmental projects.

1. Introduction

Landslide is one of the most distressing and never ending geo-environmental hazards which impose serious threats to the human life and economy (Petley, 2012). Landslide describes the down ward movement of masses including rocks and minerals, soils, organic materials etc. under the gravitational force and it is a result of the interaction between driving force and resistance force (Bera, 2008). Landslides play a significant role in geomorphological evolution of landscape and act as a common disastrous geo-environmental hazard in hilly or mountainous parts of the world and causing billion of economic losses along with thousands of injuries and causalities in each year (Galli et al., 2008). Basically, landslide is the most common geological or geomorphic hazard which is associated with economic loss (100 billion USD globally) and severe effect over natural and social environment. Fluid pressure and stress tensions come from various hydrological parameters such as changes of precipitation, infiltration, infiltration, runoff, dehydration reaction, and wastewater and thus fluid-injection can easily exaggerate different forms of sliding including stable sliding (fault slips) and unstable sliding (earthquake and rainfall induced landslides) (Iverson and major, 1987; Terzaghi, 1950; Handwerger et al., 2013; Guglielmi et al., 2015;

* Corresponding author.
E-mail address: biswajitbera00007@gmail.com (B. Bera).

https://doi.org/10.1016/j.geogeo.2022.100158

Tanyas et al., 2022). Types of sliding are classified by the rate-and-state friction models and laboratory based critical state soil mechanics (Schulz et al., 2009; Guglielmi et al., 2015). Generally, landslides are triggered by various geological, geomorphological, hydrological and anthropogenic activities and a single triggering factor can make hundreds slope failures. In 2015, the Gorkha earthquake affected the central Nepal region (more than 35000 km^2 area) and triggered around 25000 landslides in this Himalayan terrain (Roback et al., 2017). Information regarding this type of landslides and their spatial distribution is very much important for developmental planning and execution of projects, understanding the landform evolution, hazard zonation and sometimes it assists to predict the landslide probability (Tanyas et al., 2019; Galli et al., 2008). Landslide erosion quantification is very much essential to understand the relationship among physical erosion processes, chemical weathering and consumption (Gabet, 2007). Steep hill slopes in mountain areas are removed by the landslides that are driven by hydro-geomorphological and geological conditions compiled with river and glacial erosions (Whipple et al., 1999). Climate change is another secondary factor for landslides at hilly areas or high altitude regions due to degradation of permafrost and incessant melting of glaciers (Fischer et al., 2006). The timing and velocity of gravity driven landslides is controlled by the internal compositions and properties of the sliding materials (Schulz et al., 2009). Surface roughness dating and geochronological techniques are being used to determine the exact age of landslide masses and their intensity. Glade (2003) revised the evidences from New-Zealand regarding human induced landslide probabilities and its responses with land use change. The primary stage of forecasting of different geological hazards is to identify the predictor's nature and character along with geomorphological, geological and extreme climatic conditions (Kumar and Anbalagan, 2016). Road construction in hilly or mountainous terrain is an archetypical example of human induced tension on the earth surface. Constructions of road network, tunnel, bridge and culvert in hilly areas are always associated with blasting and excavation that makes the total region unstable and effectively decrease the strength and cohesive power of the hill slope as well as soil (Zhao et al., 2018). Advance remote sensing imageries can detect landslide locations, distribution and extension of the land failures (Xu et al., 2019). Recently, a wide range of landslide inventory methods or approaches has been developed and these are manually extracted of landslide areas on the basis of specialist's visual opinion, image classification approaches, multi-temporal SAR interferometry techniques, applying optical or LiDAR data from unmanned aerial vehicles and semi-automated image classification methods which are based on machine learning algorithms (Ghorbanzadeh et al., 2019). Approximately 95% landslide incidents in the entire globe were identified from developing and underdeveloped countries (Chung and Fabbri, 2003). In India, around 15% areas are landslide prone and Himalayan terrain is one of the extreme landslide prone areas (Rawat and Joshi, 2012; Bera, 2007a). Inadequate investigation of geotechnical and geological aspects unfavourably affects the regional planning and development of existing geo-environmental condition and it leads the hill slope instability (Anbalagan et al., 2008). Slope stability analysis is an essential method for the safely execution of various environmental engineering projects worldwide. In mountainous areas (like Himalaya), slopes are excavated at the time of road construction (Komadja et al., 2021). In hilly tract, it is essential to analyse the parameters that are responsible for slope instability. Generally, geology, geomorphology, human activity, climate etc. are the reasonable factors behind the slope instability (Komadja et al., 2020). Mineralogical and the geotechnical characteristics of debris material showed the particle sizes that are formed from weathering and erosion. The engineering geotechnical parameters of slope materials decline the stability of slope (Summa et al., 2018). This Himalayan range is characterised by the convergence boundary between Indian plate and Eurasia plate associated with active thrusts, lineament and faults and such neo-tectonic activities accelerate ideal land failure conditions (Sati et al., 2007). During rainy season the excess water cannot release as a surface runoff while it infiltrates through the lineaments, surface cracks, joints, cleavage, bedding and porous soils. As a result, the driving force significantly increases and ultimately landslide occurs. Generally, in mountain areas where erosion rate is greater and the rate of weathering will be kinetically limited (Dixon et al., 2012). In tectonically active mountain belts, erosions are mainly driven by bedrock land sliding and the rate depends on intense rainfall and seismicity (Dadson et al., 2003; Bera, 2009). Bedrock land sliding can promote the slow percolation of surface runoff in the exceedingly fragmented rock or soil debris and develops unfavourable condition of weathering (Shou and Chen, 2021). Exposed bed rock materials by land sliding erosion indorse effective chemical weathering. Landslides are termed as important erosion agent in mountainous region or upland landscape areas that are commonly characterised by threshold hill slopes, high-relief and relatively narrow river gorges (Montgomery and Brandon, 2002; Bera, 2008). Landslides produce massive amount of debris into river bed and dammed the river channel naturally. Korup and Montgomery (Korup and Montgomery, 2008) identified frequent glacial and debris damming in the eastern Himalayan syntax region that created ideal condition of head ward erosion of major river channels and dissected the edge of the Tibetan plateau (Korup and Montgomery, 2008). In the last decade, different ML models were employed for landslide susceptibility modelling studies such as frequency ratio (Kumar and Anbalagan, 2016); Logistic regression (Lee, 2005; Lombardo and Mai, 2018); artificial neural networks (Chen et al., 2017a); linear discriminant analysis (Pham and Prakash, 2019); neuro-fuzzy (Chen et al., 2017a); support vector machines (Xu et al., 2012); decision tree (Wu et al., 2020), classification and regression tree (Chen et al., 2017b); Bivariate and multivariate statistical analysis (Nandi and Shakoor, 2010). In this scientific study, support vector machine (SVM), extreme gradient boosting (XGBoost) and Adaptive Boosting (AdaBoost) have been considered for landslide susceptibility modelling. SVM is a widely used method for landslide susceptibility modelling that is based on non-linear transformation of the covariates in a high dimensional space and every class is linearly separable (Ballabio and Sterlacchini, 2012). AdaBoost is a homogeneous ensemble framework and significantly amplifies the performance of the prediction model (Pham et al., 2017). The main objective of the study is to identify the appropriate landslide susceptible zones in the two westernmost districts of Arunachal Pradesh (Tawang and West Kameng). Such scientific study will definitely assist to administrators and policy makers during the execution of different developmental projects like road construction, tunnel structure, building and bridge. Moreover, limitation of the study was ground based data unavailability like rock and soil strength. There are also some limitations of machine learning model based studies. Machine learning requires huge amount of data set for training and these are highly susceptible to errors (Malik, 2020).

2. Study area

Arunachal Himalaya is the eastern most sector of Himalayan mountain range with an extension of 91°30′ E to 96°E and 26°28′ N to 29°30′ N and it includes the eastern Himalayan syntaxis (Yin et al., 2006; Fig. 1). The study area has been restricted within the political boundary of Tawang and West Kameng districts of Arunachal Pradesh (Fig. 1). The study area has been divided into different tectonomorphic zones (Shiwalik Himalaya, Lesser Himalaya, Greater Himalaya, Trans Himalaya) that make

