

Chapter 16

Landslide Susceptibility Mapping Using Supervised Learning Methods – Case Study: Southwestern Colombia



N. A. Correa-Muñoz, L. J. Martinez-Martinez, and C. A. Murillo-Feo

Abstract Landslides are among the more severe geological hazards that threaten and influence infrastructure stability in populated areas. Landslide susceptibility is defined as the spatial distribution of favourable conditions for future landslide events. This research aimed to integrate explanatory variables, such as geomorphometric, soil properties, and climate, into a random forest (RF) analysis and logistic regression (LR) to identify areas susceptible to landslides. Landslide inventory data were used to develop prediction models and validate them. The prediction of landslide events with the LR had an area under the curve (AUC) of 0.91, and the more important predicting factors were road distance, soil silt, soil sand and soil clay contents, elevation, soil drainage, TRI, landscape, soil depth, and slope. The landslide type prediction based on the RF analysis had an overall accuracy of 72%, with elevation, soil silt content, slope, TRI, landscape unit, soil clay and sand contents, and roads distance as the more important predictors.

Keywords Landslide · Terrain analysis · DEM · Soil survey · Random forest · Logistic regression

N. A. Correa-Muñoz
Civil Engineering Faculty, Universidad del Cauca, Popayán, Colombia
e-mail: nico@unicauca.edu.co

L. J. Martinez-Martinez (✉)
Agricultural Sciences Faculty, Universidad Nacional de Colombia, Bogotá, Colombia
e-mail: ljmartinezm@unal.edu.co

C. A. Murillo-Feo
Engineering Faculty, Universidad Nacional de Colombia, Bogotá, Colombia
e-mail: camurillof@unal.edu.co

16.1 Introduction

Landslide hazard maps represent susceptibility, which is the likelihood of a potentially damaging landslide occurring within a given area (Department of Regional Development and Environment 1991). The purpose of susceptibility studies is to identify areas where landslides can initiate and propagate (Guzzetti et al. 2005), based on a hazard and risk evaluation. Landslides can occur because of the interaction of natural and anthropic factors and cause economic and environmental damage and human losses. The Servicio Geológico Colombiano (2017) published a hazard map for landslides, 1:100,000 scale, and concluded that approximately 4% of the country has an extremely high hazard, 20% has a high hazard, 22% has a medium hazard, and 50% has low susceptibility. According to Suárez (1998), tropical mountainous areas are very susceptible to mass movements because of the interaction of slope, seismicity, rock type, and heavy rain, which are crucial factors for landslide events.

Different landslide-conditioning factors have been identified and can be used to establish landslide susceptibility assessments, including rainfall, drainage and soil properties (Metz and Bear-Crozier 2014), landcover (Shu et al. 2019) lithology, lineaments, geomorphology, soil type, and depth, slope angle, slope aspect, curvature, altitude, properties of the lithological material, land use patterns, and drainage networks (Youssef and Pourghasemi 2021), human activities (Achour and Pourghasemi 2020), and soil moisture (Ray et al. 2010). According to Gruber et al. (2009), mass movements are strongly controlled by land-surface form. A landslide process can be triggered by heavy and prolonged rainfall, cutting into slopes for the construction of roads, mining excavation without adequate preventive management, volcanoes, land-use change, and deforestation (Guzzetti et al. 2005).

The assessment of landslide susceptibility at different scales can be accomplished with statistical methods, heuristic approaches, physical models, and spatial models. For mapping landslide susceptibility, various methods have been developed that involve the identification of causative factors and the spatial analysis of interactions, usually supported by remote sensing data. For detailed landslide vulnerability mapping, Yuvaraj and Dolui (2021) used frequency ratio and binary logistic regression, Yu et al. (2021) studied the influence of rock and soil factors on landslide susceptibility mapping with LR modelling, an artificial neural network and support vector machine, Guo et al. (2021) presented a machine learning approach based on the C5.0 decision tree model and the K-means cluster algorithm to produce a regional landslide susceptibility map, Zhou et al. (2021) developed a hybrid model to optimize the factors and enhance the predictive ability of landslide susceptibility modelling, and (S. Lee et al. 2003) carried out a landslide susceptibility analysis using an artificial neural network, weights of evidence (Q. Wang et al. 2019), and LR (Dahoua et al. 2017).

The landslide inventory for Colombia indicates a high occurrence, in mountainous area, and, although there are various statistical, physical, and heuristic models to determine susceptibility, not all of them apply to every area since several factors

that affect landslides have a local behaviour that must be analyzed. On the other hand, it is important to have methods that use information that is more commonly available in various countries, such as soil studies, digital elevation models, and climate data. This research aimed to evaluate the use of supervised learning methods for mapping susceptibility to landslides in mountainous areas to generate information for decisions in risk and hazard assessments, planning, infrastructure development, and promotion of economic activities.

16.2 Materials and Method

16.2.1 Study Area

The study area was in southwestern Colombia, specifically in Cauca (Fig. 16.1), within the coordinates 0°57'27.07" N, 77°19'48.75" W, and 2°15'57.60" N, 76°04'33.53" W, in eleven municipalities covering 8488 km². The climate is tropical, with an average annual rainfall of 2382 mm per year. The study area included the Cauca Boot, an area with important ecological and geological meaning since it includes the Santa Rosa link, which connects the Central and East mountainous ranges in Colombia, generating geographic knotting (Hubach 1982), and is a natural access to the Colombian Amazon region.

