Hybrid Machine Learning and Optimum Feature Selection Based Landslide Susceptibility Analysis

Sharma, A.,* Prakash, C., Sharma, A. and Sharma, P.
Department of Civil Engineering, National Institute of Technology, Hamirpur, India
E-mail: amol@nith.ac.in
**Corresponding Author*

Abstract
Landslide susceptibility analysis of Kullu district, India, is carried out using optimum feature selection and Frequency Ratio-Logistic Regression (FR-LR) and Frequency Ratio –Random Forest (FR-RF) hybrid models. 981 historical landslide events of the Kullu district were inventoried using information from various documented sources, high-resolution satellite imagery, and data collected by handheld GPS during field visits were analyzed and used for susceptibility analysis. Based on expert opinions, topography, hydrological setting, and data availability a set of 14 landslide causative factors (LCF) were identified for the study area. A hybrid approach for the selection of optimum LCF has been developed using feature ranking and statistical significance methods. 70% landslide events were used for training and selection of optimum LCFs, and the remaining 30% events were used for the validation of landslide susceptibility mapping. The landslide susceptibility maps (LSM) generated using two machine learning-based hybrid models i.e. FR-LR and FR-RF were compared using ROC curves and confusion matrix. Both FR-LR and FR-RF hybrid models have excellent prediction performance for landslide susceptibility analysis of the study area. Among the models, FR-RF performed better than FR-LR model.

1. Introduction

Landslides are described as the movement of rocks or debris under gravity on the earth surface. The Spatio-temporal occurrence of landslides is modulated by processes such as earthquakes, extreme rainfall, snow/glacier melting, land-use/land cover changes, and various anthropogenic activities which causes vibrations, overburden on soil, uprooting of supports laterally and change in moisture content of soil or rock structures etc. (Cutter et al., 2008, Frangov et al., 2017, Galli et al., 2008, Lei et al., 2020 and Varnes, 1984). The occurrence of landslides is a catastrophic geological natural hazard that imposes serious threats to socio-economic activities, infrastructure, settlements, and human lives in mountainous regions worldwide. The global economic expansion, unplanned haphazard development activities, and aggravated extreme weather events due to ongoing climate change in the mountainous areas have exacerbated the socio-economic impacts of landslides in recent times (Chen et al., 2019 and Li and Chen, 2020). As per the World Health Organisation (WHO), landslides caused more than 18,000 deaths and affected around 4.8 million worldwide from 1998 to 2017.

The Himalayan region of northern India is particularly prone to landslides due to its peculiar topographical conditions and the ever-increasing burden of anthropological activities on its fragile ecosystem (Ghorbanzadeh et al., 2019). Himachal Pradesh has a long history of destruction caused by landslides and faced major landslide disasters for the monsoon season of 2020 (Sharma et al., 2021b). There is a need to improve disaster mitigation and management at the regional level to minimise destructive impacts of landslides by delineating landslide susceptible areas and continuously evaluating such areas with high accuracy and prediction.

Landslide susceptibility analysis is considered the fundamental tool in a competent approach towards landslide hazards assessment, management and mitigation (Luo et al., 2019 and Thi Ngo et al., 2021). Landslide susceptibility mapping is a complex process of establishing interdependence of historical landslide events and topographical, geological, and hydrological variables that influence landslide occurrences on a regional scale (Nguyen et al., 2020).

The various stages of landslide susceptibility analysis include landslide inventory generation, identification of landslide causative factors (LCF), developing a correlation between landslide occurrences and causative factors using a modelling process and validation of the applied models (Chen et al., 2018b, Dehnavi et al., 2015 and Fell, 1993). The landslide susceptibility mapping (LSM)process has evolved over the years from simple expert-knowledge based models like Weight of Evidence (WOE) (Dahal et al., 2007), etc. towards data-driven models such as Evidential Belief Functions (EBF) (Tien Bui et al., 2012) and Certainty Factor (CF) (Devkota et al., 2013 and Zare et al., 2014) etc. These techniques have performed satisfactorily for landslide prediction but sometimes lack functional correlations between LCF. (Chen and Li, 2020 and Li and Chen, 2020). Another drawback of bivariate models is that hypothesis must be accepted before modelling (Chen and Chen, 2021).

The recent advancements in machine learning (ML) algorithms have provided an advanced framework for analysing natural hazards (Arabameri et al., 2019). Traditional ML algorithms include Logistic Regression (LR) (Aditian et al., 2018, Li and Chen, 2020 and Wang et al., 2016), Artificial Neural Networks (ANN) (Gorsevski et al., 2016 and Ngadisih et al., 2016), Decision Tree (DT) (Wang et al., 2016) Support Vector Machine (SVM), (Ada and San, 2018), Fuzzy Logic, Naïve Bayes (NB) (Lee et al., 2019, Lei et al., 2020 and Tsangaratos and Ilia, 2016) and Random Forest (Chen et al., 2018c, 2018b) etc. The ML algorithms can rearrange their internal structure according to landslide data type and have the potential to analyse and update the factor contribution automatically continuously (Saha et al., 2020 and Youssef and Pourghasemi, 2021). However, the results of ML techniques are prone to errors and sometimes lack ease of interpretation concerning the individual contribution of subclasses of LCF. Recently, hybrid machine learning models that integrate multiple statistical or machine learning algorithms have been extensively applied to generate LSM. A few examples of such studies include multi boost-based SVM (Pham et al., 2019a), frequency ratio-based ANFIS model (Dehnavi et al., 2015), Boosted Regression Tree (Saha et al., 2021), Adaptive Neuro-Fuzzy Inference System (ANFIS) (Li and Chen, 2020), etc. The integration of ML approaches with bivariate statistical models helps resolve the drawbacks of both the techniques, increases prediction accuracy and reduces errors. The ML algorithms provide high predictive potential without human intervention, whereas the results of statistical models can be easily correlated with the subclasses of LCF (Fang et al., 2020 and Pham et al., 2016). The FR model is a bivariate statistical model and was chosen because of its ease of application and good accuracy of results (Chen et al., 2019 and Pandey et al., 2018).

This method generates a spatial relationship among variables based on the percent landslide pixels in each factor class, a higher value indicating a higher correlation to landslide occurrence. A study to access landslide susceptibility using the frequency ratio method was carried out by (Pradhan et al., 2014) incorporating nine predictor variables. This study concluded that the FR model is easy to apply and understand and produces better results than similar models. Overall, the FR model for landslide susceptibility mapping is considered adequate and robust with a disadvantage of collecting and validating large datasets and the assumption of conditional independence between different variables, which is generally invalid (Dou et al., 2019). The LR machine learning model develops relationships between dependent and independent variables (Devkota et al., 2013, Li and Chen, 2020 and Yusof et al., 2015). The predictive potential of the LR model has been improved over the years with advanced geospatial and computational technologies. A modified version of the LR model called Kernel Logistic Regression (KLR) was utilized by (Chen et al., 2018a) for landslide susceptibility analysis in which two-class kernel functions were used to evaluate landslide susceptibility. The study concurred that the combination of a decision tree classifier and meta-model improves the overall prediction capability of the model. The RF is a highly accurate ML model whose prediction uses the unification of multiple decision trees to get high accuracy and prediction (Nachappa et al., 2020). A study carried out by (Youssef and Pourghasemi, 2021) compared the predictive potential of seven advanced machine learning models. They concluded that the accuracy of RF model was around 95% and RF model outperformed all other models in its prediction capabilities.

The landslide susceptibility analysis primarily depends on the inventory data and selection of the optimal set of variables influencing landslides in a particular region (Li and Chen, 2020 and Pradhan and Siddique, 2020). For reliable and accurate landslide susceptibility mapping of an area, accurate mapping of landslides and selection of causative factors should be conclusive and logical (Lee and Talib, 2005).

The causative factors must be optimized to select highly correlated factors with landslide occurrences. Factors with very low predictive potential should be avoided for robust modelling algorithms (Dou et al., 2015 and Ghosh et al., 2020). Researchers have used factor analysis, multicollinearity analysis, linear correlation, certainty factor approach, and multi-factor sets techniques for the optimal selection of LCF (Chen and Chen, 2021, Shim and Kim, 2015 and Wang et al., 2015). Most of the time, these techniques were found to be inadequate or highly time-consuming. As such, no standard guidelines are available for the optimal selection of LCF. Hence, there is a need to develop an adequate approach for optimal selection of the Landslide Causative Factors (LCF) to achieve high predictive potential from the applied models within a reasonable time.

The present study focuses on optimal selection of landslide causing factors from a given set of factors using a hybrid approach. In the first step ranking algorithms to determine the feature scores of LCF, then performing k-fold iterative prediction using LR model. Finally evaluating their pairwise statistical significance to find the optimum landslide causative factor subset to be used as an input for modelling algorithms. The study's main objective is to generate susceptibility maps by considering the optimum number of LCFs using a hybrid of a statistical model (FR) and two machine learning models (LR and RF). The performance of two hybrid models, the Frequency Ratio- Logistic Regression (FR-LR) model and the Frequency Ratio- Random Forest (FR-RF) model, is compared using prediction matrices and curves to select the most efficient model. All processes of computation and analysis were carried out in Arc GIS, ERDAS Imagine, R and SPSS software.

2. Regional Settings and Methodology

2.1 Study Area

The Kullu district lies in the central region of Himachal Pradesh and is considered a transitional zone between Lesser and Greater Himalayas. The district extends between 31°20'25" to 31°25v0" N and 76°56'30" to 77°52'20" E with a gross area of 5503 km^2. The altitude varies from 1089m in the western and southern region to 6632m in the eastern region of the district (Figure 1). The average annual rainfall between 2001-2020 is 930 mm with average maximum and minimum temperatures of 28^{0}C and 3^{0}C (https://weathershimla.nic.in/). Geologically, the region is predominated by the Salkhala formation and the Kullu formation and the Rampur group. Major rivers running through the district are Beas, Satluj, Parvati and Tirthan. The region has complex and rugged topography, heavy rainfall during the monsoon season, seismic activity, and extensive anthropological activities .

