# Chapter 14
# Predicting Landslide Susceptibility of a Mountainous Region Using a Hybrid Machine Learning-Based Model

**Amol Sharma** (iD) **and Chander Prakash** (iD)

**Abstract**  The Himalayan region of northern India has witnessed extensive devastation over the years due to numerous landslide incidents. Ever-increasing infrastructure development activities have continuously tempered the fragile Himalayan ecosystem without manifesting genuine concern for the environment, leading to increased landslide occurrences in the region. Landslide susceptibility analysis is carried out to determine the probability of landslide occurrence based on the regional geological and environmental settings. The landslide susceptibility maps are generated by integrating the historical landslide information with landslide causative factors using modeling algorithms. In this study, a hybrid integration of frequency ratio (FR) model and support vector machine (SVM) model is carried out to access the susceptibility of Mandi district of Himachal Pradesh. A set of 1723 historical landslides were mapped from various sources and split into 70% training and 30% validation datasets. A set of 11 landslide influencing factors were identified, and a spatial relationship was generated using individual FR and SVM models and their hybrid integrated model. The susceptibility maps thus generated were evaluated for performance based on receiver-operating characteristic (ROC) curves. The results concurred that the FR-SVM model had better accuracy in comparison to individual FR and SVM model having validation accuracies of 84.7%, 77.9%, and 81.2 %, respectively. Therefore, the FR-SVM model is considered suitable for analysis and is recommended in similar geophysical environment.

**Keywords**  Support vector machine · ROC · Frequency ratio · Digital elevation model

A. Sharma (✉) · C. Prakash (✉)
Civil Engineering Department, National Institute of Technology,
Hamirpur, Himachal Pradesh, India
e-mail: amol@nith.ac.in; chandermanali@nith.ac.in

## 14.1 Introduction

Landslides are an omnipresent global hazard incurring losses around the geophysical environment. The devastating impact of landslides is not limited to the socioeconomic losses but extends up to swerve environmental implications. The global disaster reports confirm that landslides are responsible for 4.9% of the total natural disasters [1]. The spatiotemporal occurrence of landslides is modulated by some triggering processes. These include earthquakes, extreme rainfall, snow/glacier melting, land-use/land cover (LULC) changes, various anthropogenic activities which cause vibrations, overburden on soil material, uprooting of supports laterally and change in moisture content of soil or rock structures, etc. [2, 3]. The global economic expansion and unplanned haphazard development activities, in the mountainous areas, have exacerbated the socioeconomic impacts of landslides in recent times [4, 5]. Hence, for effective landslide risk management, the areas susceptible to landslides should be identified accurately so that adequate response and emergency measures can be administered.

A detailed investigation of the various landslide influencing factors resulting in slope failure can provide relevant information regarding landslide occurrence. Landslide susceptibility mapping (LSM) is considered fundamental in a competent approach toward landslide hazard assessment, management, and mitigation [6, 7]. The susceptibility mapping is a complex process of establishing interdependence of historical landslide events and topographical, geological, and hydrological variables, which are expected to influence landslide occurrences on a regional scale [8]. The various stages of landslide susceptibility analysis include landslide inventory generation and identification of landslide causative factors (LCFs), spatially correlating these variables using a modeling framework and model validation [9, 10]. The predictive potential of models is correlated with the quality and accuracy of landslide inventory and the optimal selection of LCF [4, 11]. For reliable and accurate landslide susceptibility mapping of an area, accurate mapping of landslides and selection of causative factors should be conclusive and logical [12, 13]. Optimization of landslide causative factors is crucial for an effective prediction model [14, 15]. Researchers have used factor analysis, multicollinearity analysis, linear correlation, certainty factor approach, and multifactor set techniques for the optimal selection of LCF [10, 16]. 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 (LCFs) to achieve high predictive potential from the applied models within a reasonable time.

The LSM modeling process has evolved over the years from qualitative models such as weight of evidence (WOE) [17, 18], analytical hierarchy process (AHP) [19, 20], etc. toward data-driven models such as the evidential belief functions (EBF) [21, 22], frequency ratio (FR) [23, 24], certainty factor (CF) [25, 26], etc. These

techniques have performed satisfactorily for landslide prediction but sometimes lack functional correlations between LCFs [4]. Another drawback of bivariate models is that hypothesis must be accepted before modeling [27].

The recent advancements in machine learning (ML) algorithms and their integration with Python or R programming have proved to be an articulate tool for mapping and analyzing natural hazards [28]. Traditional ML algorithms include logistic regression (LR) [4, 29], artificial neural networks (ANN) [30, 31], decision tree (DT) [17, 32], support vector machine (SVM) [28, 33], fuzzy logic (FL) [34, 35], Naïve Bayes (NB) algorithm [36, 37], kernel logistic regression (KLR) [38, 39], random forest (RF) [21, 29], etc. The ML algorithms can rearrange their internal structure according to landslide data type and have the potential to analyze and update the factor contribution automatically and continuously [37, 40]. However, the results of ML techniques are prone to errors and sometimes lack ease of interpretation concerning the individual contribution of subclasses of LCF. In recent times, various ensemble or hybrid learning techniques are being used, combining multiple modeling approaches to improve the overall predictive potential of the model [41]. Some commonly used ensembling techniques in landslide susceptibility analysis include bagging, stacking, and boosting approaches. A hybrid bagging-based kernel logistic regression (BKLR) approach was adopted by [42] using two kernel functions. Hence, the objective of the current study is to integrate the frequency ratio (FR) statistical model with radial kernel-based support vector machine (SVM) learning models to analyze and predict the potential landslide-prone areas. The accuracy of prediction and validation for FR, SVM, and hybrid FR-SVM models is analyzed using ROC curves.

## 14.2   Study Area

The study area constitutes the Mandi district of Himachal Pradesh lying between 31°13′50″–32°04′30″ N and 76°37′20″–77°23′15″ E. The district has a total area of 3951 km² with a population of 901,344. A major part of the district falls in the Lesser Himalayan region comprising of steep and rugged mountain ranges and fluvial valleys. The regions' altitude varies between 500 m and 3400 m from low-lying valleys to higher-elevation mountain ranges. With a forest cover of 45%, scrub, sal, and bamboo forests are found at lower elevations, while the alpine forests are characteristics of higher elevations. The Siwaliks and Lesser Himalayan soils are mainly found in the district, which are generally high in organic matter and characterized by rugged topography. The district has two distinct and well-defined hydrogeological units, that is, the porous formations constituted by unconsolidated sediments and the fissured formations. The study area is drained by Beas and Sutlej rivers. The density of roads for the region is 155 km per 100 km², which is higher than the state's average density.