16.2.2 Data Collection

A SRTM DEM with a spatial resolution of 1 arcsec was compared in terms of accuracy to a self-produced model using an interpolated 1:25,000 topographic digital map. The fill sinks algorithm (L. Wang and Liu 2006) was applied to identify and fill surface depressions in the DEM to prepare the data for the analysis. The accuracy assessment of the DEMs was performed using the DEMANAL software developed by Leibniz University (Jacobsen 2019).

The following parameters were obtained from the DEM with the software SAGA (Conrad et al. 2012): elevation as the primary data given by the DEM; slope defined as the tangent of a plane relative to the surface topography; the aspect, which refers to the direction of slope (Olaya 2009); curvature calculated based on second derivatives for a topographic attribute that describes the convexity or concavity of a terrain surface (Romstad and Etzelmüller 2012); flow accumulation determined by accumulating the weight for all cells that flow into each downslope cell (O'Callaghan and Mark 1984) and was derived with the Top-Down method in SAGA software, as described by (Szypuła 2017), topographic wetness index (TWI), calculated as a second-order derivative of the DEM and used as an indicator of water accumulation in an area of the landscape where water is likely to concentrate through runoff

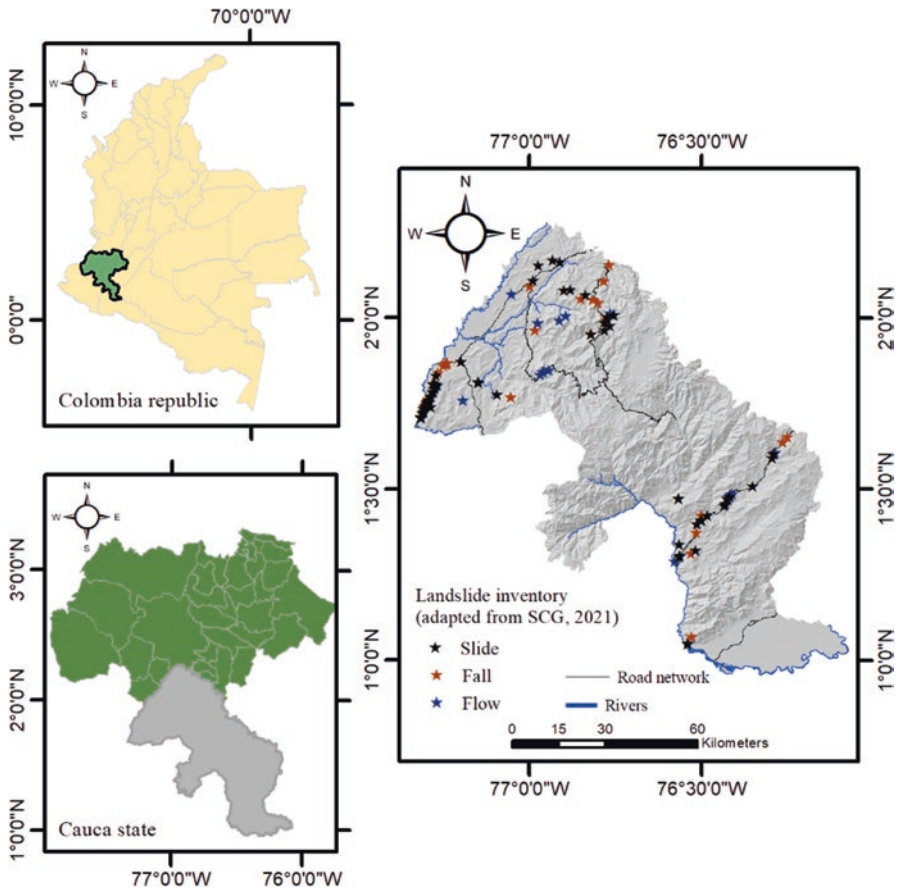


Fig. 16.1 Location of the study area

(Quinn et al. 1991), or, as described by (Vijith 2019), is a parameter that describes the tendency of a cell to accumulate water, topographic ruggedness index (TRI) expresses the elevation difference between adjacent cells of a DEM (Shawn Riley et al. 1999), stream power index (SPI) measures the erosive power of flowing water based on slope and specific catchment area (Moore et al. 1991), LS factor slope length (LS) factor as used by the Universal Soil Loss Equation (USLE) (Böhner and Selige 2006), topographic position index (TPI) indicates the altitude of each data point evaluated against its neighbours, (Guisan et al. 1999), and a description of these landform elements found in (Pike et al. 2009).

The landscape units and soil properties sand, silt and clay contents, soil depth, drainage class, and soil moisture regime were obtained from the General Soil Survey of the Cauca Department, scale 1:100,000 (Instituto Geográfico Agustín Codazzi 2009). Mean annual rainfall data were obtained from the Worldclim dataset (Fick

and Hijmans 2017), the landslide map of the study area was created with the national landslide inventory (SIMMA) (Servicio Geológico Colombiano 2021).