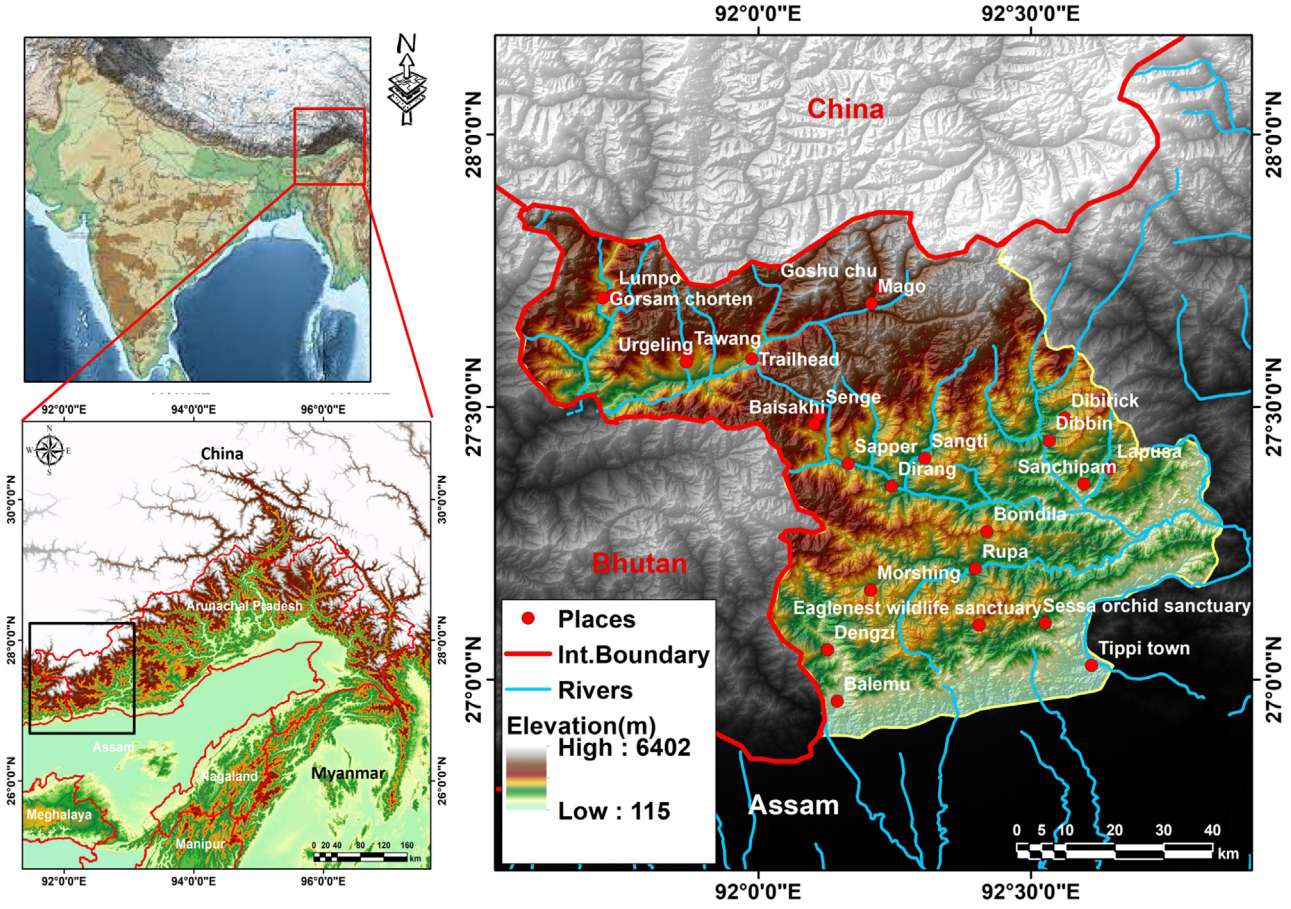

Fig. 1. Geographical location of the study area.

this area more landslide prone (Singh et al., 2014). Physiographically, this region is characterised with high altitude, steep slope, gorge, dissected valleys and ridge with mountain summits. Annual rainfall and temperature of this Himalayan tract vary from 150 cm-200 cm and 15°C-30°C respectively. The climatic characteristics of Arunachal Pradesh influence by the geographical conditions and this region comes under very high seismic zone (Zone V) due to its geographic location and tectonic activities. According to GSI guidelines, landslides are classified into three types i.e. very large active slide or more than one slide, one active landslide and dormant landslide. Dormant landslides are those where no active movements are observed. It can be identified only through the previous imprints. Active landslides are easily identified through the freshness of landslide materials in the disturbed part of the hilly area and displacement masses of landslides more than 100 cubic meters are classified into large landslide (Singh et al., 2014). Singh et al. (2014) investigated the number of landslide occurrences along the important road section of the study area (Bhalukpong-Bomdila-Tawang) and they identified 4 large active landslides, 21 active landslides and 62 dormant landslides.

2.1. Geology of Arunachal Himalayas

Arunachal Pradesh has high geological complexity and the entire region exhibits three different mountain system i.e. (i) Himalayan range (ii) Naga-Patkai-Arakan range and iii. Mishmi hills. Past geological enquiry of this area was totally carried out through field survey but structural and geochronological analyses were missing (Yin, 2006). The study area (Tawang and West Kameng district) is the westernmost part of Arunachal Pradesh that is abruptly raised from Assam plain. These two districts consist with sedimentary rocks of the Siwalik group and the MBT (main boundary thrust) divides this region from the lesser Himalaya (Fig. 2). The Lesser Himalayan range is characterised by various rocks of different geological ages and separated by thrust from one another (Singh et al., 2014). This stretch occupies by the crystalline formation of Se La group that is comprised with Kyanite-sillimanite bearing garnet-biotite schist, Psammatic gneiss, streaky gneiss, tourmaline bearing leucogranite and amphibolites. It is associated with steep hill slope and escarpment which belongs from Archaean and Palaeozoic geological age (Mishra, 2007; Fig. 2). Bomdila group is another important existing rock system in this region that is characterised by argillaceous, arenaceous, metamorphites and penecontemporaneous basic volcanic rocks and it is also categorised into three key geological formations (Bhushan et al., 1991). Dedza Formation consists of carbonate rocks along with minor quantity of black slates while Dirang formation comprises with garnet-muscovite-biotite schist, sericite quartzite, tourmaline marble and carbonate phyllite (Srivastava et al., 2011). The gneissic and schist formations of this region consist with different folds, symmetric and asymmetric structures, fault scars, shear plains and different linear and non-linear structures. Mesoscopic shear zones have been identified in this area over schists and migmatitic gneisses (Srivastava et al., 2011). Here, the Gondwana super group has been superimposed on the quartzite-carbonate-phyllite sequence that is termed as “Buxa Group” (Singh et al., 2014). A sudden change of geomorphic character has been observed along the local and regional discontinuities. Topography of Tawang section is characterized by very steep slopes with glacial geomorphic features. The

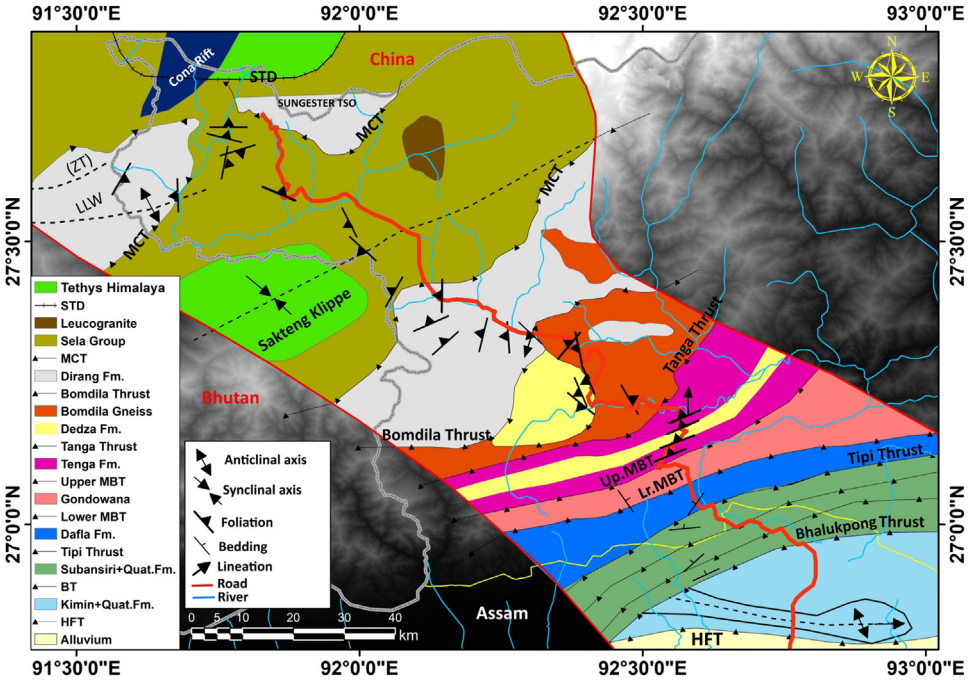

Fig. 2. Geological map showing the geological framework of the western Arunachal Pradesh.

sudden geomorphic change in this region attributed with faults and thrusts (Srivastava et al., 2011). The discontinuity between Se La and Tawang sector is known as Tawang thrust that is roughly parallel with the main central thrust (MCT). Another important thrust has been traced apart from Tawang thrust based on satellite observation in the south of Se La pass that separates the Se La sector with Senge sector. Se La thrust has also been detected on the basis of highly presence of Knick points and sudden refraction of drainage system that comes from high altitudinal areas (Kumar, 1997).

3. Methods and materials

3.1. Information gain ratio (InGR)

Information gain ratio (InGR) is a principal method for analysis of the implication of the influential factors or the parameters of a predicted model. The information gain ratio statistic can significantly select the features of the used prediction model that is accomplished to identify the high ranking parameters for the susceptibility prediction studies or researches (Bui et al., 2020). This statistical method gives the InGR values for the each causative factor to evaluate the applicability. High InGR value signifies main factor whereas the low InGR value shows low significant factor. The InGR method has been applied here to find out the usefulness of the used factors using the equation.

$$Gain\ ratio\ (x,z) = \frac{Entropy(z) - \sum_1^n \sum_{i=1}^n \frac{|Z_i|}{|Z|}\ Emtropy\ (Z_i)}{-\sum_{i=1}^n \frac{|Z_i|}{|Z|} \log \frac{|Z_i|}{|Z|}} \quad (1)$$

Here, $Z\ with\ Z_i 1 = 1, 2, 3 \ldots\ldots n$ subsets, the attribute x is belonging.

For selection of the appropriate factor in the case of predictive modelling, Johnston et al. (2018) have given the proper way to find out the existing multi-collinearity problem in the causative factors . The VIF (variance inflation factors) value greater than 10 and Tolerance value less than 0.1 specify the multi-collinearity problem (Johnston et al., 2018; Table 2). The Tolerance value and VIF value can compute by the following equation.

$$Tolerance = 1 - R^2 J \quad (2)$$

$$VIF = \left(\frac{1}{Tolerance}\right) \quad (3)$$

Here, $R^2 J$ also indicates coefficient of determination of the regression equation.

3.2. Landslide conditioning factors

To achieve better and accurate landslide susceptibility model, selection of proper landslide conditioning factors is a vital step (Table 1). For this research, total 16 conditioning factors have been selected based on previous literatures, related studies and also field investigation (geology, slope, aspect, profile curvature, plain curvature, lineament, elevation, soil, distance from road, distance from river, rainfall, NDVI, NDWI, distance from epicentre, drainage density, LULC; Figs. 3 and 4).