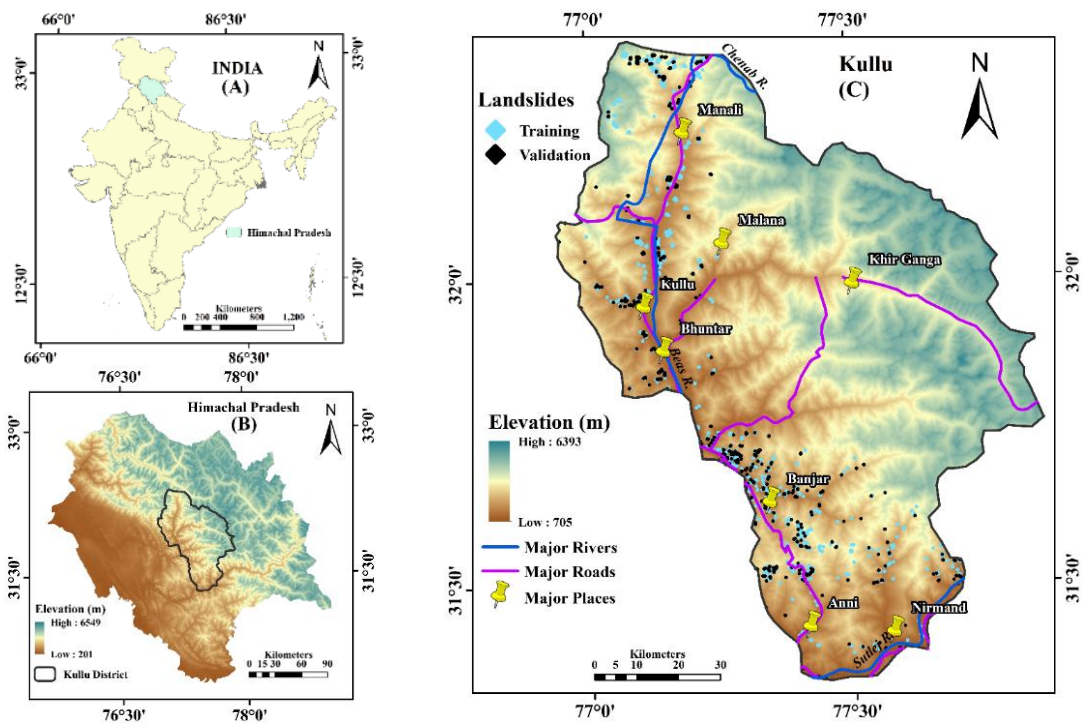

Figure 1: Study Area A. India, B. Himachal Pradesh, and C. Kullu District with Landslide Inventory

2.2 Landslide Inventory Map

The inventory of historical landslide locations can be prepared by visual interpretation, field inspections, satellite images, high-resolution DEM and available data from various documented sources (Bui et al., 2018 and Reichenbach et al., 2018). In this study, the location of historical landslides was identified using disaster and revenue reports, comprehensive field surveys using handheld GPS, and high-resolution Google imagery as an auxiliary data source. The landslide occurrences were identified primarily as rainfall-induced landslides. Out of the total 981 landslides, the maximum and minimum area of landslides was 2.9 x 10^5 m^2 and 25 m^2, respectively. The landslide inventory dataset was randomly split for 70 % (687) training and 30% (294) validation purposes as suggested in the previous studies by (Ali et al., 2021 and Roy et al., 2019). The landslide inventory map is resampled in ArcGIS and is shown in (Figure 1).

2.3 Landslide Causative Factors (LCF)

After generating a landslide inventory map, the LCFs were identified based on the Kullu district's geological, meteorological and topographic conditions (Versain, 2019). Fourteen LCF's, i.e. elevation, slope gradient, slope aspect, curvature, drainage density, geology, lithology, normalized difference vegetation index (NDVI), land use and landcover (LULC), soil, lineament density, topographic wetness index (TWI), stream power index (SPI) and distance from the roads were selected for further analysis. The details of these data sources are provided in (Table 1). ALOS-PALSAR Digital Elevation Model (DEM) of 12.5m resolution has better accuracy of prediction than SRTM and ASTER DEMs (Arabameri et al., 2019). Hence, it was the preferred source to generate the terrain characteristics of the study area. The slope gradient is considered one of the crucial factors as it quantifies the amount of stress and gravitational force acting on a specific area of sliding (Dehnavi et al., 2015). The slope map of the study area demonstrates variation and is subdivided into five classes using natural breaks classification. The slope rises from flatter terrains along the valleys towards steep gradients at higher elevations. The slope aspect directly relates to discontinuities, rock weathering, soil/rock moisture, which affects the landslide occurrences (Galli et al., 2008 and He et al., 2019). The drainage density directly influences the erodibility of slopes dissected by channels and influences the geological system and surface runoff (Demir et al., 2014 and Kavoura and Sabatakakis, 2019). The presence of rivers Beas, Tirthan and Parvati, along with their tributaries, drains the northern and central region of the district while the Satluj river fringes the southern boundary of the district. TWI is the steady-state wetness index used to quantify the topographic influence on the hydrology of an area.

Table 1: Data sources of landslide causative factors

Data	**Data Purpose**	**Data Source**	**Scale / Resolution**
Kullu District Administration	Kullu District Boundary	https://hpkullu.nic.in/map-of-district/	1:50000
State Disaster Management Authority (H.P), GPS, Google Earth	Landslide Inventory	https://hpsdma.nic.in/	1:50000
Digital Elevation Model ALOS-PALSAR	Elevation, Gradient Curvature, Drainage Density, TWI, SPI	https://search.asf.alaska.edu	12.5 m
Multispectral Landsat-8 Images	NDVI, LULC and Lineaments	http://earthexplored.usgs.gov	30 m
Geological Survey of India (GSI)	Geology, Lithology	https://bhukosh.gsi.gov.in/	1:50000
National Highway Authority of India (NHAI)	Major roads of Kullu District	https://morth.nic.in/	1:50000
National Bureau of Soil Survey and Land Use Planning (ICAR-NBSS and LUP)	Soil Classification and Properties	https://www.nbsslup.in/	1:50000

TWI map was generated using flow accumulation and direction parameters in Equatio 1 (Li and Chen, 2020 and Shahabi et al., 2014). Stream Power Index (SPI) map is used to quantify the erosive capabilities of flowing water on a particular terrain and was calculated using Equation 2 (Sahin et al., 2020). The natural breaks classification has higher accuracy than quantile and geometric mean classification (Arabameri et al., 2019 and Chen and Chen, 2021). Using the natural break classification method in ArcGIS, these hydrological parameters were categorized into five classes: very low, low, moderate, high, and very high (Pandey et al., 2018 and Yi et al., 2019).

$$\mathrm{TWI} = \frac{\mathrm{Ln}\,(\mathrm{As})}{\mathrm{Tan}\,(\beta)}$$

Equation 1

$$\mathrm{SPI} = \mathrm{As} \times \mathrm{Tan}(\beta)$$

Equation 2

Where ln is natural log, F_a = Flow Accumulation, α = Slope in Radians

Earth Explorer website (http://earthexplore-d.usgs.gov) of the United States Geological Survey (USGS) was used to procure Landsat-8 multispectral satellite images having 30-m resolution. As maximum landslide in the Kullu district occurs during the monsoon season, the multispectral images were obtained just before and after monsoon season as they were deemed suitable for this study. These images were used for generating NDVI, LULC and lineament density maps using ERDAS Imagine software. NDVI is a dimensionless index used to quantify the green vegetated area over a terrain (Brodley and Friedl, 1997 and Bhatti and Tripathi, 2014). The NDVI values were obtained using Equation 3. Low or negative values of NDVI (-0.22 to 0.12) typically suggests waterbodies or urban areas, while medium to high NDVI values (0.12 to 0.55) signify sparsely vegetated land and forest land (Deijns et al., 2020). LULC is a characterization of an area based on anthropological activities as well as natural elements of a landscape. The degradation of natural landscape due to extensive infrastructure development and deforestation leads to soil erosion (Glade, 2003 and Reichenbach et al. 2014). The LULC map of the Kullu district was categorized as water bodies, forest, vegetation cover, barren land and urban area. The lineament density of an area is defined as the total length of all the contours divided by the area under study. High lineament density often indicates a higher probability of slope failure (Saha and Saha, 2020 and Tripathi et al., 2000).

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{RED})}{(\mathrm{NIR} + \mathrm{RED})}$$

Equation 3

Where NIR (Near Infra-Red) and RED are the electromagnetic spectrum bands.

An area's lithology represents the properties of outcrop rocks, whereas geology represents Earth's changing rock formations (Nefeslioglu et al., 2008 and Dou et al., 2015b). The Salkhala and Kullu formation, followed by the Rampur group, predominates the study area's geology. The rock types in the district are phyllite, quartzite, slate, limestone, schists and granites (Patel et al., 2011). A soil's drainage properties and its depth are associated with its degree of erodibility (Baum and Godt, 2010, Merghadi et al., 2018 and Pham et al., 2016). The soils of the study area were categorized into seven classes on the basis of depth, erosion and drainage properties. Landslides are predominant in areas near roads as their construction in mountainous regions often includes excavation of natural bed slopes. This results in loss of support and cracks development and change in natural drainage corridors of the region (Pradhan et al., 2018). The study area's major roads were identified and divided into six buffer classes of (0-100, 100-200, 200-300, 300-400, 400-500 and >500 m). These LCF's were resampled into a standard raster format of 30 x 30 m in ArcGIS environment (Figure 3).

Selecting a relevant subset of LCF is paramount for generating an efficient and robust ML model and minimizing the independent variables (Dou et al., 2015). There is no accepted methodology for selecting appropriate LCF; most research studies select factors randomly or use expert opinions. Also, the availability of data in mountainous regions is also a matter of concern (Arabameri et al., 2019 and Jebur et al., 2014). The selection of reliable factors from a given dataset reduces the effects of overfitting the model (Cigdem and Demirel, 2018). The first step in this process is feature ranking, in which the influence of each factor on landslides is ranked in an ordered manner using a statistical scoring function (Santos et al., 2014). The next step in feature selection is to determine the optimum subset of factors using statistical significance tests. A hypothesis can only be accepted if the level of significance (usually summarized by a probability and denoted by p-value) is less than 0.05.