16.2.3 Method

A flowchart about the overall steps followed by the landslide susceptibility mapping is in Fig. 16.2.

All the variables were resampled at a cell size of 100 m and combined in a multidimensional raster in QGIS. The statistics of the multidimensional image were re-built using ArcCatalog of ArcGIS 10.8. Then, the RF method was run using the script adapted from the NASA-ARSET webinar in 2019 for SAR applications. Next, the accuracy metrics, such as OOB estimate of error rate, confusion matrix, mean decrease accuracy, overall accuracy, kappa, users, and producers' accuracy, were obtained.

For mapping susceptibility to landslides, a RF classification was applied using version 2021.09.1 of the R statistical software (RStudio, Inc.). The RF tree was built by training each decision tree (mtree) with a random subset of the predictor-variable (mtree) from the training dataset. The algorithm was applied with the training dataset of landslides to obtain the supervised classifier algorithm and validation dataset of landslides to assess the accuracy of the produced landslide classification map. The prediction model of the RF classifier only required the number of classification trees

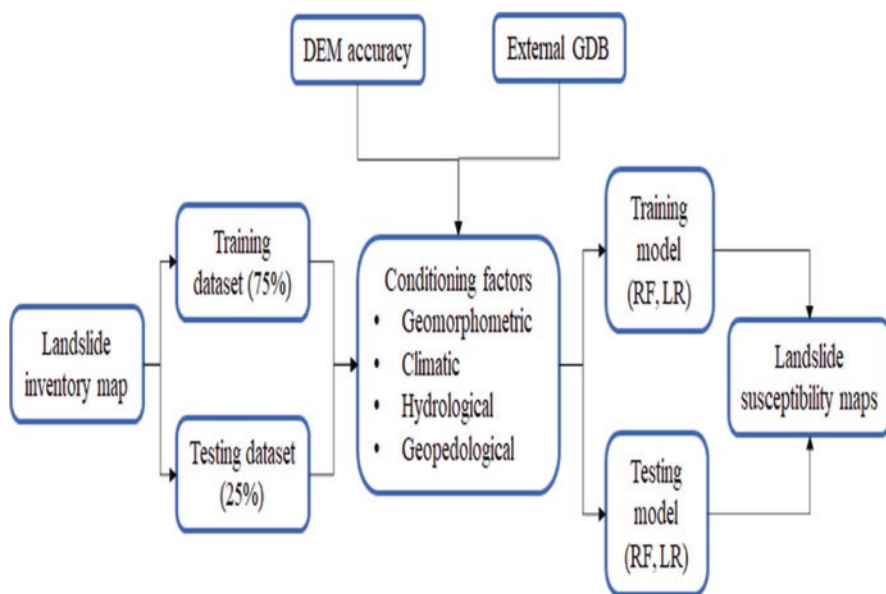


Fig. 16.2 Flowchart of the supervised learning method for landslide susceptibility mapping

(1000) and the number of prediction variables (18). The proportion used in this study was 75:25, as in the study by (Pham et al. 2018).

To map the probability of landslides occurrence, a binomial LR was applied. This statistical method has been well documented in geomorphological studies and is one of the most widespread methods for developing prediction models in geomorphology when system properties are represented by a binary variable (Schoch et al. 2018). The analysis was performed in the R software, and a training model was built using the ‘glm’ function with the binomial family. This model was assessed with the Chi-squared test, generated with 75% of the landslides inventory data and assessed with the remaining 25% of landslide inventory data using the area under the curve (AUC) as a validation metric of the prediction model (Huang and Zhao 2018).

16.3 Results and Discussion

16.3.1 SRTM DEM Accuracy Assessment

The standard deviation of the height was 11.58 m (Table 16.1), the bias was -2.3 m, and the standard deviation of the height without bias was 11.35 m. The normalized median absolute deviation (NMAD) related to bias-corrected height differences was 10.4 m. The SZ was greater than the NMAD because of a higher percentage of more significant discrepancies. This result agrees with the findings of Mukul et al.(2015), who compared the IGS and SRTM heights with the SRTM-DEM data in forest areas. The accuracy assessment of the SRTM DEM indicated an appropriate data quality for a landslide analysis since the results were equivalent to a scale about of 1:25 K, and the landslide analysis was done at a 1:100,000 scale.

16.3.2 Landslide Inventory

Table 16.2 shows the results of the landslide inventory of the study area, and its location is in Fig. 16.1. Following the Varnes classification (Hungr et al. 2014), it was found that 52.8% corresponded to slides that are displacements of material downslope, and 26.4% fit to falls that involve a collapse of material from the steepest area and accumulation in the base of the slope. 15.2% were classified as flows that are movements of materials down a hill as a fluid, 4.8% were creeps, defined as a slow downslope movement of material, and 0.8% were topples, the forward rotation and movement of material out of a slope.