A landslide inventory is a prerequisite for analyzing the spatial distribution of landslides, which is necessary to identify potential landslide zones in the study area [43]. The primary landslide inventory was generated from visual interpretation through high-resolution Google Earth images and analyzing the terrain characteristics derived from Advanced Land Observing Satellite (ALOS) Polarimetric Phased Array L-band Synthetic Aperture Radar (PALSAR) Digital Elevation Model (DEM). The published landslide inventories from Himachal Pradesh State Disaster Management Authority (HPSDMA), Geological Survey of India (GSI), and NASA along with numerous newspaper articles and previous landslide studies of the study area [44–46] were the auxiliary data sources. The spatial information of the landslides was extracted, and an inventory was compiled by incorporating geomorphologic, LULC, landslide magnitude (length and area) and other characteristics. A total of 1723 landslides with an average area of 1425 $m^2$, the area of landsides varies from a minimum of 2.5 $m^2$ to maximum $2.9 \times 10^5$ $m^2$, were mapped. The landslide inventory was further split individually for both districts into 70% training and 30% validation datasets, as suggested by [47–49] in ArcGIS (Fig. 14.1).
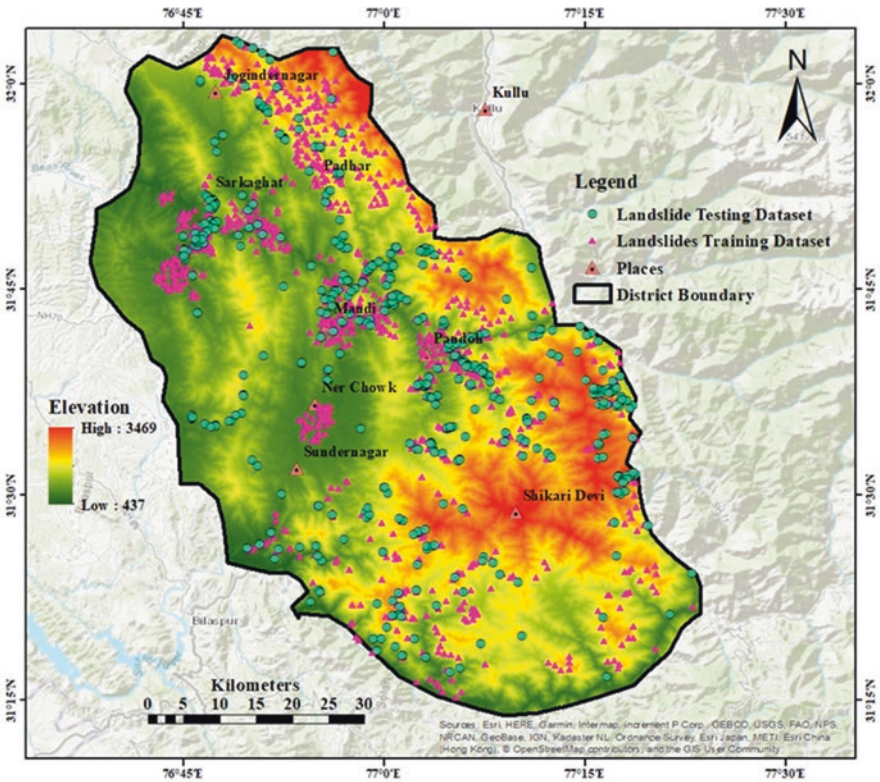


**Fig. 14.1** Study area and landslide inventory with training and testing datasets

## 14.3  Materials and Methods

### 14.3.1  Landslide Causative Factors (LCFs)

The predictive capabilities of the applied algorithms depend on the quality of processing data derived from data sources such as DEM and satellite images. This study incorporates 11 independent factors influencing landslide occurrence. ALOS-PALSAR digital elevation model (DEM) of 12.5 m resolution was used to derive elevation, slope gradient, slope aspect, curvature, topographical wetness index (TWI), and drainage density whereas the Landsat-8 (OLI) imagery was used to derive the normalized difference vegetation index (NDVI) and lineament density maps. The rest of the thematic layers, such as distance from road, geology, and soil maps, were procured and mapped using data from various government repositories [50, 51]. The summary of data products and derived information is given in Table 14.1. The slope gradient measures the steepness of the hilly slopes and was subdivided into five categories using natural break classification. The plan curvature is defined as the angle of contours generated by their intersection with the horizontal surface. The plan curvature represents the direction of maximum slope and helps in identifying the morphology of topography of the area and differentiating between valleys and ridges [52, 53]. Although the relationship between landslides and aspect is still under investigation, the aspect has a direct relationship with discontinuities, vegetation covers, and soil moisture, which affects the landslide occurrences [36, 54]. The elevation of a region signifies the altitude from mean sea level and is widely used in susceptibility analysis. The northern region of the study area has higher

**Table 14.1**  Data purpose and source along with their scale/resolution

| Data | Purpose | Source | Scale/resolution |
|---|---|---|---|
| State Authority, Himachal Pradesh | Boundary of the study area | https://hpmandi.nic.in/map-ofdistrict/ | 1:50000 |
| H.P. Disaster Revenue Reports (2015–2019), GSI, NASA, etc. | Landslide mapping | https://hpsdma.nic.in/ | 1:50000 |
| Digital elevation model (ALOS PALSAR) | Topographical and hydrological characteristics | https://search.asf.alaska.edu | 12.5 m |
| High-resolution Landsat-8 imagery | Lineament density and NDVI | http://earthexplored.usgs.gov | 30 m visual |
|  |  |  | 15 m panchromatic |
| Geological Survey of India (GSI) | Geology | https://bhukosh.gsi.gov.in/ | 1:50000 |
| MORTH, NHAI, and GSI | Road network | https://morth.nic.in/ | 1:50000 |
| ICAR-NBSS and LUP | Soil characteristics | https://www.nbsslup.in/ | 1:50000 |

elevation varying from 3000 to 6000 m. The drainage density represents stream length per unit area in a drainage basin. It directly influences the erodibility of slopes that are dissected by channels and influences the surface runoff [55]. The topographical wetness index (TWI) is the measure of accumulation of water in areas having variable elevations. Higher TWI values signify greater tendency of slopes toward erosion [5, 53]. The TWI map of the study area was prepared by mathematical augmentation of drainage parameters, using equation TWI = [ln $(F_S)$/Tan $(\alpha)$], where $F_S$ represents the accumulation of flow and $\alpha$ represents the gradient of slopes. The lineament density represents the topographical surface and the underlying faults and fractures in the structures. The lineaments of the study area were derived from Landsat-8 imagery and line density tools in GIS [56]. The NDVI is a dimensionless entity and gives information about the vegetation cover in an area. The NDVI map of the district was generated from Landsat-8 images by using image analysis in ERDAS IMAGINE software using NDVI = $(\omega_{NIR} - \omega_{RED})/(\omega_{NIR} + \omega_{RED})$, where $\omega_{NIR}$ represents near-infrared channel and $\omega_{RED}$ represents red channel of the electromagnetic spectrum.