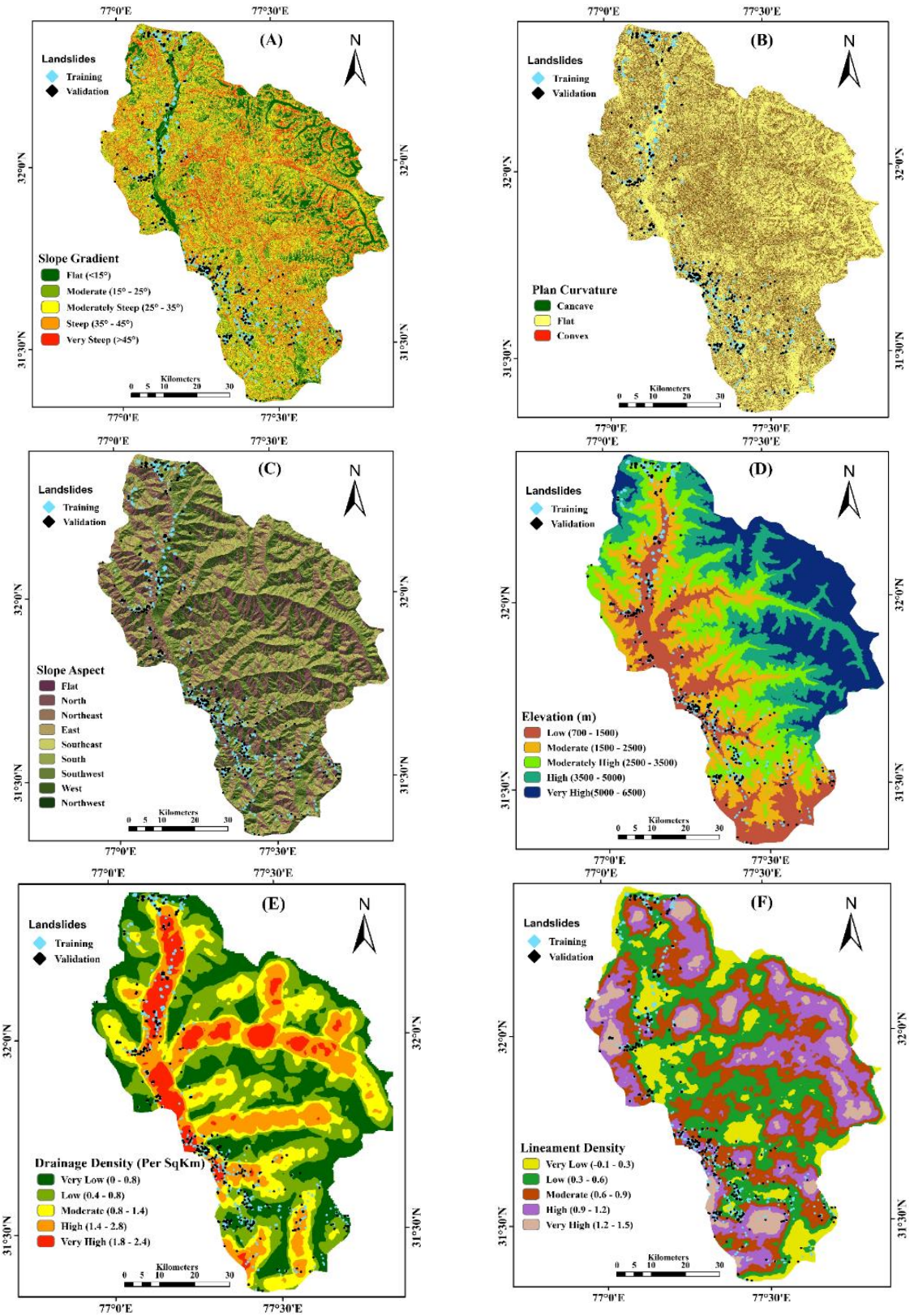

Figure 3: Landslide Causative Factors A. Slope Gradient B. Plan Curvature, C. Slope Aspect, D. Elevation, E. Drainage Density, F. Lineament Density, G. Geology, H. NDVI, I. Soil, J. Distance from Roads and K. TWI (Continue next page)

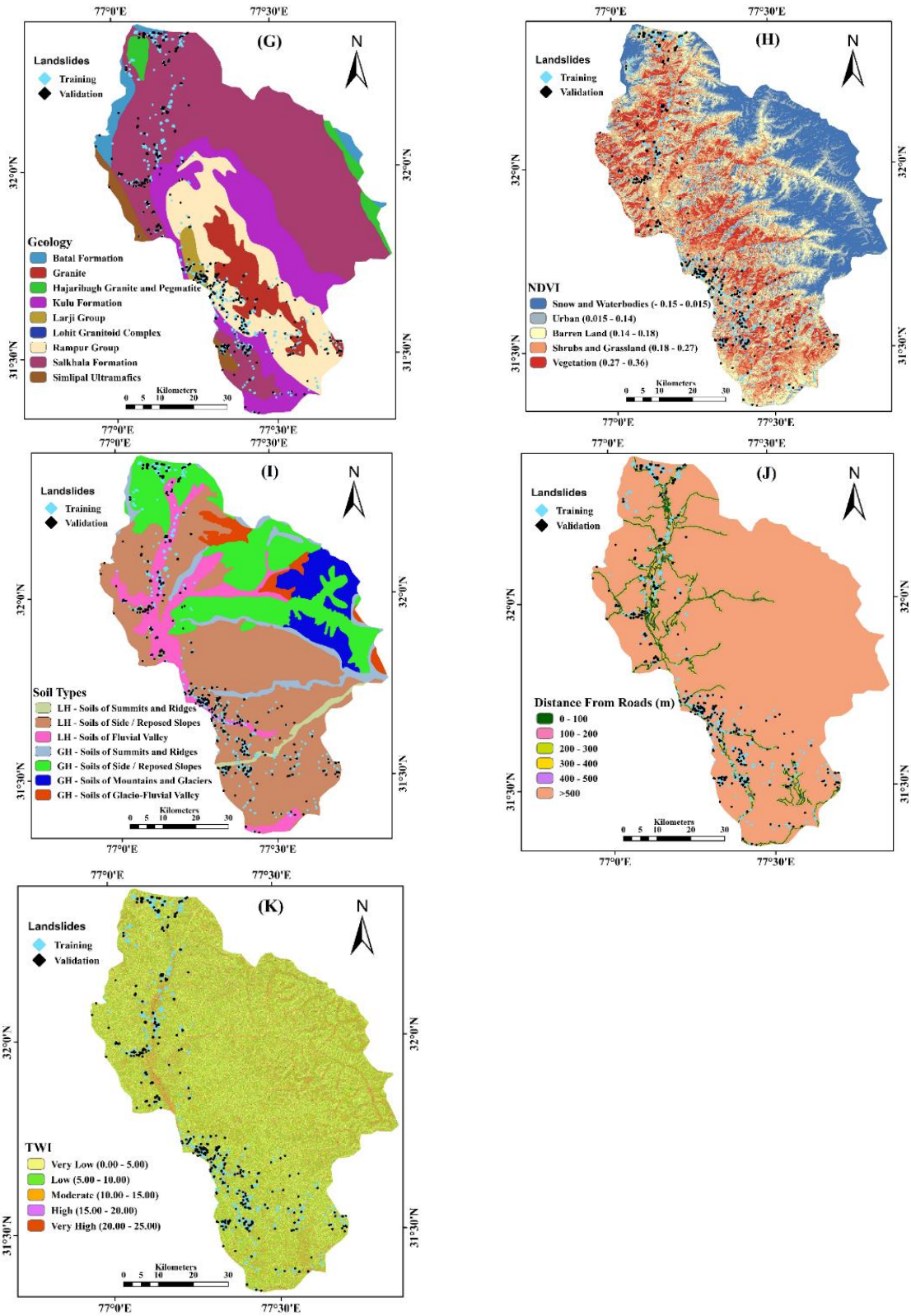

Figure 3: Landslide Causative Factors A. Slope Gradient B. Plan Curvature, C. Slope Aspect, D. Elevation, E. Drainage Density, F. Lineament Density, G. Geology, H. NDVI, I. Soil, J. Distance from Roads and K. TWI

The crossfold validation procedure was used to input multiple datasets into k-folds using repetitive random sampling. A final optimum set of LCF's was then obtained having a minimum set of parameters having maximum prediction capabilities (Merghadi et al., 2020 and Sahin, 2020).

2.4 Frequency Ratio (FR) Model

The FR model is an observation-based model which defines the quantitative spatial correlation between the distribution of landslides in an area with a set of independent LCF (Arabameri et al., 2019 and). FR is defined as the ratio of the pixel data with and without landslides with pixels of input raster data layers of causative factors. The FR value is computed for each class of causative factors using Equation 4. The correlation is high if the FR value is greater than 1, while the FR value less than 1 indicates a lower correlation with that particular causative factor (Chang et al., 2020).

$$FR(i) = \frac{Npix(li) / Npix(ci)}{\sum Npix(li) / \sum Npix(ci)}$$

Equation 4

Npix(li) = Total pixels containing landslides in class (i)
Npix(ci) = Total pixels in each class (i)
∑Npix(li) = Sum of landslide containing pixels
∑Npix(ci) = Sum of pixels in the whole study area

2.5 Logistic Regression (LR) Model

The LR model uses a logit (sigmoid) function to generate a correlation between a categorical dependent variable and a set of independent variables (Merghadi et al., 2020). The LR is widely used for LSM due to its capacity to generate a precise relationship function while introducing minimal hyper-parameters for a large sample size. The disadvantage of the LR model is that it fails to solve complex relationships due to its linear characteristics and sometimes leads to overfitting the model (Budimir et al., 2015). Quantitatively, the LR model associates landslide occurrence probability to several LCF using Equation 5.

$$\text{Logistic (P)} = \frac{1}{1 + e^{-z}}$$

Equation 5

Where P is the probability of an event within the range [0,1], z is the linear combination function that fits logistic regression Equation 6.

$$Z = b_1X_1 + b_2X_2 + \cdots b_nX_n$$

Equation 6

Where λ_i represents the partial coefficients and X_i represents the independent landslide causative variables.

2.6 Random Forest (RF) Model

The RF model is a popularly used model that uses a decision tree (DT) supervised classification algorithm and combines multiple trees and branches with random data. The RF model is considered a robust model and can be used in regression and classification problems. For regression problems, RF generates output from the mean of DT whereas for classification problems, RF provides the most probable class output (Breiman, 2001 and Lee et al., 2017). The RF model requires optimization of hyperparameters for optimal prediction performance (Catani et al., 2013 and Sharma et al., 2021a). It provides high prediction accuracy and can reduce the risk of overfitting the model but has a disadvantage of constrained interpretability of decisions (Dou et al., 2019). In the present study, three hyperparameters i.e. number of trees (ntree), terminal node size and the number of variables tested at each node (mtree) were taken as 250, 5 and 5, respectively.

3. Methodology

The methodology in the present study is based on the integration of multi-dimensional information acquired from numerous sources to generate the topographical, hydrological, geotechnical, environmental and anthropological characteristics with landslide occurrence. The primary landslide data was generated from visual interpretation through high-resolution Google Earth images and analysing the terrain characteristics derived from ALOS PALSAR DEM. The published landslide inventories from HPSDMA, GSI and NASA along with numerous newspaper articles and previous landslide studies of the study area. A set of fourteen LCF was selected with due consideration to expert opinions, published literature and data availability. The thematic layers were procured and converted to a standard raster resolution of 30x30 m. These factors were first checked for multicollinearity to identify any correlation between independent variables.