Table 16.1 Results of the comparative of the 30 m SRTM DEM against 1:25 K topo-map DEM

Reference DEM	1:25K topo-DEM	RMSZ (m)	Bias (m)	SZ without bias	NMAD
DEM for analysis	30 m SRTM-DEM	11.583	-2.302	11.353	10.435

Table 16.2 Classification of landslide inventory. (Adapted from Servicio Geológico Colombiano 2021)

Type	Frequency		Subtype	Frequency	
	Absolute	Relative		Absolute	Relative
Falls	33	26.4%	Rockfall	24	19.2%
			Debris fall	6	4.8%
			Earthfall	3	2.4%
Slides	66	52.8%	Translational	46	36.8%
			Rotational	9	7.2%
			Planar	7	5.6%
			Wedge	2	1.6%
			Block slide	2	1.6%
Flows	19	15.2%	Debris flow	7	5.6%
			Mudflow	10	8.0%
			By flow	1	0.8%
			Earth flow	1	0.8%
Creeps	6	4.8%	Soil creep	6	4.8%
Topples	1	0.8%	Rock topples	1	0.8%
Total	125	100%		125	100%

Although the landslide distribution showed two geographically separated groups, the tendency of the mass movement distribution was preserved in each group. The dominant subtypes were translational debris (36.8%), rockfalls (19,2%), debris flows (5.6%) in the south-eastern zone, and mudflow (8%) in the north-western site. The landslide susceptibility analysis was developed using the type of movement for the RF method and the presence or absence of landslides as a binary dependent variable in the LR model.

16.3.3 Landslide Conditioning Factors

In this research, 18 landslide conditioning factors were selected based on literature review and the results of Colombian landslide inventory. A statistical summary of the distribution of each analyzed variable is in Table 16.3.

Elevations varied between 224 m.a.s.l. and 4158 m.a.s.l. (Fig. 16.3a) 26% of the area was below 1000 m.a.s.l, 33% was between 1000 and 2000 m.a.s.l., 29% was between 2000 and 3000, and 12% was over 3000 m.a.s.l.

The landscape units and its main characteristics are shown in Fig. 16.3f and in Table 16.4. A mountain landscape occupies 78% of the extension, hills represent 11%, and an alluvial valley contains 9%, and plateau 2%. The mountain is characterized by slopes greater than 30%, modelled by different geological phenomena associated with volcanic, structural, erosional, and depositional activity, which determines the current landscape characteristics. Most of the mountainous area was

Table 16.3 Statistics summary of the landslides conditioning factors

Layer	Factor	Unit	Analysis	Min	Mean	SD	Max	CV (%)
1	Elevation	m	Morpho	224.0	1775.95	934.35	4157.85	53
2	Slope	Degrees	Morpho	0	16.72	10.51	75.74	63
3	Aspect	Degrees	Morpho	0.0004	187.77	101.09	360	54
4	Curvature	Ordinal	Morpho					
5	Topographic position index	Ordinal	Morpho					
6	Terrain ruggedness index	None	Morpho	0	22.52	14.05	406.30	62
7	Flow accumulation	Km ²	Hydro	0.001	7.91	102.5	5148.68	1296
8	LS factor	None	Hydro	0	26.80	46.23	3973.30	173
9	Stream power index	None	Hydro	0	5381.34	106425.39	20651448	1978
10	Topographic wetness index	None	Hydro	2.99	7.94	2.78	23.79	35
11	Landscapes	Category	Geomorpho					
12	Soil sand content	Percentage	Pedology	24.5	52.83	7.35	77	14
13	Soil clay content	Percentage	Pedology	5	20.4	10.61	47	52
14	Soil silt content	Percentage	Pedology	12.5	26.18	6.18	37	24
15	Soil drainage	Ordinal	Pedology					
16	Soil depth	Ordinal	Pedology					
17	Soil moisture regime	Ordinal	Pedology					
18	Rainfall	mm per year	Hydro	1320	2381.52	733	4705	31

Min minimum, *SD* standard deviation, *Max* maximum, *CV* coefficient of variation

developed on Cretaceous and Cenozoic sedimentary or on volcano-sedimentary and plutonic igneous rocks and is covered by volcanic ash. The hilly landscape is made up of areas with heights of less than 300 m with a slope between 7 and 12% although they can reach 50% locally, developed on Tertiary sedimentary rocks. The alluvial valley corresponds to flat areas formed by sediments transported by rivers and plateaus, which are flat areas located at the base of the hills. The dominant soils in the area are well-drained, deep to moderately deep, with loam, clay loam, sandy clay loam or sandy clay texture (Fig. 16.3d, e), and udic moisture regime. To a lesser extent, there are superficial or poorly drained soils or with an ustic or aquic moisture regime.