The geology of an area represents the topographical aspects of the underlying surface of an area along with its mineral and rock types [53]. The Jutogh formation of Mandi district comprises of slates, schists, and quartzite with hematite. Mandi-Darla volcanic is also known to occur, which represents lava flows in the past. The Shah formation is characterized by salt grit, dolomite, limestone, quartzite, and red shales [57, 58]. The soil properties influence the landslide occurrence in the region [53]. The study area's soil map procured from the National Bureau of Soil Survey and Land Use Planning (ICAR-NBSS and LUP), digitized in GIS environment. Five categories of soils were identified based on their geomorphological and erosion characteristics. The lesser Himalayan soils of side and reposed slopes are predominant in both the districts having coarse loamy and skeletal loamy soils and facilitating moderate to swerve erosion [59]. Additionally, Siwalik soils of fluvial valleys are widespread in Mandi district having sandy to loamy structure facilitating moderate erosion.

The unplanned road construction activities often involve disruption of the natural bed slopes which results in higher probability of slope failures [48, 52]. The details of road network distribution of national highways, state highways, and major district roads in the study area were procured from MORTH and GSI database, digitized in GIS software. The Euclidian distance from roads was calculated using distance from line operation in GIS and was classified into five categories from 0 to 500 m at an interval of 100 m each to analyze the influence of road construction on landslide occurrence. The methodology adopted for the study area is described through a flowchart as shown in Fig. 14.2, and the maps of thematic layers are shown in Fig. 14.3.
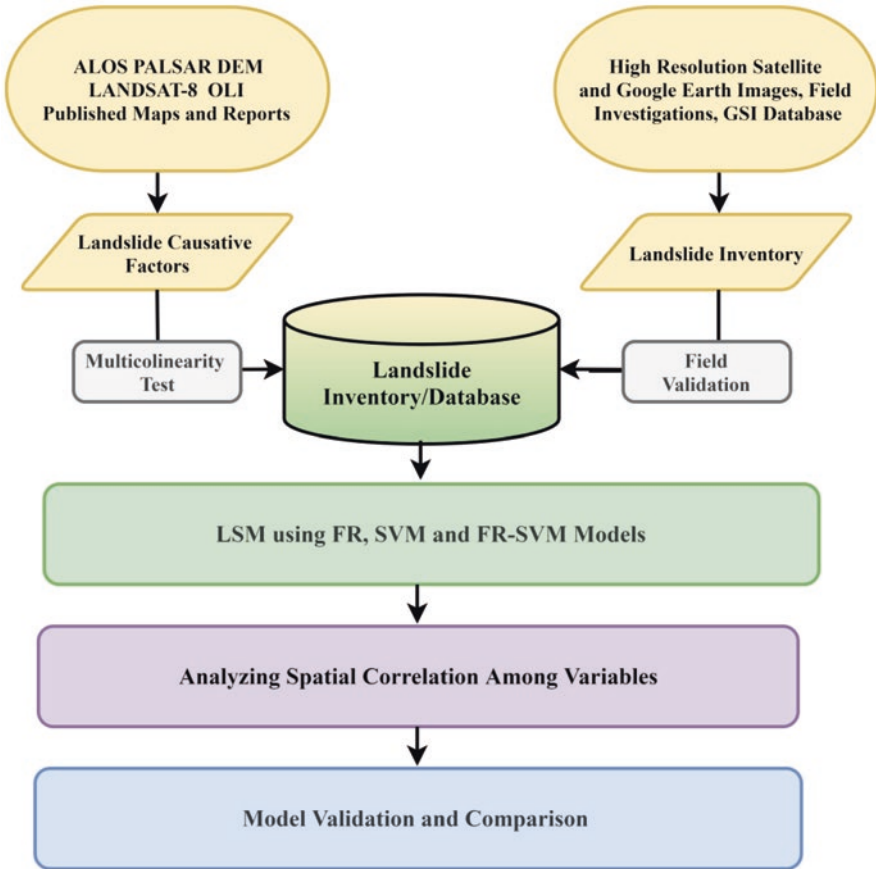
**Fig. 14.2** The methodology adopted for the susceptibility analysis

## 14.3.2   Bivariate Frequency Ratio (FR) Model

The FR represents the ratio between the pixel data with and without landslides and pixels of input raster data layers of causative factors. The FR values are computed for each class of causative factors using Eq. (14.1). The correlation is high if the FR value is greater than 1, while less than 1 value of FR represents lower correlation of landslides with causative factors [60]. A study to access landslide susceptibility using the frequency ratio method was carried out by [61], which incorporated nine predictor variables. The FR model is considered a good modeling algorithm due to its easier applicability and production of better results than similar models [62]:

$$FR(i) = \left(N_{LP} / N_{TP} / \left(\sum N_{LP} / \sum N_{TP}\right)\right)$$
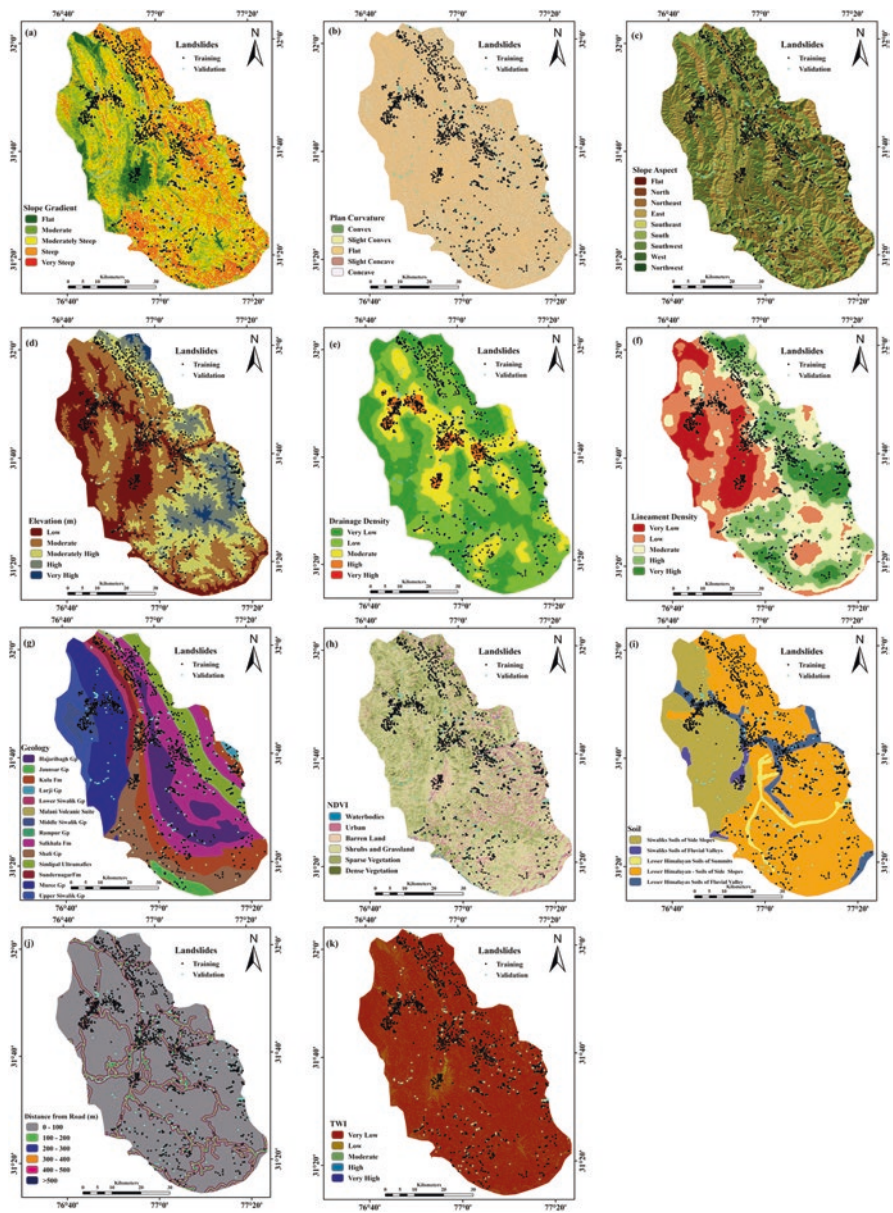
(14.1)

**Fig. 14.3** 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

where:

$N_{LP}$, landslide pixels in each class of landslide factors
$N_{TP}$, total number of pixels in each class
$\sum N_{LP}$, total landslide pixels in the study area
$\sum N_{TP}$, total pixels in the whole study area

The FR values calculated for all landslide factors are combined to produce the final LSM map using Eq. (14.2):

$$LSM_{FR} = FR_1 + FR_2 + FR_3 + \ldots + FR_n$$

(14.2)

### 14.3.3   Support Vector Machine (SVM) Model

The SVM algorithm is a set of supervised machine learning algorithms used to analyze nonlinear data for regression as well as classification [63]. SVM is a nonparametric approach which uses classification hyperplane along with a set of data points that are closer to the hyperplane called support vectors to maximize classification margin [33]. The SVM is considered a robust algorithm and is widely used in landslide susceptibility analysis.

The kernel tricks convert nonlinear datasets and project them into higher-dimensional dataset by applying Lagrangian multipliers. These functions can be categorized as radial, linear, or polynomial and sigmoid [10, 64] and are mathematically expressed as Eq. (14.3):

$$k\left(x_i, y_j\right) = \left[f\left(x_i\right), f\left(y_j\right)\right]$$

(14.3)

where $x_i$ and $y_j$ are the dimensional inputs for kernel function $k$ in an n-dimensional environment. The optimum hyperplane is generated using the decision function given in Eq. (14.4):

$$y\left(x\right) = \left(\alpha . \rho\left(x\right)\right) + b,$$

(14.4)

where $\alpha$ represents the orientation vector of the hyperplane, $\rho(x)$ is the input sample $x$ which is to be converted, and $b$ represents the distance of hyperplane from the origin.

## 14.4  Result and Discussion

### 14.4.1  Test for Multicollinearity

A test to identify multicollinearity among independent variables was carried out to check for the presence of any correlation among landslide causative factors. The problem of multicollinearity can result in reduced accuracy of the model. The tolerance values less than 0.1 and variance inflation factor (VIF) values greater than 10 suggest higher correlation between independent variables, and such variables should be removed from dataset. The results for multicollinearity of 11 landslide factors indicate that the VIF and tolerance values are within the acceptable limits as shown in Table. 14.2.

### 14.4.2  LSM Using FR Model

All the classes of landslide causative factors were rasterized and reclassified in the GIS environment for analysis. The FR values were calculated for 11 landslide causative factors, and their spatial relation to landslide factors is analyzed. The resultant FR values of each subclass of the landslide factors were calculated in GIS as shown in Table 14.3. The analysis indicates that drainage density, geology, NDVI, distance from road, and TWI were the critical factors affecting the study area's landslide susceptibility. While analyzing the hydrological parameters, areas with moderately high, high, and very high drainage density and TWI values were prone to landslides. Also, it was found that areas near the vicinity of roads were found to be more landslide susceptible. In the current study, the distance from the road class of 0–100 m

**Table 14.2** Multicollinearity analysis

| Model | Coefficients of multicollinearity | |
|---|---|---|
| | Tolerance | VIF |
| Elevation | 0.698 | 2.451 |
| Slope aspect | 0.775 | 1.165 |
| Curvature | 0.514 | 8.793 |
| Slope gradient | 0.231 | 3.037 |
| Drainage density | 0.282 | 8.525 |
| Lineament density | 0.149 | 6.785 |
| Geology | 0.336 | 2.976 |
| NDVI | 0.131 | 7.631 |
| Soil | 0.472 | 5.270 |
| Roads | 0.287 | 7.638 |
| TWI | 0.312 | 4.686 |