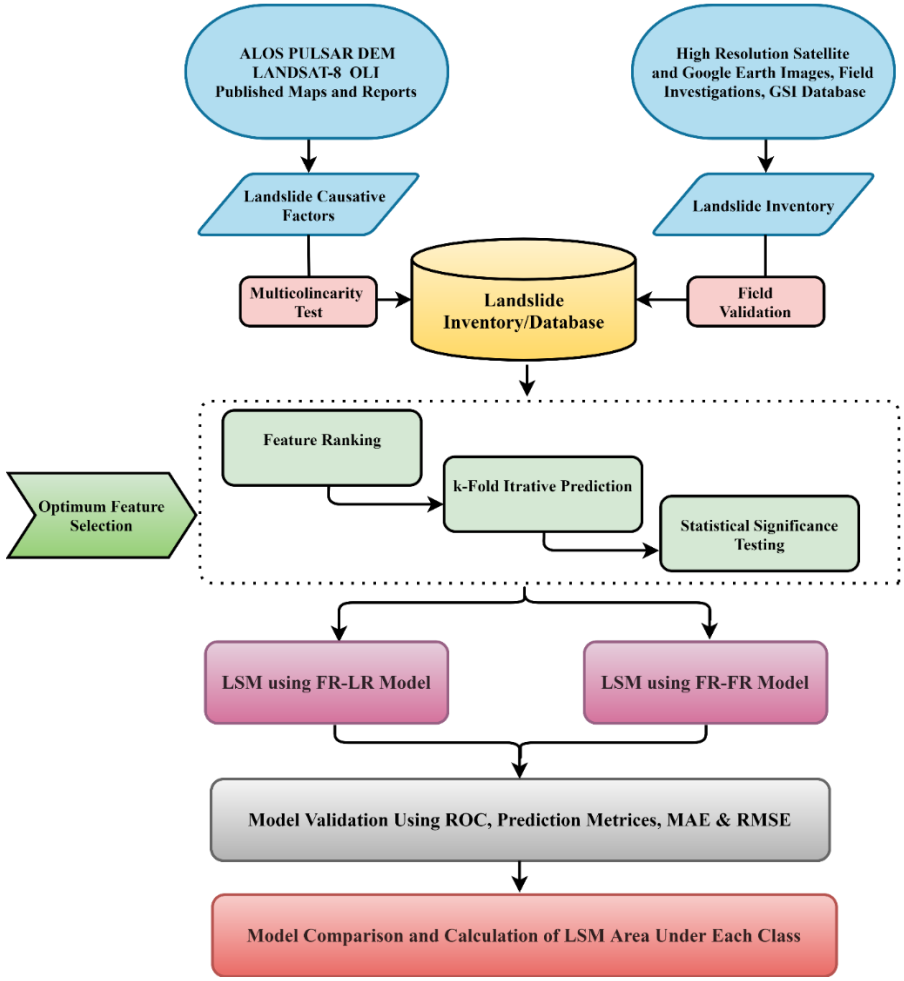

Figure 2: Flowchart depicting methodology used in the study

The optimum selection of landslide causative factors was carried out using three ranking algorithms: chi-squared, random forest feature importance and information gain. The landslide inventory dataset was resampled using k-fold cross-validation resampling procedure (Sahin et al., 2020). The hybrid models were generated by integrating a widely used statistical bivariate FR model with LR and RF machine learning models. The LR and RF are powerful ML algorithms that have high predictive potential and can classify non-linear data with minimum overfitting. The RF is an ensemble tree-based approach and is known to classify spatial data with high accuracy. The goal of integrating such models is to generate a robust model with higher predictive potential than the individual models and to generate an accurate spatial relationship among classes of LCF and landslide occurrence. The optimum selected factors were used as an input for hybrid integration of (FR-LR) and (FR-RF) models. Their results were analysed for spatial correlation of LCF and landslide occurrences. The performance of two hybrid models was using ROC curves and evaluating confusion matrix. The detailed methodology is shown in (Figure 2).

4. Results

4.1 Analysis of Multicollinearity

Multicollinearity refers to an interrelationship between two or more independent variables resulting in skewness in model results. Any relationship between these factors can decrease the model's predictability and generate an error in results (Chen and Chen, 2021). The multicollinearity analysis was carried out on a set of 14 LCSF's and the results are given in (Table 2). Generally, Variance Inflation Factor (VIF) values < 10 and tolerance (TOL) values > 0.1 are acceptable for no multicollinearity among variables (Arabameri et al., 2019).

Table 2: Analysis of multicollinearity

Model	Coefficients	
	Tolerance	VIF
TWI	0.698	2.451
Elevation	0.775	6.165
Curvature	0.514	8.793
Aspect	0.231	3.037
LULC	0.282	8.525
Lineament Density	0.149	6.785
Lithology	0.419	5.891
Geology	0.336	2.976
NDVI	0.131	7.631
Drainage Density	0.313	0.417
Soil	0.472	5.270
Distance from Roads	0.287	1.638
Slope	0.358	2.357
SPI	0.812	4.686

Table 3: Feature ranking algorithms and relative feature scores

Chi-Squared		Random Forest Feature Importance		Information Gain	
TWI	0.4774	Drainage Density	124.96	TWI	0.2031
Elevation	0.4157	TWI	111.94	Distance to Roads	0.1232
Distance to Roads	0.3298	Slope Gradient	105.32	Elevation	0.1158
Drainage Density	0.3201	Distance to Roads	102.68	Drainage Density	0.1156
Slope Gradient	0.2755	Slope Aspect	81.64	Plan Curvature	0.0971
Geology	0.2498	Elevation	70.61	Geology	0.0623
SPI	0.2123	Lineament Density	69.33	SPI	0.0489
NDVI	0.1879	Geology	65.73	NDVI	0.0357
Soil	0.0985	Plan Curvature	62.88	Slope Gradient	0.0202
Slope Aspect	0.0895	NDVI	57.73	Lineament Density	0.0149
LULC	0.8057	LULC	45.25	Lithology	0.0114
Lineament Density	0.0754	Soil	33.35	Soil	0.0047
Plan Curvature	0.0235	Lithology	9.15	Slope Aspect	0.0021
Lithology	0.0156	SPI	7.149	LULC	0.0018

The results indicate that the maximum and minimum TOL values were obtained for TWI (0.812) and distance from roads (1.638), respectively. Further, all 14 LCF had acceptable TOL and VIF values and were deemed independent and suitable for modelling.

4.2 Feature Selection Process

The present study incorporates three feature ranking and three statical significance testing models to determine an optimum subset containing relevant features. The outputs produced by these three ranking methods are shown in (Table 3). The feature ranking results indicate different feature scores for different variables. An iterative process was carried out using LR predictive model until the addition of a factor had no significant difference in the prediction performance of the statistical tests. The value of the k-fold parameter, which defines the number of splits of the input dataset, was taken as 10. These ranking factors were then tested for statistical significance, and pairwise comparison was made. The results of all the possible scenarios are shown in (Table 4). It was analysed that Model-11 had the same characteristics as Model-12 and Model-13, with more significant landslides causative factors. Hence, Model-11 was taken as the cut-off point. The Information Gain feature ranking algorithm was deemed suitable, and Wilcoxson signed-rank test was used to check statistical significance to evaluate the performance LR prediction algorithm. Finally, Case-9 with Model-11 was chosen as the optimum subset to generate LSM of the study area as it contained all relevant features and was used for further analysis.

Table 4: Optimum feature subset using feature selection process

Feature Ranking Methods	Case No.	Statistical Tests	Model and Subset Size	Excluded Features
Chi-Squared	Case - 1	F- Test	Model - 11	Curvature; Lineament Density; Slope;
	Case - 2	One-Sample T-Test	Model - 12	Slope Aspect; Lithology
	Case - 3	Wilcoxon Signed-Rank Test	Model - 11	Geology; NDVI; Lineament Density;
RF-Importance	Case - 4	F- Test	Model - 10	Drainage Density; TWI; Distance from Roads Soil;
	Case - 5	One-Sample T-Test	Model - 13	Curvature
	Case - 6	Wilcoxon Signed-Rank Test	Model - 11	LULC; Aspect; Curvature
Information Gain	Case - 7	F- Test	Model - 9	Slope; Drainage Density; Geology; Distance from Roads; NDVI
	Case - 8	One-Sample T-Test	Model - 12	Aspect Curvature
	Case - 9	Wilcoxon Signed-Rank Test	Model - 11	Lithology; LULC; SPI

4.3 Modelling Landslide Susceptibility

The FR-LR and FR-RF hybrid models were used to generate landslide susceptibility maps of the Kullu district. The spatial dependence of historical landslides and factors influencing them is given in (Table 5). The final susceptibility maps were split into five subclasses using natural breaks classification (Figure 4). The percentage of area in each subclass is shown in (Table 6). According to LSM generated using the FR-LR model, 10.68 % of the area lies in very high, and 23.89% lies in high susceptibility zones. Similarly, for the FR-RF model, 11.32 % per cent area lies in very high, and 18.19% area lies in high susceptibility zones.

The analysis of results in (Table 5) generated using the FR-LR model indicates a high correlation between slope gradient, drainage density, TWI, geology, road buffer and landslide occurrences. The moderately steep (25°-35°) and steep (35°-45°) slopes especially facing south and southwest direction are highly susceptible to landslides. The model also suggested that very high drainage density (1.8-2.4) and TWI (20-25) are the favourable hydrological parameters for landslides occurrence. Geologically, the Larji group consisting of gneiss, quartzite and schist and excessively drained loamy-skeletal soils of Lesser Himalayas (LH) positively correlate with landslides. It was found that the maximum probability of landslide occurrence was near (0-100m) road buffer and decreased with increasing distance from roads.

The LSM map generated using the FR-RF model is analysed for feature contribution using the FR values and Mean Gini index values. The results implied that drainage density, TWI, elevation, road buffer and NDVI have a high correlation. The study area's low (700–1500m) and moderate (1500-2500m) elevations have more landslide susceptibility in comparison to higher elevations. As higher elevations of the study area consist of mountain ranges, therefore, lesser landslide events are generally reported in such areas. The NDVI values suggest that urban areas (0.015-0.14) and barren lands (0.14-0.18) are more landslide-prone. The rest of the parameters, i.e., geology, lineament density and plan curvature etc., have moderate to low correlation with landslide occurrence.

4.4 Performance Evaluation and Comparison of Models

Any novel machine learning approach for generating LSM is not acceptable until the results are validated using relevant statistical methods (Ghosh et al., 2020 and Saha et al., 2021). In this study, the performance of both hybrid models was based on accuracy, precession, recall, F1, MAE, RMSE and Kappa values which were calculated using the confusion matrix.