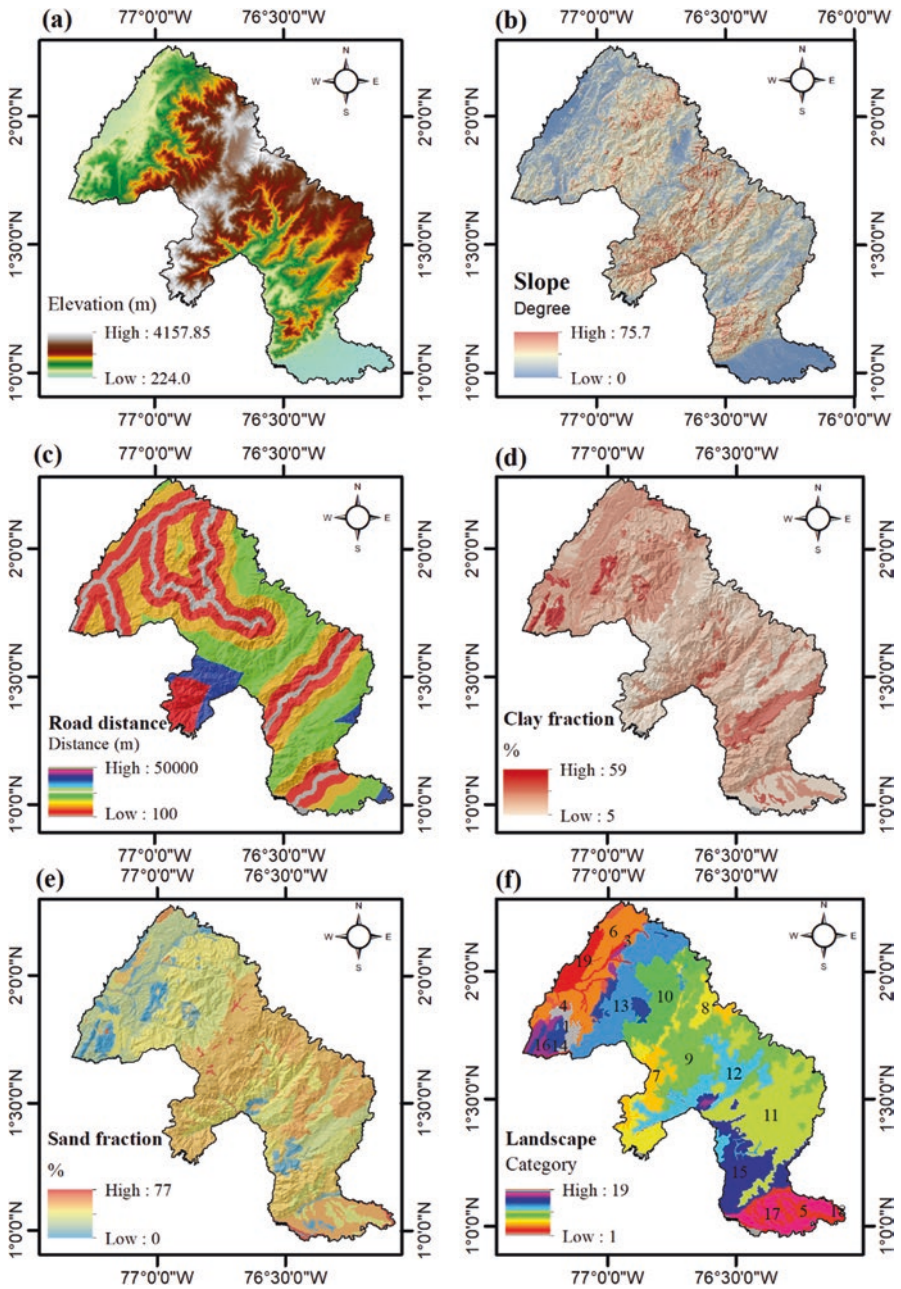


Fig. 16.3 Some factors influencing the landslide susceptibility

Table 16.4 Main characteristics of the landscape units

Landscape	Mean annual temperature (°C), mean annual rainfall (mm)	Relief type	Lithology	Number of mapping unit
Volcanic erosional plateau	18-24; 1000–2000	Plateau	Igneous rocks alternating with sedimentary materials	1
Erosional hill	18-24; 2000–4000	Hills	Volcanic ash layers over igneous rocks	3
	18-24; 1000–2000	Hills	Sedimentary rocks	4
	>24; 2000–4000	Hills	Metamorphic and sedimentary rocks	5
	>24; 1000–2000	Small valleys and hills	Colluvial-alluvial deposits, igneous rocks	6
Glacio-volcanic mountain	<12; 2000–4000	Summits	Igneous rocks	7
Fluvio-volcanic mountain	<12; 2000–4000	Summit, depressions	Volcanic ashes, organic materials	8
Volcanic-structural-erosional mountain	<12; 2000–4000	Summit, hills	Volcanic ashes	9
Volcanic mountain	12-18; 2000–4000	Summit, glacis	Volcanic ashes	10
Erosional-structural Mountain	18-24; 4000–8000	Summit, hills	Metamorphic rocks, volcanic ashes	11
Fluvio-volcanic mountain	18-24; 4000–8000	Summit, little valleys	Igneous rocks, colluvial deposits, volcanic ash	12
Erosional-structural Mountain, gravity flow	18-24; 2000–4000	Summit, hills, glacis	Sedimentary rocks, volcanic ash, igneous rocks	13
Structural erosional mountain and fluvial-gravitational mountain	18-24; 1000–2000	Summit, glacis	Volcanic ashes, metamorphic rocks	14
Erosional-structural Mountain	>24; 4000–8000	Summit, hills	Igneous rocks, volcanic and metamorphic rocks	15
Erosional-structural Mountain and small valley	>24; 1000–2000	Summit, fan, small valleys, flood plain	Sedimentary rocks, colluvial-alluvial deposits, igneous rocks	16
Alluvial valley	>24; 2000–4000	Terrace	Alluvial deposits	17
	>24; 1000–2000	Flood plain	Alluvial deposits	18
	>24; 1000–2000	Flood plain, terrace	Alluvial deposits	19