3.2.1. Topographical factors

DEM (Digital Elevation Model) is a basic element to analyse the nature of the terrain of a particular area and it is also used for precise prediction of landslide susceptibility. The characteristics of DEM have direct impact over geomorphic factors, attitude and magnitude of slope, surface characteristics (elevated and undulated surfaces), convexity and concavity (Burrough, 1988). Variation of relief can deliberately influence the intensity and degree of landslide (Pradhan, 2017a). In this research, the relief of the study area ranges from 115 m to 6402 m and plays most important controller of the occurrence of landslide hazards. Slope is another principal factor of landslide and land instability. Generally, steep slope condition can decrease the soil strength. In this research, slope ranges from 0 to 75.70 degree that indicates more potential energy available for mass movement. Terrain curvature is another geomorphic character that can regulate the soil erosion process through the divergence and convergence process of water flow (Erener and Düzgün, 2010; Bera, 2010). Previous studies

Table 1
Database (source, type and method) of the landslide conditioning factors used in this research.

Conditioning factors	Source	Variable type	Classification method
Altitude	Digital Elevation Model (ASTER DEM) (https://earthdata.nasa.gov/learn/articles/new-aster-gdem)	Continuous (DEM based)	Natural breaks
Slope		Continuous (DEM based)	Natural breaks
Aspect		Continuous (DEM based)	Azimuth classification
Plain curvature		Continuous (DEM based)	Natural breaks
Profile curvature		Continuous (DEM based)	Natural breaks
Lineament density		Continuous (DEM based)	Natural breaks
River density		Continuous (DEM based)	Natural breaks
Distance from river		Continuous (DEM based)	Manual
Soil	Food and Agricultural organization (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/)	Discrete (Digital image)	Soil classification
NDVI (Normalized Difference Vegetation Index)	Landsat 8 (OLI_TIRS) (https://earthexplorer.usgs.gov/)	Continuous (Satellite image based)	Natural breaks
NDWI (Normalized Difference Water Index)		Continuous (Satellite image based)	Natural breaks
LULC (Land use and Land Cover)		Discrete (Satellite image based)	Supervised classification
Rainfall	NASA power access (https://power.larc.nasa.gov/data-access-viewer/)	Continuous (Other factor)	Natural breaks
Distance from epi-centre	Geological Survey of India (https://www.gsi.gov.in/)	Continuous (Other factor)	Manual
Geology		Discrete (Digital image based)	Geological classification
Distance from road	Google earth pro	Continuous (Digital image based)	Manual

vividly showed that there was a high correlation between landslide and different geo-environmental factors (those influence the landslide process; Regmi et al., 2010). Different slope aspect associated factors such as, solar radiation, wind condition, rainfall condition; biological conditions can play as a triggering factor of the process of landslide (Sevgen et al., 2019).

3.2.2. Geological factors

Lithology, soil group, lineament and distance from earthquake epicentre have been taken as geological factors in this research. Lithological variations of a region have a significant impact over different types of geo-environmental hazards. The rock structure, strength, porosity and permeability also control the process of landslide (Erener et al., 2017; Bera, 2007a). Processes of Landslide are characterised by individual lithological characteristics because each unit of lithology reflects its own response to weathering and erosion or jointly the process of denudation (Duna et al., 2018). Prominent lineament including joints, faults, fractures also make rocks more fragile. The river networks follow the lineament direction and accelerate toe erosion as well as process of landslide (Pradhan and Kim, 2020). The landslide incidents are greatly controlled by soil and rock strength, porosity and permeability and cohesiveness (Althuwaynee et al., 2014).

3.2.3. Hydrological factors

Drainage density, distance from river, rainfall and NDWI are selected as hydrological factors in this research. Various hydrological factors play a vital role in the occurrence of landslides particularly in the Himalayan terrain. The drainage proximity map has been generated through Euclidean distance tool in GIS platform. Drainage density means the total length of drainage within the pixel. Higher density of drainage specifies greater surface flow and low percolation (Pradhan and Kim, 2020).Water flow can lead the transportation of sediment and reduces the strength of slope by toe erosion which allows the process of landslide in hilly region (Highland and Bobrowsk, 2008). High drainage density accelerates rapid soil and land erodibility (Pradhan et al., 2017b; Bera, 2007b). Rainfall induced landslides are the natural phenomena in the hilly terrain areas. Rainfall can extensively change the hydraulic pressure, strength of soil mass and lithology that exaggerates the downward movement of rock and debris along the slope due to the gravitational action (Wang et al., 2020). Higher magnitude of rainfall is also associated with high intensity of landslide. The quantitative analysis of rainfall threshold value considered as a landslide influencer. Landslide may occur when the rainfall crosses its certain amount of threshold value (Gariano et al., 2019; Bera, 2007b).

3.2.4. Land related factors

Distance from road, NDVI and land use land cover has been taken as land related factors here. Every area has its own land use land cover characteristics. Hilly region, bare land, road and settlement region are stated as landslide prone in past studies (Ozdemir and Altural, 2013). In this research land use land cover map has been produced using Landsat 8 satellite images along with seven distinct classes (open land, vegetation, scrub, waterbody, settlement, bare land and snow cover area). Reduction of vegetation cover (due to human activities in a regional scale) is a significant controller of the occurrence of shallow landslides (Glade, 2003). Human activities are also influential for the landslide occurrence because they always affect the land use land cover of that particular area (Al-Najjar et al., 2021). Vegetation always plays a fundamental role in stabilizing the water flow and increases the resistance force along with soil cohesion property (Sidle and Ochiai, 2006). The root system of trees acts as anchor and improves the soil strength along with shear resistance, at the same time it can reduce the moisture of soil in a certain context (Pandey et al., 2019; Bera, 2007a). Cut down of trees and excavation of soil is the common activities during the construction of roads or any developmental activities which are highly associated with landslide (Al-Najjar et al., 2021).

3.3. Application of machine learning models

3.3.1. Extreme Gradient Boosting (XGBoost)

Decision tree based methods are the cutting-edge techniques in the case of small and medium range set of data (Al-Najjar et al., 2021). The XGBoost was developed based on gradient boosting concept that is very much regularized and can overcome the overfitting problem. Gradient boosting is a boosting algorithm that is mainly used to minimize the biasness or errors in the machine learning model and it is an ensemble machine learning

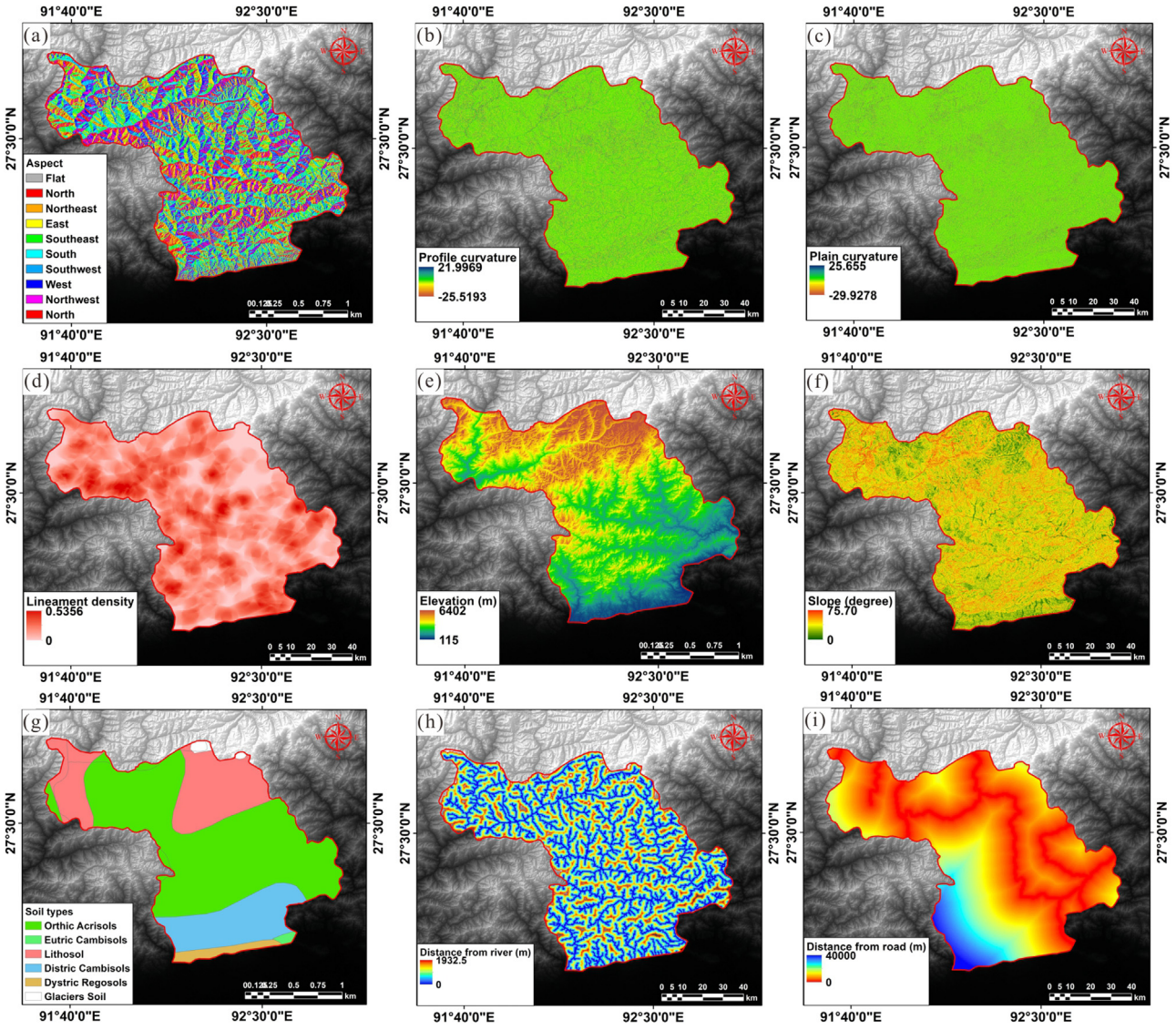

Fig. 3. Landslide conditioning factors (a) Aspect, (b) profile curvature, (c) plain curvature, (d) lineament density, (e) elevation, (f) slope, (g) soil type, (h) distance from river, (i) distance from road.

technique for the regression and classification (Dada et al., 2019). This recently developed tree based method (XGB) has gained a great significance in the recent model based susceptibility studies (Sonobe et al., 2017). This method was established in 2016 as a research project associated with high prediction ability because the weightage of this method is based on unexpected errors from the training process. XGB method has a great advantage compared with other parallel methods because it provides best usage and low computational cost (Georganos et al., 2017; Liu et al., 2018). XGBoost is an optimized extension of gradient boosting. The XGBoost model can minimize the regularized objective function that is computed using this formula.

$$L(\phi) = \sum_{i} l(\widehat{y}_i y_i) + \sum_{k} \Omega(f_k) \tag{4}$$

Here, the first term is the loss functions whereas y_i and $\widehat{y_i}$ are the target and predicted variables respectively.