Table 5: Spatial relation between landslide frequency and causative factors

	Class Pixels	Percent Pixels	Landslide Pixels	Percent Pixels	Frequency Ratio	Logistic Regression	Random Forest
					FR Values	LR Value	Mean Gini
Slope Gradient (Degree)							
Flat	411287	0.067	32.00	0.009	0.138		
Moderate	927360	0.151	382.00	0.110	0.730		
Moderately Steep	1595709	0.259	1001.00	0.288	1.111	4.068	247.279
Steep	1731153	0.281	1278.00	0.368	1.308		
Very Steep	1486151	0.242	779.00	0.224	0.929		
Plan Curvature							
Convex	1803273	0.293	1063	0.306	1.044		
Flat	2566679	0.417	1500	0.432	1.035	-2.131	108.254
Convex	1781708	0.290	909	0.262	0.904		
Slope Aspect							
Flat	32281	0.005	18	0.005	0.988		
North	692197	0.113	164	0.047	0.420		
North-East	722418	0.117	160	0.046	0.392		
East	680725	0.111	299	0.086	0.778		
South-East	763997	0.124	239	0.069	0.554	-5.457	121.356
South	798563	0.130	1059	0.305	2.350		
South-West	906111	0.147	892	0.257	1.744		
West	803628	0.131	210	0.060	0.463		
North-West	751740	0.122	431	0.124	1.016		
Elevation (m)							
Low (400 - 1000)	401095	0.065	470	0.135	2.076		
Moderate (1000 - 1500)	1507181	0.245	1710	0.493	2.010		
Moderately High (1500 - 2000)	1606521	0.261	528	0.152	0.582	-1.076	658.578
High (2000 - 2500)	1336897	0.217	699	0.201	0.926		
Very High (2500 - 3500)	1299966	0.211	65	0.019	0.089		
Drainage Density							
Very Low	1052472	0.171	328	0.094	0.552		
Low	2666622	0.433	817	0.235	0.543		
Moderate	1659702	0.270	429	0.124	0.458	-2.076	868.159
High	683898	0.111	1501	0.432	3.889		
Very High	88966	0.014	397	0.114	7.906		
Lineament Density							
Very Low	792817	0.129	632	0.182	1.412		
Low	1701868	0.277	243	0.070	0.253		
Moderate	1799995	0.293	1170	0.337	1.152	-1.599	185.988
High	1358767	0.221	1339	0.386	1.746		
Very High	498213	0.081	88	0.025	0.313		
Geology							
Larji Group	85894	0.014	228	0.066	4.703		
Salkhala Formation	3085259	0.502	2710	0.781	1.556		
Batal Formation	160607	0.026	10	0.003	0.110		
Hajaribagh Granite and Pegmatite	163203	0.027	115	0.033	1.248		
Rampur Group	1076163	0.175	214	0.062	0.352		
Granite	418090	0.068	60	0.017	0.254		
Simlipal Ultramafics	124779	0.020	16	0.005	0.227	3.521	279.611
Kulu Formation	1036990	0.169	119	0.034	0.203		
Lohit Granitoid Complex	675	0.000	0	0.000	0.000		
Larji Group	85894	0.014	228	0.066	4.703		
Salkhala Formation	3085259	0.502	2710	0.781	1.556		
Batal Formation	160607	0.026	10	0.003	0.110		
Hajaribagh Granite and	163203	0.027	115	0.033	1.248		

Pegmatite							
Rampur Group	1076163	0.175	214	0.062	0.352		
NDVI							
Waterbodies	1299601	0.211	219	0.063	0.299	-2.076	549.439
Urban	1136606	0.185	1419	0.409	2.212		
Barren Land	1362282	0.221	1043	0.300	1.357		
Shrubs and Grassland	1461005	0.237	594	0.171	0.720		
Sparse Vegetation	892166	0.145	197	0.057	0.391		
Dense Vegetation	1299601	0.211	219	0.063	0.299		
Soil Characteristics							
(LH) Soils of Side Slopes	3102536	0.50	1058	0.305	0.60		
(GH) Soils of Side / Reposed Slopes	1378935	0.22	1038	0.30	1.33		
(GH) Soils of Mountains and Valley	397621	0.06	0	0.00	0.00		
(GH) Soils of Fluvial Valleys	185131	0.03	0	0.00	0.00	-0.007	137.86
(GH) Soils of Summits and Ridges	323013	0.05	0	0.00	0.00		
(LH) Soils of Fluvial Valleys	686533	0.11	1364	0.39	3.52		
(LH) Soils of Summits and Ridges	77891	0.01	12	0.01	0.27		
TWI							
Very Low	1719028	0.28	849	0.245	0.88	8.943	778.059
Low	2369194	0.39	326	0.09	0.24		
Moderate	1433577	0.23	1035	0.30	1.28		
High	527897	0.09	872	0.25	2.93		
Very High	101964	0.02	390	0.11	6.78		
Distance to Roads (m)							
0 - 100	230150	0.04	1381	0.398	10.63	6.705	634.984
100 - 200	178367	0.03	379	0.11	3.76		
200 - 300	151784	0.02	226	0.07	2.64		
300 - 400	136214	0.02	227	0.07	2.95		
400 - 500	124618	0.02	111	0.03	1.58		
>500	5330527	0.87	1148	0.33	0.38		

Table 6: Class Wise Distribution of landslide susceptible areas

Model	**FR-LR**		**FR-RF**	
	Area (%)	**Area (Km^2)**	**Area (%)**	**Area (Km^2)**
LSM Class				
Very Low	20.11	1106.67	39.27	2160.82
Low	20.59	1133.31	15.10	831.20
Medium	24.72	1360.22	16.11	886.63
High	23.89	1314.91	18.19	1001.22
Very High	10.68	587.89	11.32	623.12
Total	100	5503	100	5503

The values of these performance metrics and errors are given in (Table 7). The visual interpretation of the accuracy of models was based on ROC curves and their AUC values (Figure 5). The outcomes of this evaluation confirm that both FR-LR and FR-RF models have good potential to predict landslide susceptibility. The FR-RF model had better performance with (AUC= 0.941) as compared to FR-LR model with (AUC= 0.902). The accuracy (0.891) and precision (0.964) values of the FR-RF model are again higher than the accuracy (0.874) and precision (0.927) values of FR-LR. The Kappa index values that measure the reliability of the model higher for FR-RF the value is (0.671) compared to the FR-LR model (0.625).

Table 7: Comparison of Models using Prediction Metrices

Model	Accuracy	AUC Prediction	AUC Validation	Kappa	Precision	Recall	F1	MAE	RMSE
FR_LR	0.8743	0.8910	0.9024	0.6258	0.9270	0.7900	0.8361	0.1287	0.3587
FR_RF	0.8917	0.9308	0.9413	0.6715	0.9641	0.8298	0.8794	0.1143	0.3381

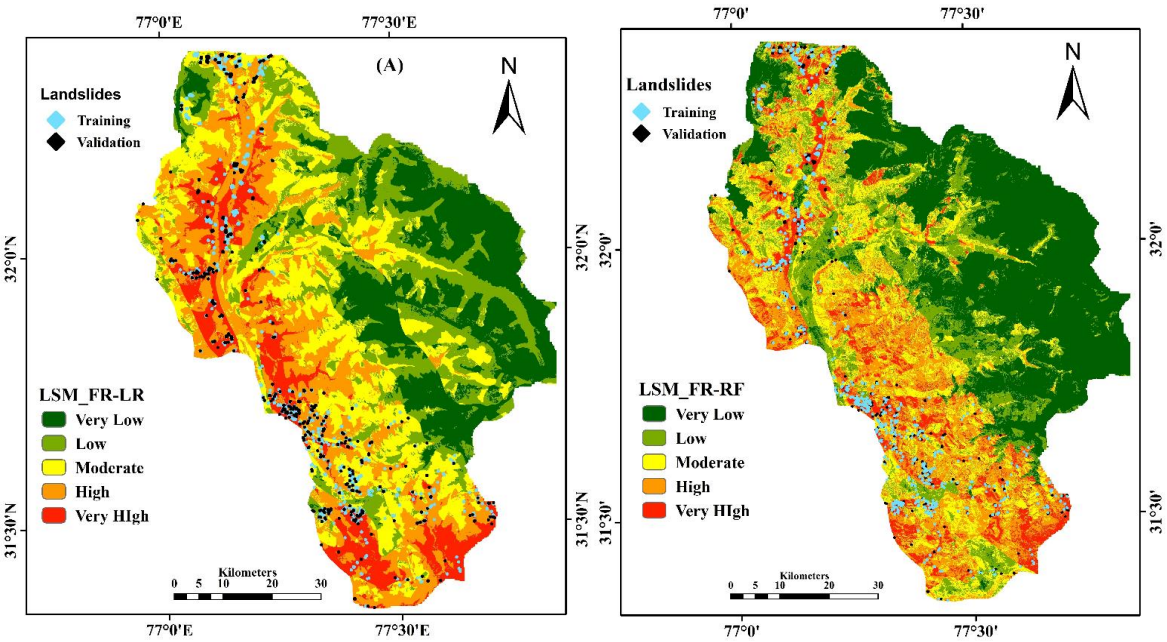

Figure 4: Landslide Susceptibility Maps A. FR-LR Model and B. FR-RF Model

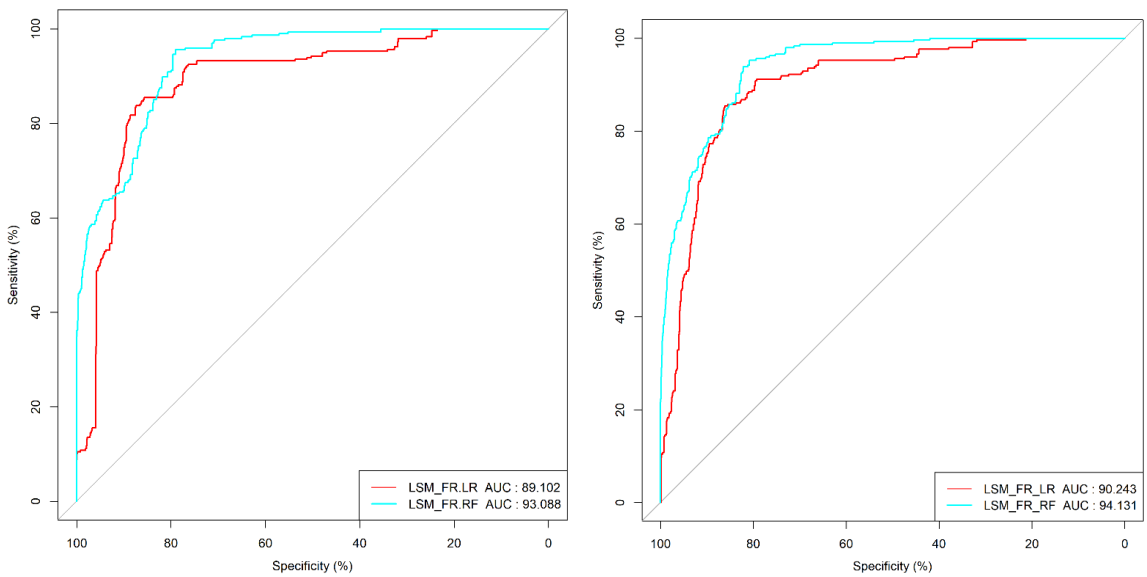

Figure 5: ROC curves with AUC values for accuracy of prediction and validation of models

The recall values indicate that the non-landslide areas are better identified by the FR-RF model (0.829) than the FR-LR model (0.790). The analysis of errors indicated that the FR-RF model MAE and RMSE values (0.114 and 0.338) were lower than the FR-LR model (0.128 and 0.358).