**Table 14.3** Landslide occurrences and their spatial relation with landslide causative factors

| | Class pixels | Percent of pixels | Landslide pixels | Percent of pixels | Frequency ratio (FR) values |
|---|---|---|---|---|---|
| **Landslide causative factors** | | | | | |
| **Slope gradient** | | | | | |
| Flat | 435014 | 0.102 | 9 | 0.008 | 0.079 |
| Moderate | 948259 | 0.222 | 85 | 0.076 | 0.341 |
| Moderately steep | 1374272 | 0.322 | 304 | 0.271 | 0.842 |
| Steep | 1047813 | 0.245 | 490 | 0.437 | 1.780 |
| Very steep | 466470 | 0.109 | 234 | 0.209 | 1.910 |
| **Plan curvature** | | | | | |
| Convex | 94610 | 0.022 | 55 | 0.049 | 2.213 |
| Flat | 1953189 | 0.457 | 254 | 0.226 | 0.495 |
| Concave | 166377 | 0.039 | 73 | 0.065 | 1.671 |
| **Slope aspect** | | | | | |
| Flat | 33660 | 0.008 | 4 | 0.004 | 0.452 |
| North | 484657 | 0.113 | 126 | 0.112 | 0.990 |
| Northeast | 515422 | 0.121 | 115 | 0.102 | 0.849 |
| East | 497821 | 0.117 | 81 | 0.072 | 0.619 |
| Southeast | 503993 | 0.118 | 108 | 0.096 | 0.816 |
| South | 545067 | 0.128 | 175 | 0.156 | 1.222 |
| Southwest | 647098 | 0.151 | 238 | 0.212 | 1.400 |
| West | 546964 | 0.128 | 195 | 0.174 | 1.357 |
| Northwest | 497146 | 0.116 | 80 | 0.071 | 0.613 |
| **Elevation (m)** | | | | | |
| Low | 995824 | 0.233 | 212 | 0.189 | 0.811 |
| Moderate | 1624309 | 0.380 | 266 | 0.237 | 0.623 |
| Moderately high | 1028156 | 0.241 | 539 | 0.480 | 1.996 |
| High | 537465 | 0.126 | 101 | 0.090 | 0.715 |
| Very high | 86074 | 0.020 | 4 | 0.004 | 0.177 |
| **Drainage density** | | | | | |
| Very low | 1299831 | 0.305 | 150 | 0.134 | 0.439 |
| Low | 1908487 | 0.448 | 229 | 0.204 | 0.456 |
| Moderate | 877782 | 0.206 | 337 | 0.300 | 1.459 |
| High | 179820 | 0.042 | 393 | 0.350 | 8.307 |
| Very high | 5908 | 0.001 | 23 | 0.020 | 14.797 |
| **Lineament density** | | | | | |
| Very low | 585993 | 0.138 | 67 | 0.060 | 0.434 |
| Low | 1093925 | 0.257 | 113 | 0.101 | 0.392 |
| Moderate | 1109204 | 0.260 | 329 | 0.293 | 1.126 |
| High | 1085918 | 0.255 | 407 | 0.363 | 1.423 |
| Very high | 396788 | 0.093 | 206 | 0.184 | 1.971 |

**Table 14.3** (continued)

| | Class pixels | Percent of pixels | Landslide pixels | Percent of pixels | Frequency ratio (FR) values |
|---|---|---|---|---|---|
| **Geology** | | | | | |
| Larji group | 17112 | 0.004 | 6 | 0.005 | 1.335 |
| Shali group | 480871 | 0.113 | 99 | 0.088 | 0.784 |
| Jaunsar group | 90819 | 0.021 | 6 | 0.005 | 0.252 |
| Middle Siwalik group | 77936 | 0.018 | 37 | 0.033 | 1.808 |
| Salkhala group | 1020010 | 0.239 | 326 | 0.291 | 1.217 |
| Hazaribagh granite | 481719 | 0.113 | 77 | 0.069 | 0.609 |
| Dharmasala group | 761109 | 0.178 | 186 | 0.070 | 1.679 |
| Upper Siwalik group | 258408 | 0.060 | 4 | 0.004 | 0.059 |
| Lower Siwalik group | 61338 | 0.014 | 3 | 0.003 | 0.186 |
| Sundernagar formation | 100192 | 0.023 | 33 | 0.119 | 0.650 |
| Malani Volcanic Suite | 15813 | 0.004 | 1 | 0.007 | 0.112 |
| Simlipal ultramafics | 368975 | 0.086 | 144 | 0.128 | 1.486 |
| Kulu formation | 534747 | 0.125 | 200 | 0.178 | 1.424 |
| **NDVI** | | | | | |
| Waterbodies | 16242 | 0.004 | 33 | 0.029 | 7.736 |
| Urban | 492012 | 0.115 | 286 | 0.255 | 2.213 |
| Barren land | 470706 | 0.110 | 152 | 0.135 | 1.230 |
| Shrubs and grassland | 1933318 | 0.453 | 399 | 0.356 | 0.786 |
| Sparse vegetation | 1204917 | 0.282 | 219 | 0.195 | 0.692 |
| Dense vegetation | 154633 | 0.036 | 33 | 0.029 | 0.813 |
| **Soil** | | | | | |
| Lesser Himalayan soils of side/reposed slopes | 2736453 | 0.641 | 899 | 0.801 | 1.251 |
| Lesser Himalayan soils of fluvial valleys | 280750 | 0.066 | 124 | 0.111 | 1.682 |
| Siwaliks soils of side/reposed slopes | 1083902 | 0.254 | 79 | 0.070 | 0.278 |
| Siwaliks soils of fluvial valleys | 62713 | 0.015 | 16 | 0.014 | 0.971 |
| Lesser Himalayan soils of summits | 108010 | 0.025 | 4 | 0.004 | 0.141 |
| **TWI** | | | | | |
| Very low | 3192586 | 0.747 | 349 | 0.311 | 0.416 |
| Low | 1031330 | 0.241 | 436 | 0.389 | 1.610 |
| Moderate | 37036 | 0.009 | 208 | 0.185 | 21.383 |
| High | 9038 | 0.002 | 105 | 0.094 | 44.232 |
| Very high | 1838 | 0.000 | 24 | 0.021 | 49.715 |
| **Distance from road (m)** | | | | | |
| 0–100 | 240721 | 0.056 | 406 | 0.362 | 6.421 |
| 100–200 | 196740 | 0.046 | 297 | 0.265 | 5.747 |
| 200–300 | 172030 | 0.040 | 111 | 0.099 | 2.456 |
| 300–400 | 156805 | 0.037 | 80 | 0.071 | 1.942 |
| 400–500 | 145918 | 0.034 | 43 | 0.038 | 1.122 |
| >500 | 3359614 | 0.787 | 185 | 0.165 | 0.210 |
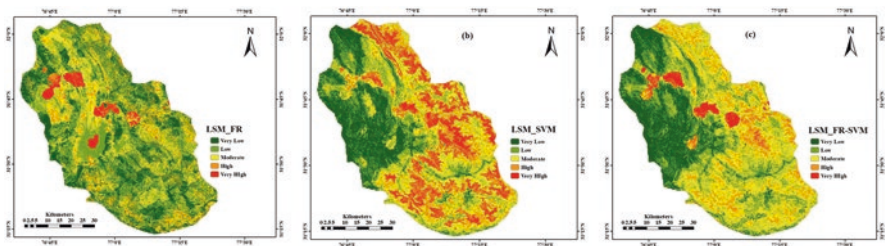
**Fig. 14.4** Landslide susceptibility maps: (**a**) FR model, (**b**) SVM model, and (**c**) FR-SVM model

had the highest FR values and hence requires special attention while planning construction activities. The LSM map generated was reclassified as areas with very low, low, moderately high, and very high susceptibility zones as shown in Fig. 14.4a.