For reduction the over fitting and complexity problem, the below mention equation has been used,

$$\Omega(f) = \gamma \mathrm{T} + \frac{1}{2}\lambda ||w||^2 \tag{5}$$

Here, T signifies the no. of leaves in the tree and w indicates the score of each leaf.

The objective function has been minimized using the given equation,

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \widehat{y_i}) + f_t(x_i)) + \Omega(f_t) \tag{6}$$

In case of accelerate the optimization process, the second order Taylor expansion has been used in the objective.

After eliminating the constant, a simple objective function has been developed applying this formula.

$$\widetilde{L^{(t)}} = \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{7}$$

Several algorithms are available in case of XGB modelling for accurate split efficiency (weighted quartile sketch, approximate, sparsity-aware split finding etc.).

3.3.2. Support Vector Machine (SVM)

Support vector machine is a kernel based method which is considered in supervised machine learning algorithm consider-

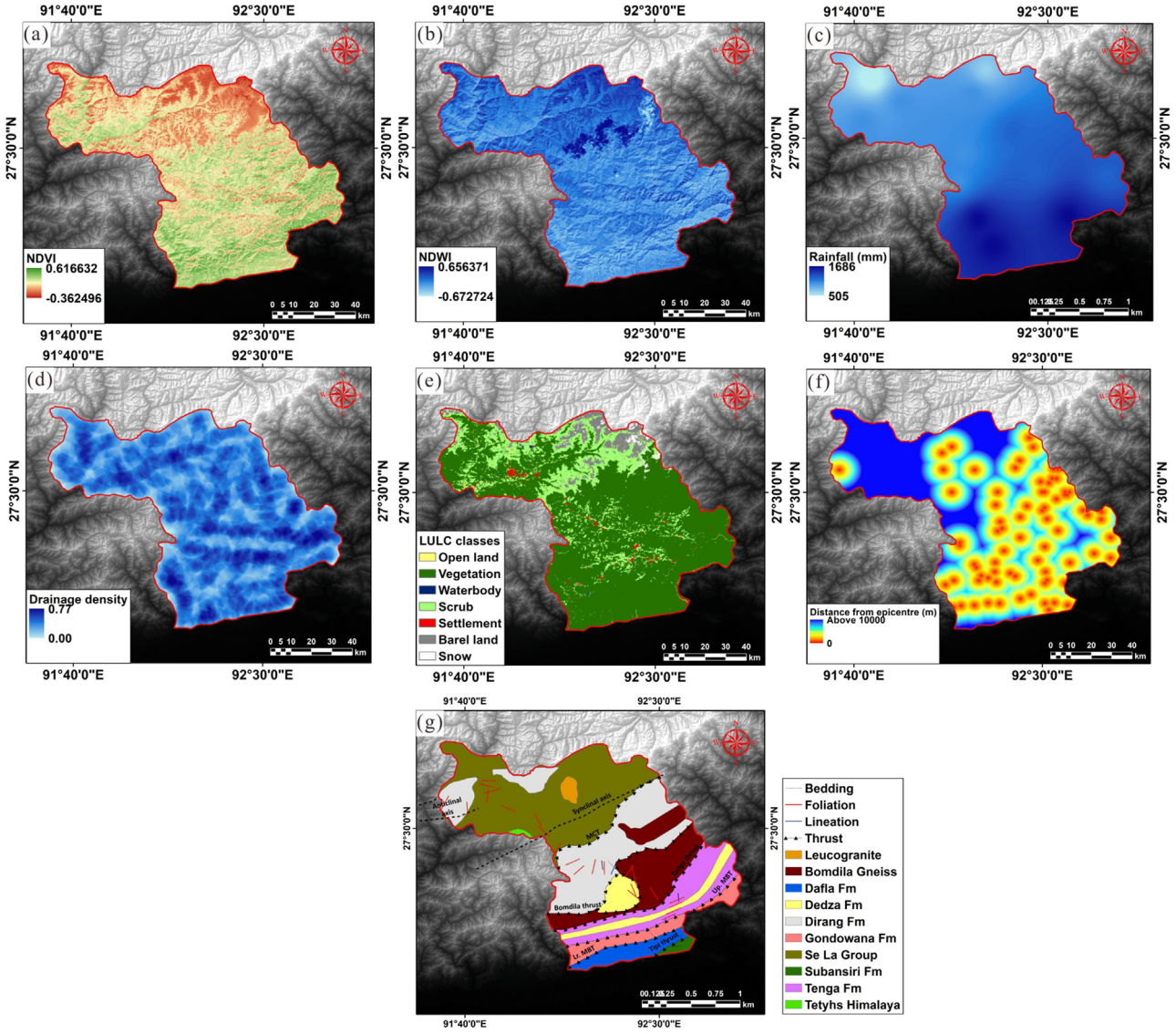

Fig. 4. Landslide conditioning factors (a) NDVI, (b) NDWI, (c) Rainfall, (d) drainage density, € LULC, (f) distance from epicentre and (g) geology.

ing Vapnik-Chervonenk is (VC) dimension and statistical theory (Vapnik and Cortes, 1995; Chervonenkis, 2013). It is classified based on optimal hyper-plain that maximizes the area of the classes and the data nearest to the hyper-plain (It completely separates the vectors into two distinct classes and maximizes the margins of training datasets) is called the support vectors (Pourghasemi et al., 2013). SVM model is suitable for the lesser number of dataset and its primary aim is to build and separating the data on the basis of hyper-plain (Xu et al., 2012). This machine learning algorithm has been used to resolve two important aspects. Firstly, it finds a proper way to compute the correlation between the input vectors. Secondly, it designs a linear structure by merging the output of training samples (Wang et al., 2020). Linear separation condition follows,

$$D = \{(x_1, y_1),\ (x_2, y_2), \ldots\ldots,\ (x_n, y_n)\} \tag{8}$$

Here, $x_1 \in X \sqsubset R^m, y_1 \in \{1, -1\},\ and\ i = 1, 2, \ldots . n.$ it can be separated by a hyper-plain, it follows,

$$(w.x) + b = 0, w \in R^n, b \in R \tag{9}$$

Here, b signifies the scaler base whereas (.) represents the scaler operation.

Similar input sample has greater contribution in output. In the case of original nearest neighbour classifier, it is expressed as,

$$\sum_{i=1}^{l} y_i, \lambda_i * K(x, x_i) \tag{10}$$

Here, l signifies training sample and y_i denotes the result of training samples.

3.3.3. Adaptive Boosting (AdaBoost)

AdaBoost is widely used ensemble machine learning algorithm that was proposed by Freund and Schapire (1997). It is a boosting algorithm that consists a series of classifiers and every classifier in this ensemble learning method attempts to precisely classify the dataset (Kadavi et al., 2018). AdaBoost is an ensemble boosting algorithm that is notably applied to amplify the predictive capability in various machine learning methods such as, support vector machine, random forest, tree based models etc. (Wang et al., 2009). This classifier uses adaptive resampling technique at the time of selection of training samples. AdaBoost algorithm has been achieved by the changes of data distribution that determines the weight of every sample based on correct classification of the samples in the training set and exactness of the previous entire clas-

sification (Wu et al., 2020). For the iteration, this algorithm develops a new training dataset that is the sample dataset of the given training dataset. Subsequently, a base classifier (S_t) also uses in new training dataset for learning. The error of S_t (E_s) has been calculated by the equation

$$E_s = \sum_{i:s_t \neq y_i} w_s \tag{11}$$

The following steps have been maintained during learning process,

$$w_{i+1} = w_i . \exp(-\beta . z_i) \tag{12}$$

Here, β and z_i have been determined using the following equation,

$$\beta = 0.5 \ln\left(\frac{1 - E_s}{E_s}\right) \tag{13}$$

$$z_i = \begin{matrix} 1 \ if \ S_t(x_i) = y_i \\ -1 \ if \ S_t(x_i) \neq y_i \end{matrix} \tag{14}$$

The calculated weights are normalized by the following equation,

$$w_i + 1 = \frac{w_i + 1}{\sum_{i=1}^{n} w_i + 1} \tag{15}$$

In the last stage, the boosting algorithm integrates all the results of the classification.