Based on the performance evaluation strategy results, it can be stated that even though both FR-RF and FR-RF models have good values of prediction metrics, the FR-RF model outperformed the FR-LR model in its prediction capabilities.

5. Discussion

The analysis of landslide susceptibility of an area is crucial in landslide hazard management. The continuous updating of the LSM of a high-risk area using modern algorithms and approaches to increase assessment accuracy and decrease the risk potential of landslide hazards is of utmost importance (Saha et al., 2021). The selection of LCF affects the predictive potential of the LSM model (Li and Chen, 2020 and Romer and Ferentinou, 2016). The current trend of generating LSM using hybrid bivariate and machine learning approaches helps to eliminate the limitations of both techniques (Pandey et al., 2018 and Saha and Saha, 2020). However, due to the lack of consistency in the results of various hybrid models, further investigations regarding the integration of these approaches are needed (Wang et al., 2020). An optimum subset of 11 LCF from 14 factors using the feature selection process was chosen based on feature ranking and statistical significance results. Two hybrid models, namely, FR-LR and FR-RF, were used to generate LSM of Kullu district, Himachal Pradesh. The FR statistical model was used due to its high accuracy and ease of result interpretation, whereas LR and RF algorithms were used for their higher prediction potential and ability to model complex non-linear interactions between the variables (Arabameri et al., 2020, He et al., 2019 and Saha et al., 2021).

The results of hybrid FR-LR and FR-RF models implied that the hydrological parameters such as drainage density and TWI are the most influential landslide causative parameters for analysing the landslide susceptibility of an area. The increased anthropogenic activities, particularly slope excavation for road construction, increase the study area's landslide susceptibility (Chen and Chen, 2021, Li and Chen, 2020). The Beas River is the predominant river of the district, and the portion of under construction NH-154 (Ner Chowk to Manali) runs alongside this river resulting in excessive drainage near-road network of the study area. This has led to an increased landslide activity in the vicinity of both due to complex hydrological and anthropological interactions. The visual inspection of both LSM maps indicates that the study area's north, eastern, and southern parts with lower to moderate elevations, southward facing slope and slope steepness of 25° to 45° have higher correlation landslides. In contrast, many parts of the eastern region of the Kullu district had lower landslide susceptibility due to higher elevations and low density of drainage networks. The geological formations of the Larji group and soils of fluvial valleys having high permeability and lower depth were found to be highly landslide susceptible. The remaining variables, like lineament density, plan curvature etc., have minimal influence on landslide occurrences.

The performance of the two-hybrid models was compared based on accuracy, precision, F1, MAE, RMSE kappa and ROC curves etc. Although both models have excellent accuracy, the ROC curves suggest that the FR-RF hybrid model outperformed the FR-LR hybrid model in terms of accuracy of prediction and reduction in errors relatively. The hybrid integration of FR statistical and LR and RF machine learning models increased the model's accuracy and reduced errors in both the hybrid model as compared to previous studies which have applied these models individually. Pradhan and Kim (2017) concluded that the best fit FR model could have a maximum accuracy of around 80%. Similarly, a study carried out by (Tsangaratos and Ilia, 2016) found the accuracy of LR model to be around 84%. In a similar study carried out by (Sahin, 2020) predicted the accuracy of RF model close to 88%. Hence it is evident that the process of integrating bivariate and ML models increases the accuracy and predictive potential of the models. In this study, the FR-RF model shows +3.89 % AUC validation, +1.7 % accuracy. +3.71% precision and +4.5 % kappa improvements compared to the FR-LR model. These results are consistent with previous similar studies (Arabameri et al., 2019b, Arca et al., 2019, Chang et al., 2020, He et al., 2019, Li and Saha, 2020; Pham et al., 2019b, Pourghasemi et al., 2020 and Saha and Saha, 2020).

6. Conclusions

The ever-increasing landslide incidences in the Kullu district of Himachal Pradesh calls for continuous evaluation of landslide susceptibility in the region. The advancement in remote sensing and machine learning techniques is crucial in accurately predicting landslide-prone areas. The present study provides an upgraded landslide susceptibility map by hybrid integration of statistical (FR) and ML (LR and RF) algorithms. A total of 981 landslide incidences were recorded and grouped into a training dataset of 687 (70%) and a validation dataset of 294 (30%). Fourteen LCF were assessed through the multicollinearity and feature selection process, and 11 optimum factors were selected for generating LSM using hybrid FR-LR and FR-RF models.

The comparison of the results of the individual model confirmed that the FR-RF model outperforms the FR-LR model with an overall 4.5% better accuracy. The result of the present proposed model indicates that the process of feature selection minimized the errors and reduced the overfitting of models. The integration of bivariate and ML models increases the predictive potential of individual models, and more focus should be given to such integration of models in areas with a similar regional setting. In the Kullu district of Himachal Pradesh, excessive drainage, steep slope angles, and road construction were identified as primary factors for landslide occurrences. It is advisable that futuristic modelling of such areas can be checked by incorporating time series analysis based on hydrological and anthropogenic data. Also, the feature selection process could be refined by incorporating new indices for better performance and faster computation of the models. Finally, this study will be beneficial for the mitigation of landslide hazards and the planning of developmental activities in similar mountainous regions.

References

Ada, M. and San, B. T., 2018, Comparison of machine-Learning Techniques for Landslide Susceptibility Mapping Using Two-Level Random Sampling (2LRS) in Alakir Catchment Area, Antalya, Turkey. *Natural Hazards*, Vol. 90, 237–263. DOI: 10.1007/S11069-017-3043-8.

Aditian, A, Kubota, T. and Shinohara, Y., 2018, Comparison of GIS-based Landslide Susceptibility Models Using Frequency Ratio, Logistic Regression, and Artificial Neural Network in a Tertiary Region of Ambon, Indonesia. *Geomorphology*, Vol. 318, 101–111. DOI: 10.1016/J.GEOMORPH.2018.06.006.

Ali, S, A,, Parvin, F., Vojteková, J., Costache, R., Linh, N. T. T., Pham, Q. B., Vojtek, M., Gigović, L., Ahmad, A and Ghorbani, M. A., 2021, GIS-based Landslide Susceptibility Modeling: A Comparison Between Fuzzy Multi-Criteria and Machine Learning Algorithms. *Geoscience Frontiers*, Vol. 12, 857–876. DOI: 10.1016/J.GSF.2020.09.004.

Arabameri, A., Pradhan, B., Rezaei, K. and Lee, C. W. W., 2019, Assessment of Landslide Susceptibility Using Statistical- and Artificial Intelligence-Based FR-RF Integrated Model and Multiresolution DEMs. *Remote Sensing*, Vol. 11, DOI: 10.3390/rs11090999.

Arabameri, A., Saha, S., Roy, J., Chen, W., Blaschke, T. and Bui, D. T., 2020, Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in the Gallicash River Watershed, Iran. *Remote Sensing*, Vol. 12. DOI: 10.3390/rs12030475.

Arca, D., Keskin, Citiroglu, H. and Tasoglu, I. K., 2019, A Comparison of GIS-based Landslide Susceptibility Assessment of the Satuk Village (Yenice, NW Turkey) by Frequency Ratio and Multi-Criteria Decision Methods. *Environmental Earth Sciences*, Vol. 78(3), 1–13. DOI: 10.1007/S12665-019-8094-6.

Baum, R. L. and Godt, J. W., 2010, Early Warning of Rainfall-Induced Shallow Landslides and Debris Flows in the USA. *Landslides*, Vol. 7, 259–272. DOI: 10.1007/s10346-009-0177-0.

Banshtu, R. S.; Prakash, C. Application of Remote Sensing and GIS Techniques in Landslide Hazard Zonation of Hilly Terrain. Landslide Sci. a Safer Geoenvironment Vol. 2 Methods Landslide Stud., 2014, 313–317. https://doi.org/10.1007/978-3-319-05050-8_49.

Bhatti, S. S. and Tripathi, N. K., 2014, Built-up Area Extraction Using Landsat 8 OLI Imagery. *GIScience & Remote Sensing*, Vol. 51, 445–467. DOI: 10.1080/15481603.2014.939539.

Breiman, L., 2001, Random Forests. *Machine Learning*, Vol. 45, 5–32. DOI: 10.1023/A:10-10933404324.

Brodley, C. E and Friedl, M. A., 1997, Decision Tree Classification of Land Cover from Remotely Sensed Data. *Remote Sensing of Environment*, Vol. 61, 399–409. DOI: 10.1016/S0034-4257(97)00049-7.

Budimir, M. E. A., Atkinson, P. M. and Lewis, H. G., 2015, A Systematic Review of Landslide Probability Mapping Using Logistic Regression. *Landslides*, Vo1. 12(3), 419–436. DOI: 10.1007/S10346-014-0550-5.

Bui, D. T., Shahabi, H., Shirzadi, A., Chapi, K., Alizadeh, M., Chen, W., Mohammadi, A., Ahmad, B. B., Panahi, M., Hong, H. and Tian, Y., 2018, Landslide Detection and Susceptibility Mapping by AIRSAR Data Using Support Vector Machine and Index of Entropy Models in Cameron Highlands, Malaysia. *Remote Sensing*, Vol. 10, DOI: 10.3390/RS10101527.

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V., 2013, Landslide Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues. *Natural Hazards and Earth System Sciences*, Vo. 13, 2815–2831. DOI: 10.5194/NHESS-13-2815-2013.

Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W. and Guo, Z., 2020, Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models. *Remote Sensing*, Vol. 12. DOI: 10.3390/rs12030502.

Chen, W., Fan, L., Li, C. and Pham, B. T., 2019, Spatial Prediction of Landslides Using Hybrid Integration of Artificial Intelligence Algorithms with Frequency Ratio and Index of Entropy in Nanzheng County, China. *Applied Sciences*, Vol. 10, DOI: 10.3390/APP10010029.

Chen, W. and Li, Y., 2020, GIS-based Evaluation of Landslide Susceptibility Using Hybrid Computational Intelligence Models. *Catena*, Vol. 195, DOI: 10.1016/j.catena.2020.104777.

Chen, W., Shahabi, H., Zhang, S., Khosravi, K., Shirzadi, A., Chapi, K., Pham, B. T., Zhang, T., Zhang, L., Chai, H., Ma, J., Chen, Y., Wang, X., Li, R. and Ahmad, B. B., 2018a, Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Applied Sciences*, Vol. 8, DOI: 10.3390/APP8122540.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., Duan, Z., Li, S., Zhu, A. X., 2018b, GIS-based Landslide Susceptibility Evaluation Using a Novel Hybrid Integration Approach of Bivariate Statistical Based 2 Random Forest Method. *Catena*, Vol. 164(7), DOI:10.1016/j.catena.2018.01.012.