Adapted from IGAC (2009)

The slope varied between 0° and 76.4° (Fig. 16.3b). A classification of the landscape by its slope indicated a flat area in 13.6% of the study area, sloping areas in 68.3% of the extension, and steep areas with slopes greater than 30° in 18.1% of the zone. The slope aspect, indicating the flow-line direction, was distributed at 14.2% north-eastern, 31.4% south-eastern, 32.6% southwestern, and 21.8% north-western. The curvature plan indicated a concave surface in 48.7% of the area and a convex surface in 51.3% of the study area. The curvature profile indicated a concave form in 52.0% of the area and a convex form in 48.0% of the zone. The tangential curvature was concave in 48.2% of the cases and convex in 51.8% of the area.

The terrain ruggedness index varied between 0 and 275 and classified 53.6% of the area as smooth terrain, 42.0% as rough terrain, and 4.4% as irregular. The flow accumulation indicated that 92.5% of the drainage proportion was less than 0.47 km²; this accumulation reached 2163 km². The road distance (Fig. 16.3c) shows areas contiguous to the roads and others located up to 50 km.

Slope length (LS) is a topographic parameter used in soil erosion. Its mean value was 38 with a positively skewed distribution, which means that 88% of the distribution was less than 42.6. Stream power index describes potential flow erosion; its distribution was highly positively skewed. The topographic wetness index had a light, positively skewed distribution where about 88% of the distribution had an index less than 8.5 or a moderate wetness index. Finally, rainfall ranged from 1322 mm to 4705 mm per year, with a favourable bias distribution of 80.7% of the study area, less than 2584 mm per year.

16.3.4 LR Model for Landslide Probability Occurrence

The LR method was used to predict the probability of landslide occurrence based on the presence or absence of landslide events as a binary dependent variable based on 18 landslide conditioning factors as explanatory variables.

16.3.4.1 Training Model

Table 16.5 shows the results of the generalized linear model developed with a R script software using the glm-function and the logit-family of the binomial method. As with RF, the model was developed with 75% of the landslide data and validated with the remaining 25%. The results showed that road distance, highly significant relationship with the occurrence of landslides and in the limit, at a significance level of 0.01 are elevation and slope. The above indicated that the probability to obtain the coefficient of the model with respect to the hypothesis the true coefficient is zero was low. The coefficients of the other conditioning factors were not significant different to 0 they had no effect on the probability of landslide occurrence.

The chi square test (Table 16.6) indicated that the variables road distance, soil sand content and slope had a highly statistically significant association between the

Table 16.5 Results of the logit function in the landslide susceptibility analysis

Coefficients	Estimated coefficient	Code	Coefficients	Estimated coefficient	Code
Intercept	-5.291e+00	ns			
Clay content	-6.159e-02	ns	Silt content	2.380e-02	ns
Sand content	-6.326e-02	ns	Landscape	-5.514e-02	ns
Roads Distance	-5.039e-04	***	Soil depth	1.246e-01	ns
Drainage	-6.352e-02	ns	Slope	2.084e-01	*
Elevation	-5.660e-04	*	TRI	4.178e-02	ns

Significance: *** (p-value =0), *(p-value <0.01), ns: no significance

Table 16.6 Results of the Chi-square test of the prediction model of landslide susceptibility

Coefficients	Deviance	Code	Coefficients	Deviance	Code
Null					
Clay content	4.741	*	Silt content	0.000	
Sand content	54.469	***	Landscape	3.898	*
Roads Distance	187.748	***	Soil depth	0.381	ns
Drainage	0.033	ns	Slope	17.477	***
Elevation	1.316	ns	TRI	0.193	ns

Significance: *** (p-value =0), ** (p-value < 0.001), *(p-value <0.01), ns: no significance

observed and estimated values, soil clay content, and landscape unit had statistically effect on the landslide prediction.

The probability of landslide occurrence map was prepared with the LR equation using map algebra in ESRI's ArcMap v10.8 and reclassified with four classes with the quantile method (Fig. 16.4). The highest probability of occurrence was found near roads, the remaining area presents medium probability.

Additionally, the information library of the R software was applied to the training geodatabase to compute the weight of evidence and information value metrics. Distance to roads was the most important variable to explain the occurrence of landslides (Fig. 16.5), followed by TRI, soil silt, sand and clay content, elevation, soil drainage, TRI, and landscape type. The greater probability of occurrence of landslides in certain areas is related to natural factors such as edaphic, geomorphometric and climatic that facilitate the occurrence of events and with the anthropic activities, in this case roads construction, which are the trigger that activates the landslide phenomena.