3.4. Comparison of the implemented models

To compare the implemented models, two different non-parametric tests have been considered in this research (Wilcoxon Signed Rank test and Friedman test). The non-parametric test does not require any special distribution pattern of the datasets. If the data does not follow the normal distribution pattern, this kind of tests will be very helpful to interpret the inferential statistics (Martínez-Álvarez et al., 2013). Friedman test assumes that there is no substantial change in the model efficiency if the significance level of Alpha is 0.05 (Beasley and Zumbo, 2003). Wilcoxon Signed Rank test has been performed to assess the pair-wise statistical differences between the models. In the case of this research p-value and *z*-value have been utilized to evaluate the significant differences among them. Here, the null hypothesis will be rejected if the value of p is greater than 5% level of significance and the value of z is greater than the critical value of z (+1.96 and -1.96).

Friedman test follows,

$$F_R = [\frac{12}{NK(K+1)} \sum_{j=1}^{K} R_j^2] - 3N(K+1) \tag{16}$$

Here, R_j refers sum of the ranks. N & K refer number of blocks and treatments

The Wilcoxon signed rank test follows,

$$W = \sum_{i=1}^{N_r} [sgn(x_{2,i} - x_{1,i}) . R_i \tag{17}$$

sgn refers the Sign function. $x_{2,i}, x_{1,i}$ signify corresponding rank pairs.

3.5. Accuracy assessment

In this research, different model performance methods have been evaluated using confusion matrix. Various statistical indices such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), prevalence, detection rate were used to assess the model accuracy. Total four types of possibility (true positive (TP), true negative (TN), false negative (FN), false positive (FP)) have been applied to estimate the statistics. Higher values of the statistics signify better accuracy of the model and vice versa.

$$PPV = \frac{TP}{FP + TP} \tag{18}$$

$$NPV = \frac{TP}{TP + FN} \tag{19}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{20}$$

$$Specificity = \frac{TN}{FP + TN} \tag{21}$$

Here, TP shows true positive, TN shows true negative, FN signifies false negative and FP shows false positive.

ROC is another important statistical device designed for the examination of the degree of validation for applied machine learning models that contain two dimensions such as events and non-event phenomena. This curve is plotted on the basis of 1-specificity (x axis) and sensitivity (y axis). The predictive capability of the executed model can be classified using the value of area under curve (AUC-ROC). The ROC-AUC follows the given equation,

$$S_{AUC} = \sum_{k=1}^{n} (X_{K+1} - X_K)\left(S_K + 1 - S_{K+1} - \frac{S_K}{2}\right) \tag{22}$$

Here, S_{AUC} signifies the area under curve (AUC). X_K and S_K represent 1-specificity and sensitivity respectively.

4. Results

4.1. Multi collinearity assessment

Multi collinearity assessment is knowingly used to diminish the biasness of the model. The outcomes of multi collinearity assessment test illustrate that there is no existence of multi collinearity among the variables because the tolerance and VIF values of the factors are not surpassed there threshold limit or critical limit. Highest and lowest values of variance inflation factors are 4.59 and 1.16 respectively on the other hand the highest and lowest tolerance values are 0.86 and 0.22 respectively (Table 2). It suggested that altitude has highest importance value among the other landslide conditioning factors (72.72) followed by river density (68.91), distance from road (60.85), geology (57.48), NDVI (54.19). Profile

Table 2
Result of the multi-collinearity statistics of the used causative factors.

Factors	Multi collinearity statistics	
	Tolerance	VIF
River density	0.699	1.431
Altitude	0.317	3.151
NDWI	0.243	4.117
Aspect	0.446	2.014
LULC	0.669	1.495
Lineament density	0.862	1.16
Distance from river	0.571	1.904
Distance from road	0.723	1.384
NDVI	0.218	4.592
Plain curvature	0.529	1.891
Distance from epi-centre	0.329	3.781
Profile curvature	0.55	1.819
Rainfall	0.293	3.413
Geology	0.556	1.887
Soil	0.565	1.77
Slope	0.614	1.678

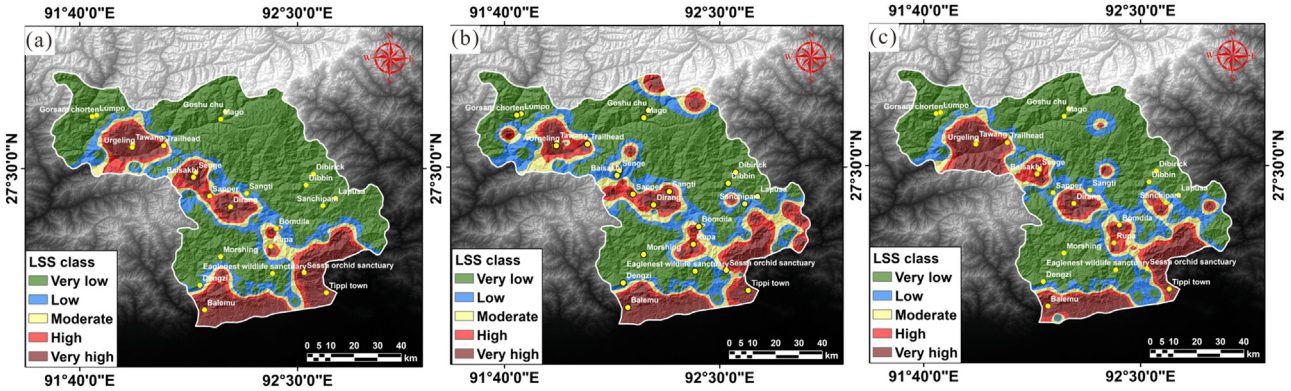

Fig. 5. Landslide susceptibility maps using machine learning models, (a) Adaptive Boosting (AdaBoost), (b) Support Vector Machine (SVM) and (c) Extreme Gradient Boosting (XGBoost).

Table 3
Analysis the landslide conditioning factors using information gain ratio (InGR).

Landslide conditioning factors	Average merit (AM)
River density	0.141
Altitude	0.187
NDWI	0.021
Aspect	0.007
LULC	0.097
Lineament density	0.011
Distance from river	0.021
Distance from road	0.114
NDVI	0.057
Plain curvature	0.003
Distance from epi-centre	0.009
Profile curvature	0
Rainfall	0.064
Geology	0.089
Soil	0.014
Slope	0.078

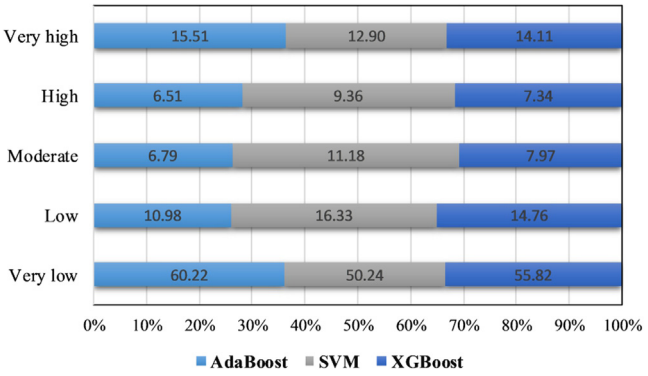

Fig. 6. Bar graph showing the area wise proportion of the susceptibility classes.

curvature has the lowest amount of importance for landslide (2.17) particularly in landslide susceptible study. The influence of the landslide conditioning factors has been appraised through InGR statistics. Altitude (0.187), river density (0.141) etc. have high InGR value that indicates better predictive capability whereas profile curvature (0.000) and aspect (0.007) have low InGR value that indicates minimal relevancy (Table 3).

4.2. Landslide vulnerability analysis

In this scientific study, various machine learning methods (SVM, XGBoost, AdaBoost) have been applied to prepare the actual landslide susceptibility maps in two western districts of Arunachal Himalayas. The landslide susceptibility maps have been classified into five different landslide probable zones in ArcGIS platform on the basis of natural break raster classification method (very low, low, moderate, high and very high) (Fig. 5). In the case of support vector machine model, very low, low, moderate, high and very high susceptibility classes have been occupied 3682.06 (50.24%), 1196.51 (16.33%), 819.53 (11.18%), 685.7 (9.36%) and 945.26 (12.90%) respectively (Figs. 5b, 6). In case of extreme gradient boosting method, very low, low moderate, high and very high susceptibility classes have been occupied 4091.41 (55.82%), 1081.57 (14.76%), 584.04 (7.97%), 538.19 (7.34%) and 1033.85 (14.11%) respectively (Figs. 5c, 6). Similarly, adaptive boosting model shows that very low, low moderate, high and very high susceptibility classes have been occupied 4413.42 (60.22%), 804.38 (10.98%), 497.73 (6.79%), 476.78 (6.51%), 1136.76 (15.51%) respectively (Figs. 5a, 6). The machine learning models significantly identified some middle part and south and south eastern parts which are more vulnerable for landslide hazard particularly the areas of Urgeling, Tawang, Rump, Bomdila, Dirang, Senge, Balemu, Tippi town, Baisakhi etc.

4.3. Comparison of landslide susceptible models

In this scientific study, test datasets of landslide susceptibility models have been employed to compare the models with each other. The Friedman test reveals that the level of confidence (p value) is less than 0.05 which clearly indicates the rejection of null hypothesis and models are statistically difference in terms of their performance (Table 5). AdaBoost model has the highest mean rank that may be considered finest performing model here. For pairwise comparison of the applied models, Wilcoxon-Signed rank test has been applied. The result of Wilcoxon-Signed rank clearly portrays that pairs of the models are statistically different in terms of their performance (p value is less than 0.05 and z value surpass the threshold or critical value -1.96 to +1.96) (Table 6).

4.4. Model validation

Validation of the models as well as the accuracy can be shown through the receiver operating curve (ROC) and the value of area under curve (AUC). The value of AUC can easily symbolize the prediction capability of the model and it is considered as a quantitative value for model validation and comparison (Table 4). The AUC values of the testing dataset of the models SVM, XGBoost and AdaBoost are 0.85, 0.89 and 0.92, respectively (Fig. 7). Although all the models provide reasonable result but AdaBoost performs better accuracy in terms of AUC value. Other statistical accuracy in-

Table 4
Performance evaluation of the machine learning based landslide susceptibility models (AUC values) with respective standard error and confidence interval.