Chen, W., Zhang, S., Li, R. and Shahabi, H., 2018c, Performance Evaluation of the GIS-based Data Mining Techniques of Best-First Decision Tree, Random Forest, and Naïve Bayes Tree for Landslide Susceptibility Modeling. *Science of the Total Environment*, Vol 644, 1006–1018. DOI: 10.1016/J.SCITOTENV.2018.06.389.

Chen, X. and Chen, W., 2021, GIS-based landslide Susceptibility Assessment Using Optimized Hybrid Machine Learning Methods. *Catena*, Vol. 196, DOI: 10.1016/j.catena.2020.104833.

Cigdem, O. and Demirel, H., 2018, Performance Analysis of Different Classification Algorithms Using Different Feature Selection Methods on Parkinson's Disease Detection. *Journal of Neuroscience Methods*, Vol. 309, 81–90. DOI: 10.1016/j.jneumeth.2018.08.017.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E. and Webb, J., 2008, A Place-Based Model for Understanding Community Resilience to Natural Disasters. *Global Environmental Change*, Vol. 18, 598–606. DOI: 10.1016/j.gloenvcha.2008.07.013.

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T. and Nishino, K., 2007, GIS-based Weights-of-Evidence Modelling of Rainfall-Induced Landslides in Small Catchments for Landslide Susceptibility Mapping. *Environmental Geology*, Vol., 54(2), 311–324. DOI: 10.1007/S00254-007-0818-3.

Dehnavi, A., Aghdam, I. N., Pradhan, B., Morshed Varzandeh, M. H., 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*, Vol. 135, 122–148. DOI: 10.1016/J.CATENA.2015.07.020.

Deijns, A. A. J., Bevington, A. R., van Zadelhoff, F., de Jong, S. M., Geertsema, M. and McDougall, S., 2020, Semi-automated Detection of Landslide Timing Using Harmonic Modelling of Satellite Imagery, Buckinghorse River, Canada. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 84, DOI: 10.1016/j.jag.2019.101943.

Demir, G., Aytekin, M. and Akgun, A., 2014, Landslide Susceptibility Mapping by Frequency Ratio and Logistic Regression Methods: An Example from Niksar–Resadiye (Tokat, Turkey). *Arabian Journal of Geosciences*, Vol. 8(3), 1801–1812. DOI: 10.1007/S12517-014-1332-Z.

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R. and Althuwaynee, O. F., 2013, Landslide Susceptibility Mapping Using Certainty Factor, Index of Entropy and Logistic Regression Models in GIS and their Comparison at Mugling-Narayanghat Road Section in Nepal Himalaya. *Natural Hazards*, Vol. 65, 135–165. DOI: 10.1007/s11069-012-0347-6.

Dou, J., Bui, D. T., Yunus, A. P., Jia, K., Song, X., Revhaug, I., Xia, H. and Zhu, Z., 2015, Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PLoS ONE*, Vol. 10, DOI: 10.1371/journal.pone.0133262.

Dou, J., Yunus, A. P., Bui, D. T., Sahana, M., Chen, C. W., Zhu, Z., Wang, W. and Pham, B. T., 2019, Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. *Remote Sensing*, Vol. 11, DOI: 10.3390/RS11060638.

Fang, Z., Wang, Y., Peng, L. and Hong, H., 2020, Integration of Convolutional Neural Network and conventional Machine Learning Classifiers for Landslide Susceptibility Mapping. *Computers and Geosciences*, Vol. 139, DOI: 10.1016/j.cageo.2020.104470.

Fell, R., 1993, Landslide Risk Assessment and Acceptable Risk. *Canadian Geotechnical Journal*, Vol. 31,(2), https://doi.org/10.1139/t94-031.

Frangov, G., Petkova, V., Stoyanov, V., Kadiyski, M., Kostov, V. and Papaliangas, T., 2017, Landslide Risk Assessment and Mitigation along a Road in Sw Bulgaria. *Fresenius Environmental Bulletin*, Vol. 26, 244–253.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P., 2008, Comparing Landslide Inventory Maps. *Geomorphology*, Vol. 94, 268–289. DOI: 10.1016/J.GEOMORP-H.2006.09.023.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D. and Aryal, J., 2019, Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing*, Vol. 11, DOI: 10.3390/RS-11020196.

Ghosh, T., Bhowmik, S., Jaiswal, P., Ghosh, S. and Kumar, D., 2020, Generating Substantially Complete Landslide Inventory using Multiple Data Sources: A Case Study in Northwest Himalayas, India. *Journal of the Geological Society of India*, Vol. 95(1), 45–58. DOI: 10.1007/S12594-020-1385-4.

Glade, T., 2003, Landslide occurrence as a response to land use change: a review of evidence from New Zealand. *Catena*, Vol. 51(3-4), 297-314, DOI:10.1016/S0341-8162(02)00170-4.

Gorsevski, P. V., Brown, M. K., Panter, K., Onasch, C. M., Simic, A. and Snyder, J., 2016, Landslide Detection and Susceptibility Mapping Using LiDAR and An Artificial Neural Network Approach: A Case Study in the Cuyahoga Valley National Park, Ohio. *Landslides*, Vol. 13, 467–484. DOI: 10.1007/S10346-015-0587-0.

He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., Chai, H., Bian, H., Ma, J., Chen, Y., Wang, X., Chapi, K. and Ahmad, B. B., 2019, Landslide Spatial Modelling Using Novel Bivariate Statistical Based Naïve Bayes, RBF Classifier, and RBF Network Machine Learning Algorithms. *Science of The Total Environment*, Vol. 663, 1–15. DOI: 10.1016/J.SCITOTEN-V.2019.01.329.

Jebur, M. N., Pradhan, B. and Tehrany, M. S., 2014, Optimization of Landslide Conditioning Factors Using Very High-Resolution Airborne Laser Scanning (LiDAR) Data at Catchment Scale. *Remote Sensing of Environment*, Vol, 152, 150–165. DOI: 10.1016/J.RSE.2014.05.013.

Kavoura, K. and Sabatakakis, N., 2019, Investigating Landslide Susceptibility Procedures in Greece. *Landslides*, Vol. 17(1), 127–145. DOI: 10.1007/S10346-019-01271-Y.

Lee, S., Kim, J. C., Jung, H. S., Lee, M. J. and Lee, S., 2017, Spatial Prediction of Flood Susceptibility Using Random-Forest and Boosted-Tree Models in Seoul Metropolitan City, Korea. *http://www.tandfonline.com/a-ction/journalInformation?show=aimsScope&jou rnalCode=tgnh20#.VsXodSCLRhE* 8: 1185–1203. DOI: 10.1080/19475705.2017.1308971.

Lee, S., Lee, M. J., Jung, H. S. and Lee, S., 2019, Landslide Susceptibility Mapping Using Naïve Bayes and Bayesian Network Models in Umyeonsan, Korea. *Geocarto International*, Vol. 35(15), 1665-1679, https://doi.org/10.1-080/10106049.2019.1585482.

Lee, S. and Talib, J. A., 2005, Probabilistic Landslide Susceptibility and Factor Effect Analysis. *Environmental Geology*, Vol. 47(7), 982–990. DOI: 10.1007/S00254-005-1228-Z.

Lei, X., Chen, W. and Pham, B. T., 2020, Performance Evaluation of GIS-Based Artificial Intelligence Approaches for Landslide Susceptibility Modeling and Spatial Patterns Analysis. *ISPRS International Journal of Geo-Information*, Vol. 9, DOI: 10.3390/IJGI9070443.

Li, Y. and Chen, W., 2020, Landslide Susceptibility Evaluation Using Hybrid Integration of Evidential Belief Function and Machine Learning Techniques. *Water (Switzerland)*, Vol. 12. DOI: 10.3390/w12010113.

Luo, X., Lin, F., Chen, Y., Zhu, S., Xu, Z., Huo, Z., Yu, M. and Peng, J., 2019, Coupling Logistic Model Tree and random Subspace to Predict the Landslide Susceptibility Areas with Considering the Uncertainty of Environmental Features. *Scientific Reports* , Vol. 9(1), 1–13. DOI: 10.1038/s41598-019-51941-z.

Merghadi, A., Abderrahmane, B. and Bui, D. T., 2018, Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods. *ISPRS International Journal of Geo-Information 2018*, Vol. 7, DOI: 10.3390/IJGI7070268.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R. and Abderrahmane, B., 2020, Machine Learning Methods for Landslide Susceptibility Studies: A Comparative Overview of Algorithm Performance. *Earth-Science Reviews*, Vol. 207, DOI: 10.1016/j.earscirev.2020.103225.

Nachappa, T. G., Ghorbanzadeh, O., Gholamnia, K. and Blaschk, T., 2020. Multi-hazard Exposure Mapping Using Machine Learning for the State of Salzburg, Austria. *Remote Sensing*, Vol. 12, 1–24. DOI: 10.3390/RS12172757.

Nefeslioglu, H. A., Duman, T. Y. and Durmaz, S., 2008, Landslide Susceptibility Mapping for a Part of Tectonic Kelkit Valley (Eastern Black Sea region of Turkey). *Geomorphology*, Vol. 3(4), 401–418. DOI: 10.1016/J.GEOMORP-H.2006.10.036.

Ngadisih, Bhandary, N. P., Yatabe, R., Dahal, R. K., 2016, Logistic Regression and Artificial Neural Network Models for Mapping of Regional-Scale Landslide Susceptibility in Volcanic Mountains of West Java (Indonesia). *AIP Conference Proceedings*, Vol. 1730(1), DOI: 10.1063/1.4947407.

Nguyen, H. D., Pham, V. D., Nguyen, Q. H., Pham, V. M., Pham, M. H., Vu, V. M. and Bui. Q. T., 2020, An Optimal Search for Neural Network Parameters Using the Salp Swarm Optimization Algorithm: A Landslide Application. Remote Sensing Letters, Vol. 11(4), 353–362. DOI: 10.1080/2150704X.2020.1716409.

Pandey, V. K., Pourghasemi, H. R., Sharma, M. C., 2018, Landslide Susceptibility Mapping Using Maximum Entropy and Support Vector Machine Models along the Highway Corridor, Garhwal Himalaya. *Geocarto International*, Vol. 35, 168–187. DOI: 10.1080/10106049.2018.1510038.

Patel, R. C., Adlakha, V., Singh, P., Kumar, Y. and Lal, N., 2011, Geology, Structural and Exhumation History of the Higher Himalayan Crystallines in Kumaon Himalaya, India. *Journal of the Geological Society of India*, Vol. 77, 47–72. DOI: 10.1007/S12594-011-0008-5.