### 14.4.3   LSM Using SVM Model

The SVM model was incorporated in GIS environment using R-integration and was utilized to calculate the spatial prediction of landslides in the study area. The LSM map generated was reclassified as areas with very low, low, moderately high, and very high susceptibility zones, as shown in Fig. 14.4b. It was observed that the slope gradient, drainage density, lineament density, soil, and distance from the road were the vital parameters that influence landslide occurrences. The slope gradient influences the landslide occurrence as steeper slope facilitates maximum slope failures. The results also show that the slope gradient classes, namely, steep (35°–45°) and very steep (>45°), are susceptible to landslides, whereas landslide occurrences were minimum for flatter slopes (<15°). Similarly, due to modest terrains at lower elevations (400 m–1000 m), these regions also witness less landslides. The highest probability of landslides was observed for moderate (1000 m–1500 m) to moderately high (1500 m–2000 m) elevations, but at high (2000 m–2500 m) to very high elevations (2500 m–3500 m), the probability of landslides again decreases.

### 14.4.4   LSM Using FR-SVM Model

The landslide causative factors were reclassified using FR model values, and the landside causative factors were reclassified according to these FR values. The radial-based SVM algorithm was applied to these factors to generate final LSM$_{FR-SVM}$ map. The LSM produced using the FR-SVM model was classified into five categories as shown in Fig. 14.4c. The LSM$_{FR-SVM}$ map analysis indicated that 15.28% of the total area falls into high landslide susceptibility zone, whereas 6.49% of the total area falls into very high landslide susceptibility zone. The analysis confirmed that TWI, drainage density, and NDVI were highly correlated to landside occurrence in the region.

## 14.5   Discussion

The disaster caused by landslides is not limited only to the socioeconomic loss but has a profound and everlasting impact on the overall demographic upliftment of the area. The effective management of landslide risk relies predominantly on the accuracy of the area's landslide susceptibility maps. This study aims to carry out the comparison of FR, SVM, and hybrid FR-SVM models for their accuracies in predicting landslides.

The analysis of the LSM's produced using FR, SVM, and hybrid FR-SVM models indicated that areas in the vicinity of roads (0–100 m) having high drainage density and having steeper slopes are more susceptible to landslides. In hilly regions such as Mandi district, the unplanned road construction activities continuously employ with the natural bed slope of the region. The areas with lower elevation have extensive developmental activities in comparison to higher regions of the study area. Since the regions constituting higher elevations are less approachable, therefore, less landslide incidences are reported.

The remaining landslide factors, that is, slope aspect, NDVI, geology, etc., had moderate to low influence on landslide occurrence. The results of this study are in accordance with similar landslide susceptibility studies of mountainous regions [30, 38, 51, 65].

The SVM is considered as a robust model and has been applied in wide varieties of landslide susceptibility analyses. However, it lacks in assigning effective relative weights to subclasses of causative factors. The SVM predictive power is maximized when the sample datasets are nonlinear and uses kernel functions for classification which sometimes leads to overfitting. The FR is a quantitative statistical model that can easily assign factor importance to subclasses of causative factors in minimal time.

While comparing the results obtained from the models, the FR-SVM model performed better than individual FR and SVM models. The FR-SVM model obtained a higher AUC value (84.7%) for model prediction as compared to FR model (77.9%) and SVM model (81.2%) as shown in Fig. 14.5. Hence, the use of hybrid model for predicting landslide susceptibility helps in eliminating the shortcoming of individual methods and improves the overall accuracy of the model.
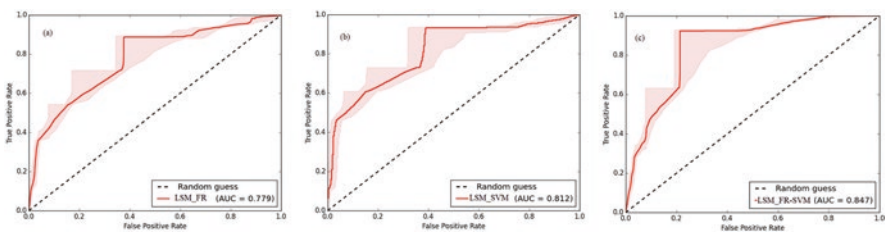


**Fig. 14.5**  ROC curves with AUC values: (**a**) FR model, (**b**) SVM model, and (**c**) FR-SVM model

## 14.6    Conclusion

Landslides constantly threaten the communities residing in landslide-prone mountainous regions. The planning of landslide management and mitigation activities in such regions requires adequate knowledge of the location and probable impact due to landslides. The Mandi district of Himachal Pradesh is highly prone to landslides and has witnessed landslide-induced destruction for many years. In the present study, 1723 landslides were documented from various sources and field observations, out of which 1206 (70%) landslides were included in the training dataset and 517 (30%) landslides were included in the testing dataset using random sampling. Also, it should be noted that all these landslide incidences were assumed to be triggered from rainfall only, and other triggering factors like earthquakes, volcanic eruptions and rapid snowmelt, etc. were neglected. In this study, although independent, the landslide causative parameters were chosen manually based on topographical and hydrological conditions of the study area. Eleven factors having high correlation to landslide occurrences were selected for identification of areas prone to landslides. The results of all models indicated that drainage density, slope gradient, and road distance are the three variables to be highly correlated to landslide occurrence. The results confirmed that the use of hybrid FR-SVM model produces robust model having better AUC values as compared to individual FR and SVM models. The process of integration of a bivariate and machine learning models provides highly accurate spatial correlation among variables with minimal overfitting of model. The outcomes of the present study should be considered during the planning of mitigation strategies in mountainous regions with similar topographic and hydrological conditions.

## References

1. M.J. Froude, D.N. Petley, Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. **18**(8), 2161–2181 (2018). https://doi.org/10.5194/nhess-18-2161-2018
2. S.L. Cutter, M. Gall, C.T. Emrich, Toward a comprehensive loss inventory of weather and climate hazards. Clim. Extrem. Soc. **9780521870**(October 2015), 279–295 (2008). https://doi.org/10.1017/CBO9780511535840.016
3. P. Reichenbach, C. Busca, A.C. Mondini, M. Rossi, The influence of land use change on landslide susceptibility zonation: The briga catchment test site (Messina, Italy). Environ. Manage. **54**(6), 1372 (2014, Nov). https://doi.org/10.1007/S00267-014-0357-0
4. W. Chen, L. Fan, C. Li, B.T. Pham, Spatial prediction of landslides using hybrid integration of artificial intelligence algorithms with frequency ratio and index of entropy in Nanzheng County, China. Appl. Sci. **10**(1), 29 (2019, Dec). https://doi.org/10.3390/APP10010029
5. Y. Li, W. Chen, Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. Water (Switzerland) **12**(1) (2020). https://doi.org/10.3390/w12010113
6. H. Wang, L. Zhang, K. Yin, H. Luo, J. Li, Landslide identification using machine learning. Geosci. Front. **12**(1), 351–364 (2021, Jan). https://doi.org/10.1016/J.GSF.2020.02.012
7. S. Saha et al., Prediction of landslide susceptibility in Rudraprayag, India using novel ensemble of conditional probability and boosted regression tree-based on cross-validation method. Sci. Total Environ. **764**, 142928 (2021). https://doi.org/10.1016/j.scitotenv.2020.142928