Models	AUC value	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
AdaBoost	0.921	0.016	0	0.887	0.961
SVM	0.855	0.032	0	0.793	0.917
XGBoost	0.893	0.02	0	0.859	0.938

Table 5
Result of the Friedman test of LSMs associated with mean rank along with chi-square and significance value.

Models	Mean ranks	Chi-square	Significance
XGBoost	1.99	7.473	0.024
AdaBoost	2.12		
SVM	1.88		

Table 6
Comparison of the applied machine learning models using Wilcoxon signed-rank test.

Models	XGBoost - AdaBoost	SVM - XGBoost	SVM - AdaBoost
Z	-2.599[b]	-2.331[c]	-3.148[c]
Asymp. Sig. (2-tailed)	0.009	0.02	0.002

[b] Based on positive ranks.
[c] Based on negative ranks.

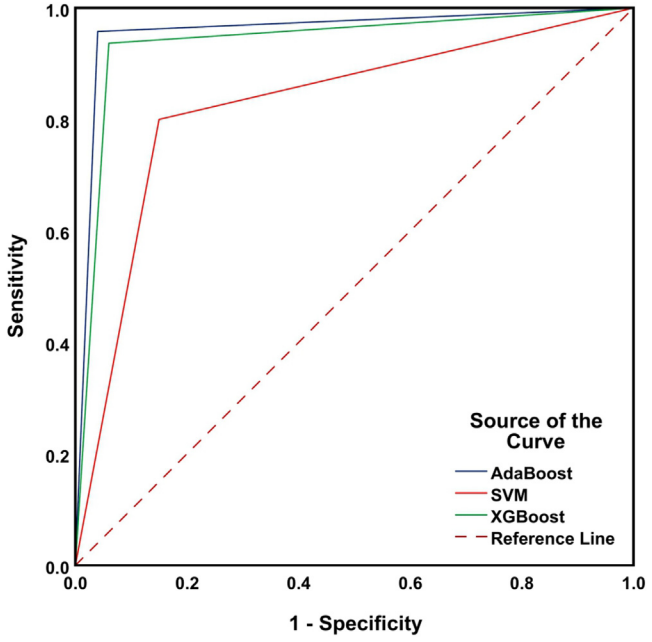

Fig. 7. Validation assessment of the applied ML models through ROC curve and AUC analysis.

dicators such as sensitivity, specificity, positive pred value, negative pred value, prevalence, detection rate have been applied here. The values of these statistical indicators are 0.8133, 0.9014, 0.8948, 0.8219, 0.5972 and 0.4593; 0.9200, 0.8700, 0.8942, 0.8978, 0.6372 and 0.5338; 0.9267, 0.9171, 0.9263, 0.9067, 0.5772 and 0.5083 for the model SVM, XGBoost and AdaBoost respectively (Table 7). AdaBoost has relatively higher values of the performance indicator statistics that indicating the ability of AdaBoost model to learn the complex relationship as well as correlation among the geospatial characteristics and the incidence of landslides.

Table 7
Comparison of different accuracy indicators of the applied models.

Accuracy indicators	SVM	XGBoost	AdaBoost
Sensitivity	0.8133	0.92	0.9267
Specificity	0.9014	0.87	0.9171
Positive pred value	0.8948	0.8942	0.9263
Negative pred value	0.8219	0.8978	0.9067
Prevalence	0.5972	0.6372	0.5772
Detection rate	0.4593	0.5338	0.5083

5. Discussion

Identification technique of future probable landslide areas or future landslide potential areas is an suitable tool for land use planning and decision making particularly in hilly or mountainous terrain (Park et al., 2013). In this research three highly accepted machine learning algorithms (SVM, XGBoost and AdaBoost) have been applied for spatial analysis of landslide susceptibility. Huge number of methods is being applied in different parts of the world for precise mapping of landslide. These are analysed to find out the complex relationship between predictors or independent factors with the dependent variable. Landslide factors are selected on the basis of literature and field study then multicollinearity check has been adopted to investigate whether there is any existence of multi-collinearity problem. Model comparison has been done through different statistical test (Friedman test and Wilcoxon-Signed rank) and it significantly shows that models have a statistical significant different in context of their statistical analysis and outcomes. It has identified that altitude; river density and distance from road are the most influencing factors for landslide. All statistical model prediction indicators including AUC, sensitivity, specificity, positive pred vale, negative pred value, prevalence, detection rate show that AdaBoost model has a better prediction ability and high goodness-of-fit than other applied models. The advantage of AdaBoost model is that it can act as a Meta classifier and maintains a balance between diversity and accuracy and it reduces the data over fitting problem in the training dataset. In the case of supervise machine learning algorithm, the model overfitting problem cannot be avoided completely. Based on the bias training data the overfitting problem will occur and it can reduce the accuracy of the model (Bu and Zhang, 2020). This problem will arise when the model components are appraised against the wrong reference distribution. Kawabata and Bandibas (2009) considered that geology is the most vital parameter in the occurrence of landslide. Kıncal et al. (2009) also showed that lithology has the greatest impact for the occurrence landslide in hilly terrain. Except lithology other import completing factors are altitude, river density, slope, vegetation etc. Hong et al. (2007) also identified that altitude, geology; slope, soil texture and structure are most substantial factors of landslide. Like other parts of Himalaya, Arunachal Himalaya exhibits different well-known tectonomorphic zones (lesser, greater and Trans-Himalayan zones) and all these segments exist along the Bhalukpong-Bomdila-Tawang road section (Singh et al., 2014). Land use and land cover has a direct

Fig. 8. Some active landslide incidents happened in various pockets of western Arunachal (Source: Arunachal 24.in).

impact over the triggering and activation of landslide (Lu et al., 2015). With the changing pattern of human activity, the land use of a region is being changed simultaneously and high impact of human activity leads slope instability processes (Lu et al., 2015). In the western parts of Arunachal Pradesh, the higher Himalayan rocks area termed as Se La group which is distinct with the rock of Bomdila group (the crystallines of lesser Himalayan) in the south by the MCT (main central thrust) (Singh et al., 2014). The Bomdila group consists of low grade penecontemporaneous basic volcanic rocks, argillaceous metamorphites and large scale igneous intrusive (Bhushan et al., 1991). This region consists with Subansiri formation, Tenga formation, Dedza formation, Dirang formation etc. and during field survey it has been observed that Bomdila group has the highest probability of landslide occurrence. The Tenga formation is characterised by mafic metavolcanics, quartzite and phyllites while the Dedza formation is compacted with carbonate rocks along with a little quantity of black slate. The Dirang formation is composed by muscovite-biotite schist, tourmaline marble, sericite quartzite and carbonate phyllite (Srivastava et al., 2011). MCT in Arunachal Pradesh is under strain and recognized as high ductile shaving areas that are unveiling more than a single thrust plane (Bhattacharjee and Nandy, 2008). Multi-dimensional first and second generation structural imprints have been exposed on these lithological formations along Bhalukpong-Bomdila-Tawang road section. These formations are highly landslide prone due to low degree of strength and excessive stress (heavey loaded vehicles) is being imposed throughout year due to important international tourist destination. Rainwater can easily penetrate through these vulnerable structures during rainy months as a result most of the landslides occur during the monsoonal period. Dry spell is also partially responsible for landslide particularly western part of Arunachal Himalayas. The MBT passes through the study area and the continuous pressure of Indian plate towards the northern side is always fallen in this region that makes huge landslide prone condition. Most of the lineaments are trending parallel to MBT and playing a vital role for landslides incidents (Rawat and Joshi, 2012). During the field survey, it has been observed that the upslope areas as well as steep areas are recurrent landslide prone and the outcrops are much crumbled (Fig. 8). The incised river terrace, hanging valley, steep slope, escarpment, fault scarp and gorges are contained the signature of tectonic activities within the study area. In Arunachal Pradesh, Jhum cultivation and high rate of deforestation are the significant anthropogenic factors which accelerate landslide (Rawat and Joshi, 2012). Selection of proper landslide types and assessment of appropriate landslide models are very much difficult work and these will highly control the final outcome map of landslide susceptibility. The selected causative factors are carefully considered based on their relative significance and data adequacy. The accurate landslide susceptibility mapping on a regional scale depends on various influencers but it will be quite difficult in terms of different perspective of the researchers.

6. Conclusions

This study exhibits that remote sensing based landslide probability models play a vital role in the precise mapping of landslide susceptibility zonation mapping. Here, three different machine learning based approaches (SVM, AdaBoost and XGBoost) have been used for landslide susceptibility mapping in Tawang and West Kameng district of Arunachal Pradesh. There are some limitations of machine learning model based studies. Machine learning requires huge amount of data set for training and these are highly susceptible to errors (Malik, 2020). The results attain from machine learning may have errors due to the statistical reasoning. The systems generate the data based on previous results and the data that was loaded previously. Landslides are considered as

one of the gravest natural hazard that imposes threats for both life and property. Short term and long term solutions are more necessary to combat against these daunting challenges. The landslide vulnerability map recently becomes popular in the context of landslide prone area delineation and management. This kind of studies can definitely help to implement different kinds of management strategies such as proper land use management, afforestation, policy making and barrier construction etc. The landslide susceptibility study of western Arunachal reveals that relief, river density, geology are the most influencing factors than anthropogenic activities. Various unconsolidated based geological formations in this region are very much susceptible for landslide occurrence and in the recent years, shifting (Jhum cultivation) cultivation along with high rate of deforestation on this rugged Himalayan terrain exaggerates the magnitude of landslide. The AdaBoost model has the highest prediction rate or success rate compared to other used models. Therefore, AdaBoost model has been considered most optimized model here. From this scientific study, researchers will get different future research directions and wings. Consequently, new hybrid models will develop to reduce the direct and indirect landslide risk particularly in Himalayan terrain. These research findings as well as strategies will be very much important for future development and sustainable livelihood of the people. Selection of appropriate method can assist to combat against this kind of geo-environmental hazards. It is very difficult to stop completely such geo-hazard in tectonically active mountainous belt of Himalaya. Therefore, various management techniques such as hard, soft and bio-engineering should be implemented for very high and high landslide prone zones particularly at the vicinity of build environment (Bera, 2007).