Pham, B. T., Jaafari, A., Prakash, I. and Bui, D. T., 2019a, A novel hybrid Intelligent Model of Support Vector Machines and the MultiBoost Ensemble for Landslide Susceptibility Modeling. *Bulletin of Engineering Geology and the Environment*, Vol. 78, 2865–2886. DOI: 10.1007/S10064-018-1281-Y/FIGURES/11.

Pham, B. T., Shirzadi, A., Shahabi, H., Omidvar, E., Singh, S. K., Sahana, M., Asl, D. T., Ahmad, B. B., Quoc, N. K. and Lee, S., 2019b, Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms. *Sustainability (Switzerland)*, Vo1. 1, 1–25. DOI: 10.3390/su11164386.

Pham, B. T., Bui, D. T., Prakash, I. and Dholakia, M. B., 2016, Rotation Forest Fuzzy Rule-Based Classifier Ensemble for Spatial Prediction of Landslides Using GIS. *Natural Hazards*, Vol. 83, 97–127. DOI: 10.1007/S11069-016-2304-2.

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T. and Cerda, A., 2020, Assessing and Mapping Multi-Hazard Risk Susceptibility Using a Machine Learning Technique. *Scientific Reports*, Vol. 10, 1–11. DOI: 10.1038/s41598-020-60191-3.

Pradhan, A. M. S. and Kim, Y. T., 2017, Spatial Data Analysis and Application of Evidential Belief Functions to Shallow Landslide Susceptibility Mapping at Mt. Umyeon, Seoul, Korea. *Bulletin of Engineering Geology and the Environment*, Vol. 76, 1263–1279. DOI: 10.1007/S10064-016-0919-X.

Pradhan, B., Abokharima, M. H., Jebur, M. N. and Tehrany, M. S., 2014, Land Subsidence Susceptibility Mapping at Kinta Valley (Malaysia) Using the Evidential Belief Function Model in GIS. *Natural Hazards*, Vol. 73, 1019–1042. DOI: 10.1007/S11069-014-1128-1.

Pradhan, S. P. and Siddique, T., 2020, Stability Assessment of Landslide-Prone Road Cut Rock Slopes in Himalayan terrain: A Finite Element Method Based Approach. *Journal of Rock Mechanics and Geotechnical Engineering*, Vol. 12, 59–73. DOI: 10.1016/J.JRMGE.2-018.12.018.

Pradhan, S. P., Vishal, V., Singh, T. N., 2018, Finite Element Modelling of Landslide Prone Slopes Around Rudraprayag and Agastyamuni in Uttarakhand Himalayan Terrain. *Natural Hazards*, Vol. 94, 181–200. DOI: 10.1007/s11069-018-3381-1.

Reichenbach, P., Busca, C., Mondini, A. C. and Rossi, M., 2014, The Influence of Land Use Change on Landslide Susceptibility Zonation: The Briga Catchment Test Site (Messina, Italy). *Environmental Management*, Vol. 54, DOI: 10.1007/S00267-014-0357-0.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., Guzzetti, F., 2018, A Review of Statistically-Based Landslide Susceptibility Models. *Earth-Science Reviews*, Vol. 180, 60–91. DOI: 10.1016/j.earscirev.2018.03.001.

Romer, C. and Ferentinou, M., 2016, Shallow landslide Susceptibility Assessment In A Semiarid Environment - A Quaternary Catchment of KwaZulu-Natal, South Africa. *Engineering Geology*, Vol. 201, 29–44. DOI: 10.1016/J.ENGGEO.2015.12.013.

Roy, J., Saha, S., Arabameri, A., Blaschke, T. and Bui, D. T., 2019, A Novel Ensemble Approach for Landslide Susceptibility Mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. *Remote Sensing*, Vol. 11, DOI: 10.3390/RS11232866.

Saha, A. and Saha, S., 2020, Comparing the Efficiency of Weight of Evidence, Support Vector Machine and their Ensemble Approaches in Landslide Susceptibility Modelling: A Study on Kurseong Region of Darjeeling Himalaya, India. *Remote Sensing Applications: Society and Environment*, Vol. 19, DOI: 10.1016/J.RSASE.-2020.100323.

Saha, S., Arabameri, A., Saha, A., Blaschke, T., Ngo, P. T. T., Nhu, V. H. and Band, S. S., 2021, Prediction of Landslide Susceptibility in Rudraprayag, India Using Novel Ensemble of Conditional Probability and Boosted Regression Tree-Based on Cross-Validation Method. *Science of the Total Environment*, Vol. 764, 142928. DOI: 10.1016/j.scitotenv.2020.142928.

Saha, S., Paul, G. C., Pradhan, B., Maulud, K. N. A. and Alamri, A. M., 2020, Integrating Multilayer Perceptron Neural Nets with Hybrid Ensemble Classifiers for Deforestation Probability Assessment in Eastern India. *Geomatics, Natural Hazards and Risk*, Vol. 12(1), 29-62, https://doi.org/10.1080/19475705.2020.1860139.

Sahin, E. K., 2020, Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. *SN Applied Sciences*, Vol. 2, 1–17. DOI: 10.1007/S42452-020-3060-1/TABLES/4.

Sahin, E. K., Colkesen, I., Acmali, S. S., Akgun, A., Aydinoglu, A. C., 2020, Developing Comprehensive Geocomputation Tools for Landslide Susceptibility Mapping: LSM Tool Pack. *Computers and Geosciences*, Vol. 144, DOI: 10.1016/j.cageo.2020.104592.

Santos, V., Datia N., Pato, M. P M., Santos, V., Datia, N., Pato, M. P. M., 2014, Ensemble Feature Ranking Applied to Medical Data. *Procedia Technology* Complete, Vol. 17, 223–230. DOI: 10.1016/J.PROTCY.2014.10.232.

Shahabi, H., Khezri, S., Ahmad, B. B. andHashim, M., 2014, Landslide Susceptibility Mapping at Central Zab Basin, Iran: A Comparison Between Analytical Hierarchy Process, Frequency Ratio And Logistic Regression models. *CATENA* 115: 55–70. DOI: 10.1016/J.CATENA.2013.11.014

Sharma A, Prakash C, Manivasagam VS. 2021a. Entropy-Based Hybrid Integration of Random Forest and Support Vector Machine for Landslide Susceptibility Analysis. *Geomatics*, Vol. 1, 399–416. DOI: 10.3390/GEOMATIC-S1040023.

Sharma, A., Prakash, C. and Nautiyal, A., 2021b, Recent Landslide News of Himachal Pradesh, India. *Landslides*, 1–4. DOI: 10.1007/S10346-021-01802-6.

Shim, J. H. and Kim, C. Il., 2015, Measuring Resilience To Natural Hazards: Towards Sustainable Hazard Mitigation. *Sustainability (Switzerland)*, Vol. 7, 14153–14185. DOI: 10.3390/su71014153.

Thi Ngo, P. T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A. and Lee, S., 2021, Evaluation of Deep Learning Algorithms for National Scale Landslide Susceptibility Mapping of Iran. *Geoscience Frontiers*, Vol. 12, 505–519. DOI: 10.1016/j.gsf.2020.06.013.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I. and Dick, O. B., 2012, Spatial Prediction of Landslide Hazards in Hoa Binh Province (Vietnam): A Comparative Assessment of the Efficacy of Evidential Belief Functions and Fuzzy Logic Models. *Catena*, Vol. 96, 28–40. DOI: 10.1016/J.CATENA.2012.04.001.

Tripathi, N. K., Gokhale, K. V. G K. and Siddiqui, M. U., 2000, Directional Morphological Image Transforms for Lineament Extraction from Remotely Sensed Images. *International Journal of Remote Sensing*, Vol. 21, 3281–3292. DOI: 10.1080/014311600750019895.

Tsangaratos, P. and Ilia, I., 2016, Comparison of a Logistic Regression and Naïve Bayes Classifier in Landslide Susceptibility Assessments: The Influence of Models Complexity and Training Dataset Size. *Catena*, Vol. 145, 164–179. DOI: 10.1016/J.CATENA.2016.06.004.

Varnes, D. J., 1984, *Landslide Hazard Zonation A Review of Principles and Practice, Natural Hazards. UNESCO, Paris.* References - Scientific Research Publishing.

Versain, L. D., 2019, Bi-Variate Statistical Approach in Landslide Hazard Zonation: Central Himalayas of Himachal Pradesh, India. *International Journal of Applied Engineering Research*, Vol. 14: 415–428.

Wang, G., Lei, X., Chen, W., Shahabi, H., Shirzadi, A., 2020, Hybrid Computational Intelligence Methods for Landslide Susceptibility Mapping. *Symmetry*, Vol. 12, DOI: 10.3390/SYM12030325.

Wang, L. J., Guo, M., Sawada, K., Lin, J., Zhang, J., Wang, L. J., Guo, M., Sawada, K., Lin, J., Zhang, J., 2016, A Comparative Study of Landslide Susceptibility Maps Using Logistic Regression, Frequency Ratio, Decision Tree, Weights of Evidence and Artificial Neural Network. *Gesc. J.*, Vol. 20, 117–136. DOI: 10.1007/S12303-015-0026-1.

Wang, Q., Li, W., Chen, W., Bai, H., 2015, GIS-based assessment .f Landslide Susceptibility Using Certainty Factor and Index of Entropy Models for the Qianyang County of Baoji City, China. *Journal of Earth System Science*, Vol. 124, 1399–1415. DOI: 10.1007/s12040-015-0624-3.

Yi, Y., Zhang, Z., Zhang, W., Xu, Q., Deng, C. and Li, Q., 2019, GIS-based Earthquake-Triggered-Landslide Susceptibility Mapping with an Integrated Weighted Index Model in Jiuzhaigou Region of Sichuan Province, China. *Natural Hazards and Earth System Sciences*, Vol., 19, 1973–1988. DOI: 10.5194/nhess-19-1973-2019.

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*, Vol. 12: 639–655. DOI: 10.1016/j.gsf.2020-.05.010.

Yusof, N. M., Pradhan, B., Shafri, H. Z. M., Jebur, M. N. and Yusoff, Z., 2015, Spatial landslide hazard Assessment along the Jelapang Corridor of the North-South Expressway in Malaysia Using High Resolution Airborne LiDAR Data. *Arabian Journal of Geosciences*, Vol. 8, 9789–9800, DOI: 10.1007/s12517-015-1937-x.

Zare, M., Jouri, M. H., Salarian, T., Askarizadeh, D., 2014, Comparing of Bivariate Statistic, AHP and Combination Methods to Predict the Landslide Hazard in Northern Aspect of Alborz Mt. (Iran). *International Journal of Agriculture and Crop Sciences*, Vol., 7(9), 543–554.