8. H.-D. Nguyen et al., An optimal search for neural network parameters using the Salp swarm optimization algorithm: A landslide application. Remote Sens. Lett. **11**(4), 353–362 (2020, Apr). https://doi.org/10.1080/2150704X.2020.1716409

9. R. Fell, Landslide risk assessment and acceptable risk, no. 4 (1993)

10. W. Chen, H.R. Pourghasemi, S.A. Naghibi, A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull. Eng. Geol. Environ. **77**(2), 647–664 (2018, May). https://doi.org/10.1007/S10064-017-1010-Y

11. S.P. Pradhan, T. Siddique, Stability assessment of landslide-prone road cut rock slopes in Himalayan terrain: A finite element method based approach. J. Rock Mech. Geotech. Eng. **12**(1), 59–73 (2020, Feb). https://doi.org/10.1016/J.JRMGE.2018.12.018

12. S. Lee, J.A. Talib, Probabilistic landslide susceptibility and factor effect analysis. Environ. Geol. **47**(7), 982–990 (2005, Mar). https://doi.org/10.1007/S00254-005-1228-Z

13. O.H. Ozioko, O. Igwe, GIS-based landslide susceptibility mapping using heuristic and bivariate statistical methods for Iva Valley and environs Southeast Nigeria. Environ. Monit. Assess. **192**(2) (2020, Feb). https://doi.org/10.1007/S10661-019-7951-9

14. J. Dou et al., Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One **10**(7), e0133262 (2015, July). https://doi.org/10.1371/JOURNAL.PONE.0133262

15. T. Ghosh, S. Bhowmik, P. Jaiswal, S. Ghosh, D. Kumar, Generating substantially complete landslide inventory using multiple data sources: A case study in Northwest Himalayas, India. J. Geol. Soc. India **95**(1), 45–58 (2020, Jan). https://doi.org/10.1007/S12594-020-1385-4

16. Q. Wang, W. Li, W. Chen, H. Bai, GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang county of Baoji city, China. J. Earth Syst. Sci. **124**(7), 1399–1415 (2015). https://doi.org/10.1007/s12040-015-0624-3

17. L.-J. Wang et al., A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Gesc. J. **20**(1), 117–136 (2016, Feb). https://doi.org/10.1007/S12303-015-0026-1

18. X. Lei, W. Chen, B.T. Pham, Performance evaluation of GIS-based artificial intelligence approaches for landslide susceptibility modeling and spatial patterns analysis. ISPRS Int. J. Geo-Inf. **9**(7), 443 (2020, July). https://doi.org/10.3390/IJGI9070443

19. T. Xiong, I.G.B. Indrawan, D.P. Eka Putra, Landslide susceptibility mapping using analytical hierarchy process, statistical index, index of enthropy, and logistic regression approaches in the TinalahWatershed, Yogyakarta. J. Appl. Geol. **2**(2), 67 (2018). https://doi.org/10.22146/jag.39983

20. E. Kutlug Sahin, C. Ipbuker, T. Kavzoglu, Investigation of automatic feature weighting methods (Fisher, Chi-square and Relief-F) for landslide susceptibility mapping. Geocarto Int. **32**(9), 956–977 (2017, Sept). https://doi.org/10.1080/10106049.2016.1170892

21. H.R. Pourghasemi, N. Kerle, Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environ. Earth Sci. **75**(3), 1–17 (2016, Jan). https://doi.org/10.1007/S12665-015-4950-1

22. N.M. Yusof, B. Pradhan, H.Z.M. Shafri, M.N. Jebur, Z. Yusoff, Spatial landslide hazard assessment along the Jelapang Corridor of the North-South Expressway in Malaysia using high resolution airborne LiDAR data. Arab. J. Geosci. **8**(11), 9789–9800 (2015). https://doi.org/10.1007/s12517-015-1937-x

23. A. Arabameri, B. Pradhan, K. Rezaei, C.-W. Lee, Assessment of landslide susceptibility using statistical- and artificial intelligence-based FR–RF integrated model and multiresolution DEMs. Remote Sens. **11**(9), 999 (2019, Apr). https://doi.org/10.3390/RS11090999

24. B. Pradhan, M.H. Abokharima, M.N. Jebur, M.S. Tehrany, Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS. Nat. Hazards **73**(2), 1019–1042 (2014). https://doi.org/10.1007/S11069-014-1128-1

25. M. Zare, M.H. Jouri, T. Salarian, D. Askarizadeh, Comparing of bivariate statistic, AHP and combination methods to predict the landslide hazard in northern aspect of Alborz Mt. (Iran). Int. J. Agric. Crop Sci. (JANUARY), 543–554 (2014)

26. W. Chen et al., GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based 2 random forest method (2018)

27. A. Mohan, A.K. Singh, B. Kumar, R. Dwivedi, Review on remote sensing methods for landslide detection using machine and deep learning. Trans. Emerg. Telecommun. Technol. **32**(7), e3998 (2021, July). https://doi.org/10.1002/ETT.3998

28. J. Roy, S. Saha, A. Arabameri, T. Blaschke, D.T. Bui, A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India. Remote Sens. **11**(23), 2886 (2019)

29. L.-L. Liu, C. Yang, X.-M. Wang, Landslide susceptibility assessment using feature selection-based machine learning models. GECE, 25–28 (2020)

30. C. Romer, M. Ferentinou, Shallow landslide susceptibility assessment in a semiarid environment – A quaternary catchment of KwaZulu-Natal, South Africa. Eng. Geol. **201**, 29–44 (2016, Feb). https://doi.org/10.1016/J.ENGGEO.2015.12.013

31. J. Dou et al., An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat. Hazards **78**(3), 1749–1776 (2015, May). https://doi.org/10.1007/S11069-015-1799-2

32. B. Pradhan, A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. CG **51**, 350–365 (2013, Feb). https://doi.org/10.1016/J.CAGEO.2012.08.023

33. N. Micheletti et al., Machine learning feature selection methods for landslide susceptibility mapping. Math. Geosci. **46**, 33–57 (2014). https://doi.org/10.1007/s11004-013-9511-0