Table 16.7 shows the most significant importance (values > +1) of the bivariate method of the weight of evidence. A positive weight indicates a positive correlation between the presence of the predictable variable and landslides (Jaafari et al., 2015). The conditioning factors within this category were road distance between 0 and 1000 m, soil sand content between 14.5% and 37%, soil drainage moderately well-drained, elevation in the range 607 m to 850 m, soil silt content in the range 36% to

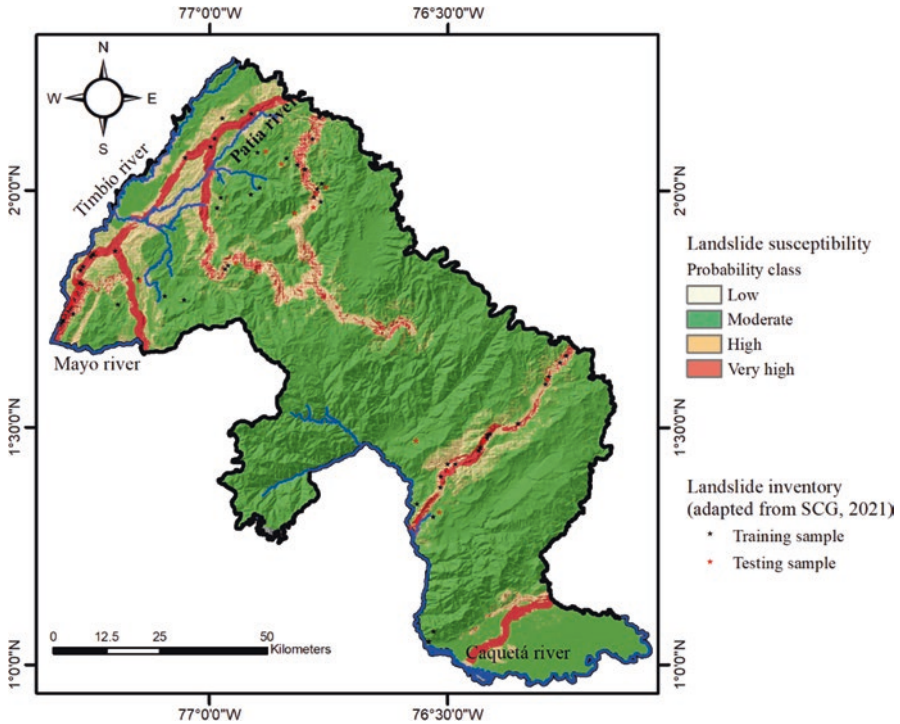


Fig. 16.4 Probability of landslide occurrence based on the LR method

46%, and soil clay content between 30.5% to 32, landscape units structural erosional mountain, mainly in ridges and back-slopes other factor showed also positive values.

16.3.4.2 Performance of the LR Model

The ROC curve and the AUC are standard measures for binary classifier performance. The ROC plot (Fig. 16.6) was obtained by plotting the valid positive rate (TPR) against the false positive rate (FPR), while AUC is the area under the ROC curve. As a rule of thumb, a model with good predictive ability should have an AUC greater than 0.5. The AUC of the landslide susceptibility model with regression analysis was 0.91. An analysis of scaling land-surface variables for landslide detection obtained AUC values between 0.73 and 0.80 (Sirbu et al. 2019). AUC values between 0.7 and 0.9 indicate a reasonable agreement between the predicted landslides and test landslides (Lee et al. 2018).

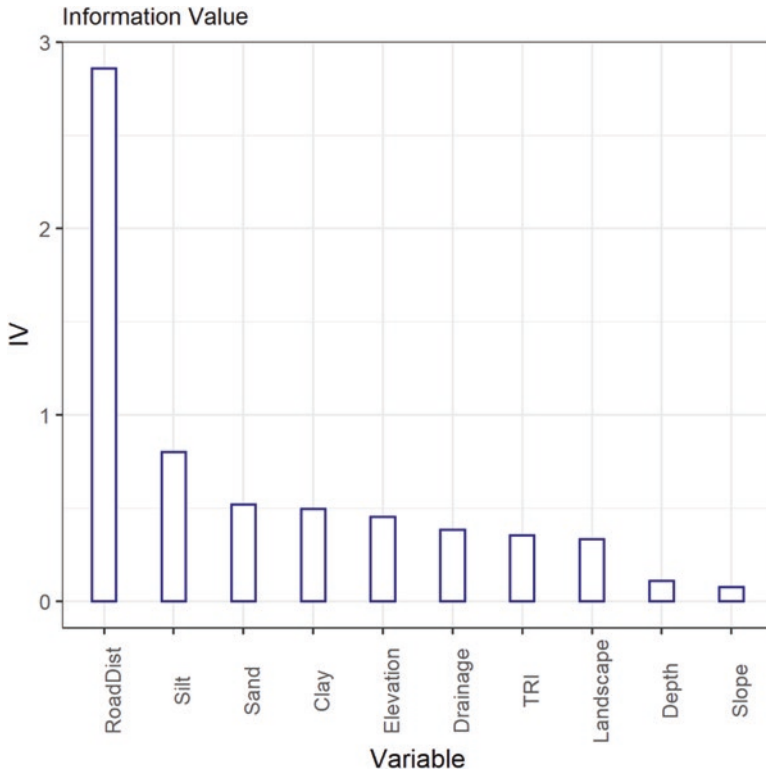


Fig. 16.5 Importance the conditioning factors in the LR model

Table 16.7 Weight of evidence of the variables and range of the category

Landslide conditioning factor	Unit	Area	WOE	IV	Category of the variable
Road distance	m	9.4%	2.11	1.43	0 to 1000
Soil sand content	%	9.6%	1.26	0.30	14.5 to 37
Soil drainage	Category	10.5%	1.25	0.32	2 to 3 (moderately well-drained)
Elevation	m a.s.l.	10.0%	1.18	0.28	606.8 to 850
Landscape	Category	10.3%	1.02	0.33	16 to 19
Soil silt content	%	12.8%	0.88	0.80	36 to 46
Soil clay content	%	9.3%	0.84	0.48	30.5 to 32