Declaration of Competing Interest

The auhors declare no competing interest.

References

Al-Najjar, H.A.H., Pradhan, B., Kalantar, B., Sameen, M.I., Santosh, M., Alamri, A., 2021. Landslide susceptibility modeling: an integrated novel method based on machine learning feature transformation. Remote Sens. 13, 3281. doi:10.3390/rs13163281.

Althuwaynee, O.F., Pradhan, B., Park, H.J., Lee, J.H., 2014. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides 11, 1063–1078. doi:10.1007/s10346-014-0466-0.

Anbalagan, R., Chakraborty, D., Kohli, A., 2008. Landslide hazard zonation (LHZ) mapping on meso-scale for systematic town planning in mountainous terrain. J. Sci, Ind. Res. 67, 486–497.

Ballabio, C., Sterlacchini, S., 2012. Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study. Italy. Math Geosci. 44, 47–70. doi:10.1007/s11004-011-9379-9.

Beasley, T.M., Zumbo, B.D., 2003. Comparison of aligned Friedman rank and parametric methods for testing interactions in split-plot designs. Computat. Stat. Data Anal. 42 (4), 569–593.

Bera, B., 2007a. A geo technical investigation into the causes and management of landslides in Gangtok town of Sikkim Himalayas. xxi, 238p. http://hdl.handle.net/10603/155289

Bera, B., 2007b. Landslide hazard zonation mapping. Contemporary Issues and Techniques in Geography 24-35. https://www.researchgate.net/publication/327721447_Landslide_Hazard_Zonation_Mapping

Bera, B., 2008. A Geotechnical Evaluation of the Sichey Landslide of Gangtok of Sikkim Himalayas Geographical Review of India, 69 (4), 434-441. https://www.researchgate.net/publication/327578564_A_Geotechnical_Evaluation_of_the_Sichey_Landslide_of_Gangtok_of_Sikkim_Himalayas_Geographical_Review_of_India

Bera, B., 2009. A Geotechnical Appraisal of the Chanmari Landslide of Gangtok, Sikkim. Geomorphology in India 199-206. https://www.researchgate.net/publication/327721644_A_Geotechnical_Appraisal_Of_the_Chanmari_Landslide_of_Gangtok_Sikkim

Bera, B., 2010. Landslide management in eastern Himalayan scenario. Indian J. Landsc. Syst. Ecol. Stud. 33 (1), 601–606. https://www.researchgate.net/publication/327558977_Landslide_management_in_eastern_Himalayan_scenario_Indian_Journal_of_Landscape_System_and_Ecological_Studies.

Bhattacharjee, S., Nandy, S., 2008. Geology of the Western Arunachal Himalaya in parts of Tawang and west Kameng District, Arunachal Pradesh. J. Geol. Soc. India 72, 199–207. doi:10.1007/s12594-009-0043-7.

Bhushan, S.K., Bindal, C.M., Aggrawal, R.K., 1991. Geology of Bomdila Group in Arunachal Pradesh. Himalayan Geol. 2 (2), 207–214.

Bu, C., Zhang, Z., 2020. Research on overfitting problem and correction in machine learning. J. Phys. Conf. Ser. 1693. https://iopscience.iop.org/article/10.1088/1742-6596/1693/1/012100.

Bui, D.T., Hoang, N.D., Martínez-Álvarez, F., Ngo, P.T.T., Hoa, P.V., Pham, T.D., Samui, P., Costache, R., 2020. A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. Sci. Total Environ. 701, 134413. doi:10.1016/j.scitotenv.2019.134413.

Burrough, P.A., 1988. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford.

Chen, W., Panahi, M., Pourghasemi, H.R., 2017a. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), diferential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. Catena 157, 310–324. doi:10.1016/j.catena.2017.05.034.

Chen, W., Xie, X., Wang, J., et al., 2017b. A comparative study of logistic model tree, random forest, and classifcation and regression tree models for spatial prediction of landslide susceptibility. Catena 151, 147–160. doi:10.1016/j.catena.2016.11.032.

Chervonenkis, A.Y., 2013. Early history of support vector machines. Empirical Inference Festschrift in Honor of Vladimir N. Vapnik, pp 13–20

Chung, C.J.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30 (3), 451–472. doi:10.1023/B:NHAZ.0000007172.62651.2b.

Dada, E.G., Bassi, J.S., Chiroma, H., Adetunmbi, A.O., Ajibuwa, O.E., et al., 2019. Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 5 (6), e01802. doi:10.1016/j.heliyon.2019.e01802.

Dadson, S., Hovius, N., Chen, H., Dade, W., 2003. Links between erosion, runoff variability and seismicity in the Taiwan orogen. Nature 426, 648–651. doi:10.1038/nature02150.

Dixon, J.L., Hartshorn, A.S., Heimsath, A.M., DiBiase, R.A., Whipple, K.X., 2012. Chemical weathering response to tectonic forcing: a soils perspective from the San Gabriel Mountains, California. Earth Planet. Sci. Lett. 323-324, 40–49. doi:10.1016/j.epsl.2012.01.010.

Duna, C.R., D'Arcy, M., McDonald, J., Whittaker, C.A., 2018. Lithological controls on hillslope sediment supply: insights from landslide activity and grain size distributions. Earth Surf. Process. Landf. 43, 956–977. doi:10.1002/esp.4281.

Erener, A., Duzgun, H.S.B., 2010. Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). Landslides 7, 55–68. doi:10.1007/s10346-009-0188.

Erener, A., Sivas, A.A., Selcuk-Kestel, A.S., Du¨zgu¨n, H.S., 2017. Analysis of training sample selection strategies for regression-based quantitative landslide susceptibility mapping methods. Comput. Geosci. 104, 62–74. doi:10.1016/j.cageo.2017.03.022.

Fischer, L., Kaab, A., Huggel, C., Noetzli, J., 2006. Geology, glacier retreat and permafrost degradation as controlling factors of slope instabilities in a high-mountain rock wall: the Monte Rosa east face. Nat. Hazards Earth Syst. Sci. 6, 761–772. doi:10.5194/nhess-6-761-2006.

Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences,55,, 119–139.

Gabet, E.A., 2007. Theoretical model coupling chemical weathering and physical erosion in landslide-dominated landscapes. Earth. Planet. Sci. Lett. 264, 259–265. doi:10.1016/j.epsl.2007.09.028.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. Geomorphology 94, 268–289. doi:10.1016/j.geomorph.2006.09.023.

Gariano, S.L., Sarkar, R., Dikshit, A., Dorji, K., Brunetti, M.T., Peruccacci, S., Melillo, M., 2019. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. Bull. Eng. Geol. Environ. 78, 4325–4332. doi:10.1007/s10064-018-1415-2.

Georganos, S., Grippa, T., VanHuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., Wolff, E., 2017. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GISci. Remote. Sens. 55, 221–242. doi:10.1080/15481603.2017.1408892.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sens. 11, 196. doi:10.3390/rs11020196.

Glade, T., 2003. Landslide occurrence as a response to land use change: a review of evidence from New Zealand. Catena 51, 297–314. doi:10.1016/S0341-8162(02)00170-4.

Guglielmi, Y., Cappa, F., Avouac, J.P., Henry, P., Elsworth, D., 2015. Seismicity triggered by fluid injection–induced aseismic slip. Science 348, 1224–1226. doi:10.1126/science.aab0476.

Handwerger, A.L., Roering, J.J., Schmidt, D.A., 2013. Controls on the seasonal deformation of slow-moving landslides. Earth Planet. Sci. Lett. 377, 239–247. doi:10.1016/j.epsl.2013.06.047.

Highland, L., Bobrowsky, P.T., 2008. The Landslide Handbook: a Guide to Understanding Landslides 129 (In the U.S).

Hong, Y., Adler, R., Huffman, G., 2007. Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat. Hazards 43 (2), 245–256. doi:10.1007/s11069-006-9104-z.

Iverson, R.M., Major, J.J., 1987. Rainfall, ground-water flow, and seasonal movement at Minor Creek landslide, northwestern California: physical in-