34. R.A. El-Rashidy, S.M. Grant-Muller, An assessment method for highway network vulnerability. J. Transp. Geogr. **34**, 34–43 (2014). https://doi.org/10.1016/j.jtrangeo.2013.10.017

35. H. Abedi Gheshlaghi, B. Feizizadeh, An integrated approach of analytical network process and fuzzy based spatial decision making systems applied to landslide risk mapping. J. African Earth Sci. **133**, 15–24 (2017, Sept). https://doi.org/10.1016/J.JAFREARSCI.2017.05.007

36. Q. He et al., Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF classifier, and RBF network machine learning algorithms. Sci. Total Environ. **663**, 1–15 (2019, May). https://doi.org/10.1016/J.SCITOTENV.2019.01.329

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

38. X. Chen, W. Chen, GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. Catena **196**(August 2020), 104833 (2021). https://doi.org/10.1016/j.catena.2020.104833

39. W. Chen, X. Xie, J. Peng, J. Wang, Z. Duan, H. Hong, GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. Geomatics Nat. Hazards Risk **8**(2), 950–973 (2017, Dec). https://doi.org/10.1080/19475705.2017.1289250

40. A. Arabameri et al., Novel credal decision tree-based ensemble approaches for predicting the landslide susceptibility. Remote Sens. **12**(20), 3389 (2020, Oct). https://doi.org/10.3390/RS12203389

41. I. Gandhi, M. Pandey, Hybrid ensemble of classifiers using voting. Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015, 399–404 (2016, Jan). https://doi.org/10.1109/ICGCIOT.2015.7380496

42. W. Chen et al., Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. Appl. Sci. **8**(12), 2540 (2018, Dec). https://doi.org/10.3390/APP8122540

43. A.A. Othman, R. Gloaguen, Automatic extraction and size distribution of landslides in kurdistan region, NE Iraq. Remote Sens. **5**(5), 2389–2410 (2013). https://doi.org/10.3390/rs5052389

44. R.S. Banshtu, L.D. Versain, An inventory study on Landslide Hazard Zonation of Kullu Valley of Central Himalayan zone, Himachal Pradesh, India. Int. Academy Engin. **1**, 8–11 (2015). https://doi.org/10.15242/iae.iae0315417

45. V. Bandhu, S. Chandel, Geo-physical disasters in Himachal Pradesh: A spatial perspective (2016, Apr)

46. S. Kahlon, V.B.S. Chandel, K.K. Brar, Landslides in Himalayan mountains: A study of Himachal Pradesh, India. Int. J. IT, Eng. Appl. Sci. Res. Int. Res. J. Consort. **3**(9), 2319–4413 (2014)

47. J. Dou et al., Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One **10**(7) (2015). https://doi.org/10.1371/journal.pone.0133262

48. D. Tien Bui, B. Pradhan, O. Lofman, I. Revhaug, O.B. Dick, 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 **96**, 28–40 (Sep. 2012). https://doi.org/10.1016/J.CATENA.2012.04.001

49. B.C. Sujeewon, R. Sarkar, Landslide susceptibility mapping using GIS-based frequency ratio approach in Part of Kullu District, Himachal Pradesh, India, in *Geohazard Mitigation*, (2022), pp. 185–200. https://doi.org/10.1007/978-981-16-6140-2_16

50. I. Peshevski et al., Preliminary regional landslide susceptibility assessment using limited data. Geol. Croat. **72**(1), 81–92 (2019). https://doi.org/10.4154/gc.2019.03

51. A. Aditian, T. Kubota, Y. Shinohara, Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology **318**, 101–111 (2018, Oct). https://doi.org/10.1016/J.GEOMORPH.2018.06.006

52. K.C. Devkota et al., 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. Nat. Hazards **65**(1), 135–165 (2013, Jan). https://doi.org/10.1007/S11069-012-0347-6

53. H.R. Pourghasemi, B. Pradhan, C. Gokceoglu, Remote sensing data derived parameters and its use in landslide susceptibility assessment using Shannon's entropy and GIS. Appl. Mech. Mater. **225**, 486–491 (2012). https://doi.org/10.4028/www.scientific.net/AMM.225.486

54. B. Pradhan, Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. AdSpR **45**(10), 1244–1256 (2010, May). https://doi.org/10.1016/J.ASR.2010.01.006

55. G. Demir, M. Aytekin, A. Akgun, Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar–Resadiye (Tokat, Turkey). Arab. J. Geosci. **8**(3), 1801–1812 (2014, Mar). https://doi.org/10.1007/S12517-014-1332-Z

56. C.L. Salui, Methodological validation for automated lineament extraction by LINE method in PCI geomatica and MATLAB based hough transformation. J. Geol. Soc. India **92**(3), 321–328 (2018, Sept). https://doi.org/10.1007/S12594-018-1015-6

57. V.M. Choubey, P.K. Mukherjee, B.S. Bajwa, V. Walia, Geological and tectonic influence on water–soil–radon relationship in Mandi–Manali area, Himachal Himalaya. Environ. Geol. **52**(6), 1163–1171 (2006, Nov). https://doi.org/10.1007/S00254-006-0553-1

58. R.C. Patel, V. Adlakha, P. Singh, Y. Kumar, N. Lal, Geology, structural and exhumation history of the Higher Himalayan Crystallines in Kumaon Himalaya, India. J. Geol. Soc. India **77**(1), 47–72 (2011, Jan). https://doi.org/10.1007/S12594-011-0008-5

59. B. Stres, W.J. Sul, B. Murovec, J.M. Tiedje, Recently deglaciated high-altitude soils of the Himalaya: Diverse environments, heterogenous bacterial communities and long-range dust inputs from the upper troposphere. PLoS One **8**(9), Sep (2013). https://doi.org/10.1371/JOURNAL.PONE.0076440

60. Z. Chang et al., Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models. Remote Sens. **12**(3) (2020, 10.3390/rs12030502)

61. S. Lee, B. Pradhan, Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides **4**(1), 33–41 (2007). https://doi.org/10.1007/s10346-006-0047-y
62. J. Dou et al., Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. Remote Sens. **11**(6), 638 (2019, Mar). https://doi.org/10.3390/RS11060638
63. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995, Sept). https://doi.org/10.1007/BF00994018
64. A.X. Zhu, Y. Miao, L. Yang, S. Bai, J. Liu, H. Hong, Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. Catena **171**, 222–233 (2018, Dec). https://doi.org/10.1016/J.CATENA.2018.07.012
65. D.T. Bui, P. Tsangaratos, V.T. Nguyen, N. Van Liem, P.T. Trinh, Comparing the prediction performance of a deep learning neural network model with conventional machine learning models in landslide susceptibility assessment. Catena **188**(July 2019), 104426 (2020). https://doi.org/10.1016/j.catena.2019.104426