terpretation of empirical relations. Geol. Soc. Am. Bull. 99, 579–594 doi:10.1130/0016-7606(1987)99%3C579:RGFASM%3E2.0.CO;2.
Johnston, R., Jones, K., Manley, D., 2018. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Qual. Quant. 52 (4), 1957–1976.
Kadavi, P.R., Lee, C.W., Lee, S., 2018. Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sens. 10, 1252. doi:10.3390/rs10081252.
Kawabata, D., Bandibas, J., 2009. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN). Geomorphology 113 (1–2), 97–109. doi:10.1016/j.geomorph.2009.06.006.
Kıncal, C., Akgun, A., Koca, Y., 2009. Landslide susceptibility assessment in the Izmir (West Anatolia, Turkey) city center and its near vicinity by the logistic regression method. Environ. Earth Sci. 59 (4), 745–756. doi:10.1007/s12665-009-0070-0.
Komadja, G.C., Pradhan, S.P., Oluwasegun, A.D., Roul, A.R., et al., 2021. Geotechnical and geological investigation of slope stability of a section of road cut debris-slopes along NH-7, Uttarakhand, India. Results Eng. 10, 100227. doi:10.1016/j.rineng.2021.100227.
Komadja, G.C., Pradhan, S.P., Roul, A.R., Adebayo, B., Habinshuti, J.B., Glodji, L.A., Onwualu, A.P., 2020. Assessment of stability of a Himalayan road cut slope with varying degrees of weathering: a finite-element-model-based approach. Heliyon 6. doi:10.1016/j.heliyon.2020.e05297.
Korup, O., Montgomery, D.R., 2008. Tibetan plateau river incision inhibited by glacial stabilization of the Tsangpo gorge. Nature 455, 786–789. doi:10.1038/nature07322.
Kumar, G., 1997. Geology of Arunachal Pradesh. Geological society of India, Bangalore, p. 217.
Kumar, R., Anbalagan, R., 2016. Landslide susceptibility mapping using analytical hierarchy process (AHP) in Tehri reservoir rim region, Uttarakhand. J. Geol. Soc. India 87, 271–286. doi:10.1007/s12594-016-0395-8.
Lee, S., 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. Int. J. Remote Sens. 26, 1477–1491. doi:10.1080/01431160412331331012.
Liu, T., Abd-Elrahman, A., Morton, J., Wilhelm, V.L., 2018. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. GIScience Remote. Sens. 55, 243–264. doi:10.1080/15481603.2018.1426091.
Lombardo, L., Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results. Eng. Geol. 244, 14–24. doi:10.1016/j.enggeo.2018.07.019.
Lu, P., Bai, S., Casagli, N., 2015. Spatial relationships between landslide occurrences and land cover across the Arno River Basin (Italy). Environ. Earth Sci. 74 (7), 5541–5555. doi:10.1007/s12665-015-4569-2.
Malik, M., 2020. A hierarchy of limitations in machine learning. 10.48550/arXiv.2002.05193
Martínez-Álvarez, F., Reyes, J., Morales-Esteban, A., Rubio-Escudero, C., 2013. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. Knowl. Based Syst. 50, 198–210. doi:10.1016/j.knosys.2013.06.011.
Mishra, D.K., 2007. Evidence of Neotectonic activity along active faults in Arunachal Himalay, NE India. Himalayan Geol. 28 (2), 75–78.
Montgomery, D.R., Brandon, M.T., 2002. Non-linear controls on erosion rates in tectonically active mountain ranges. Earth Planet. Sci. Lett. 201, 481–489. doi:10.1016/S0012-821X(02)00725-2.
Nandi, A., Shakoor, A., 2010. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol. 110 (1-2), 11–20. doi:10.1016/j.enggeo.2009.10.001.
Ozdemir, A., Altural, T., 2013. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: sultan mountains, SW Turkey. J. Asian Earth Sci. 64, 180–197. doi:10.1016/j.jseaes.2012.12.014.
Pandey, V.K., Sharma, K.K., Pourghasemi, H.R., Bandooni, S.K., 2019. Sedimentological characteristics and application of machine learning techniques for landslide susceptibility modelling along the highway corridor Nahan to Rajgarh (Himachal Pradesh), India. Catena 182, 104150. doi:10.1016/j.catena.2019.104150.
Park, S., Choi, C., Kim, B., Kim, J., 2013. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artifcial neural network methods at the Inje area, Korea. Environ. Earth Sci. 68 (5), 1443–1464. doi:10.1007/s12665-012-1842-5.
Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40, 927–930. doi:10.1130/G33217.1.
Pham, B.T., Nguyen, V.T., Ngo, V.L., Trinh, P.T., Ngo, H.T.T., Bui, D.T., 2017. A novel hybrid model of rotation forest based functional trees for landslide susceptibility mapping: a case study at Kon Tum Province, Vietnam. In: Proceedings of the International Conference on GeoSpatial Technologies and Earth Resources.
Pham, B.T., Prakash, I., 2019. Evaluation and comparison of LogitBoost ensemble, fsher's linear discriminant analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. Geocarto Int. 34, 316–333. doi:10.1080/10106049.2017.1404141.
Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., Gokceoglu, C., 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province. Iran. J. Earth Syst. Sci. 122, 349–369. doi:10.1007/s12040-013-0282-2.
Pradhan, A.M.S., Kang, H.S., Kim, Y.T., 2017a. Hybrid Landslide Warning Model for Rainfall Triggered Shallow Landslides in Korean Mountain. In: Proceedings of the Advancing culture of living with landslides. Advances in landslide technology, 3, pp. 193–200.
Pradhan, A.M.S., Kang, H.S., Lee, S., Kim, Y.T., 2017b. Spatial model integration for shallow landslide susceptibility and its run out using a GIS-based approach in Yongin, Korea. Geocarto Int. 32, 420–441. doi:10.1080/10106049.2016.1155658.
Pradhan, A.M.S., Kim, Y.T., 2020. Rainfall-induced shallow landslide susceptibility mapping at two adjacent catchments using advanced machine learning algorithms. ISPRS Int. J. Geo Inf. 9, 569. doi:10.3390/ijgi9100569.
Rawat, J., Joshi, R., 2012. Remote-sensing and GIS-based landslide-susceptibility zonation using the landslide index method in Igo River Basin, Eastern Himalaya, India. Int. J. Remote Sens. 33, 3751–3767. doi:10.1080/01431161.2011.633121.
Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology 115 (1–2), 172–187. doi:10.1016/j.geomorph.2009.10.002.
Roback, K., et al., 2017. The size, distribution, and mobility of landslides caused by the 2015 M w 7.8 Gorkha earthquake Nepal. Geomorphology 301, 121–138. doi:10.1016/j.geomorph.2017.01.030.
Sati, S.P., Sundriyal, Y.P., Rawat, G.S., 2007. Geomorphic indicators of eotectonic activity around Srinagar (Alaknanda basin), Uttarakhand. Curr. Sci. 92, 824–829.
Schulz, W.H., McKenna, J.P., Kibler, J.D., Biavati, G., 2009. Relations between hydrology and velocity of a continuously moving landslide—evidence of pore-pressure feedback regulating landslide motion? Landslides 6, 181–190. doi:10.1007/s10346-009-0157-4.
Sevgen, E., Kocaman, S., Nefeslioglu, H.A., Gokceoglu, C.A., 2019. Novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. Sensors 19, 3940. doi:10.3390/s19183940.
Shou, K., Chen, J., 2021. On the rainfall induced deep-seated and shallow landslide hazard in Taiwan. Eng. Geol. 288, 106156. doi:10.1016/j.enggeo.2021.106156.
Sidle, R.C., Ochiai, H., 2006. Landslides: Processes, Prediction, and Land Use. American Geophysical Union, Washington, DC, USA.
Singh, C.D., Kohli, A., Kumar, P., 2014. Comparison of results of BIS and GSI guidelines on macrolevel landslide hazard zonation — a case study along highway from Bhalukpong to Bomdila, West Kameng district, Arunachal Pradesh. J. Geol. Soc. 83, 688–696. doi:10.1007/s12594-014-0101-7.
Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K.I., 2017. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. GISci. Remote. Sens. 54, 918–938. doi:10.1080/15481603.2017.1351149.
Srivastava, H.B., Srivastava, V., Srivastava, R.K., et al., 2011. Structural analyses of the crystalline rocks between Dirang and Tawang, West Kameng district, Arunachal Himalaya. J. Geol. Soc. India 78, 45. doi:10.1007/s12594-011-0066-8.
Summa, V., Margiotta, S., Medici, L., Tateo, F., 2018. Compositional characterization of fine sediments and circulating waters of landslides in the southern Apennines – Italy. Catena 171, 199–211. doi:10.1016/j.catena.2018.07.009.
Tanyas, H., Hill, K., Mahoney, L., Fadel, I., Lombardo, L., 2022. The world's second-largest, recorded landslide event: Lessons learnt from the landslides triggered during and after the 2018 Mw 7.5 Papua New Guinea earthquake. Eng. Geol. 297, 106504. doi:10.1016/j.enggeo.2021.106504.
Tanyas, H., Rossi, M., Alvioli, M., van Westen, C.J., Marchesini, I., 2019. A global slope unit-based method for the near real-time prediction of earthquake-induced landslides. Geomorphology 327, 126–146. doi:10.1016/j.geomorph.2018.10.022.
Terzaghi, K., Paige, S., 1950. Mechanism of landslides. In: Application of Geology to Engineering Practice. Geological Society of America, pp. 83–123.
Vapnik, V., Cortes, C., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
Wang, S.J., Mathew, A., Chen, Y., Xi, L.F., Ma, L., Lee, J., 2009. Empirical analysis of support vector machine ensemble classifiers. Expert Syst. Appl. 36, 6466–6476. doi:10.1016/j.eswa.2008.07.041.
Wang, Y., Sun, D., Wen, H., Zhang, H., Zhang, F., 2020. Comparison of random forest model and frequency ratio model for Landslide Susceptibility Mapping (LSM) in Yunyang County (Chongqing, China). Int. J. Environ. Res. Public Health 17, 4206. doi:10.3390/ijerph17124206.
Whipple, K., Kirby, E., Brocklehurst, S., 1999. Geomorphic limits to climate-induced increases in topographic relief. Nature 401, 39–43. doi:10.1038/43375.
Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., Hong, H., 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena 187, 104396. doi:10.1016/j.catena.2019.104396.
Xu, C., Dai, F., Xu, X., Lee, Y.H., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology 145, 70–80. doi:10.1016/j.geomorph.2011.12.040.
Xu, L., Wang, X., Liu, J., He, Y., Tang, J., Nguyen, M., Cui, S., 2019. Identifying the trade-offs between climate change mitigation and adaptation in urban land use planning: an empirical study in a coastal city. Environ. Int. 133, 105162. doi:10.1016/j.envint.2019.105162.
Yin, A., Dubey, C., Kelty, T.K., Gehrels, G., Chou, C.Y., Grove, M., Lovera, O., 2006. Structural evolution of the Arunachal Himalaya and implications for asymmetric development of the Himalayan orogen. Curr. Sci. 90 (2).
Zhao, Y., Huang, Y., Liu, H., et al., 2018. Use of the Normalized Difference Road Landside Index (NDRLI)-based method for the quick delineation of road-induced landslides. Sci. Rep. 8, 17815. doi:10.1038/s41598-018-36202-9.