16.3.5 Landslide Susceptibility Zoning Based on RF Analysis

16.3.5.1 Training Model

To develop the predictive model of the landslide susceptibility the slides and falls types were selected since these were the more frequent events, and the 16 factors selected as the most relevant for the study area. From the inventory of landslides,

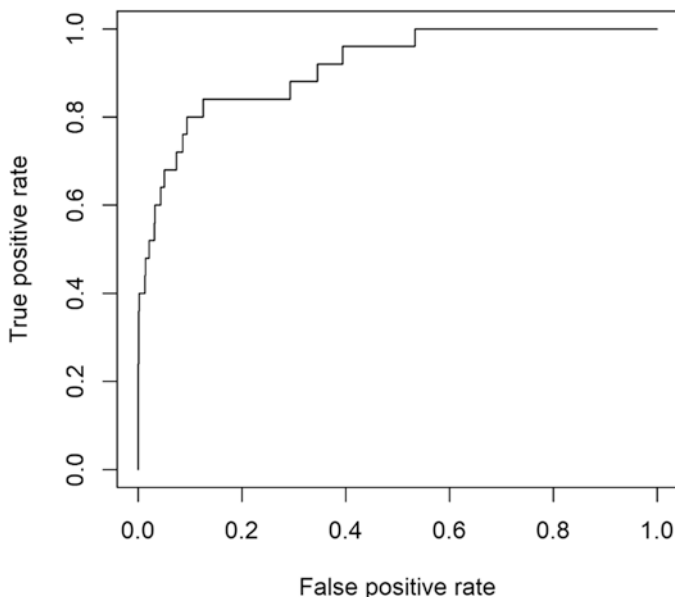


Fig. 16.6 ROC curve of the LR-landslide susceptibility model

Table 16.8 Results of the accuracy RF classification of the landslides

Type of movement	Class accuracy	
	User	Producer
Falls	56.0%	66.7%
Slides	85.7%	79.2%
Overall accuracy	75.7%	

75% were selected to generate the model, and the remaining 25% were used to validate it. To guarantee the stability of the model, 1000 trees were used, as recommended by Lagomarsino et al. (2017).

The overall classification accuracy of developed model (Table 16.8) was 75.7%, slides were the mass movements that were better classified with a user accuracy of 85.7%, which means the percentage of landslides that were correctly classified as compared to the landslide inventory. The producer accuracy refers to the commission error, which was 79.2%. The falls and flows of mass movements had lower user and producer accuracy. The general accuracy depended on the frequency of landslide events, the more frequent the occurrence of a landslide type, the greater accuracy obtained in the prediction, consistent with other researches (Tansey et al. 2004).

Although the general accuracy was low, when the classification for each landslide type was analyzed, good prediction accuracy was found for the landslides that are more frequent in the study area. The analysis was based on existing data, which

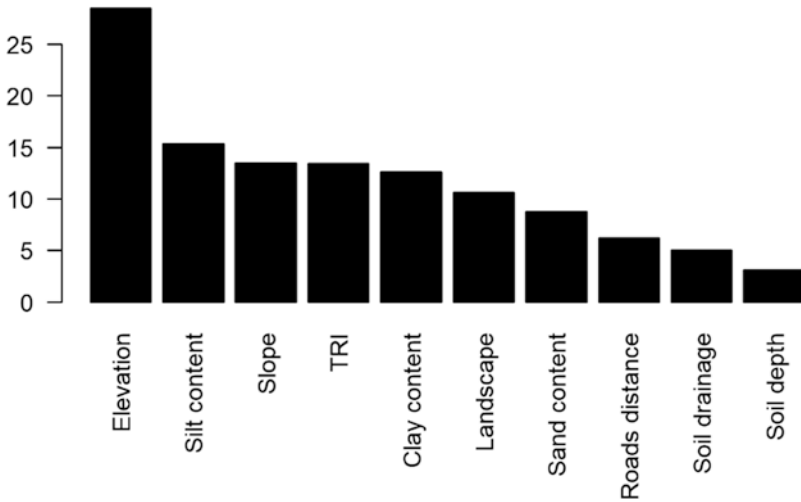


Fig. 16.7 Mean Decrease Accuracy in the study area

is one of the main limitations since some, such as climate and geology, were very general for the scale of the study.

16.3.5.2 Mean Decrease Accuracy (MDA)

The MDA was one of the outcomes of the RF analysis and indicated the degree of importance of each of the variables in the prediction. Figure 16.7 displays the MDA results, ranking the variables by importance. The more important variables in the RF prediction were elevation, soil silt content, slope, TRI, soil clay, landscape unit, soil sand content and roads distance. It was found that there was a relationship between some edaphic and geomorphometric characteristics with the presence of the main types of landslides. Most landslides were found in the structural erosional mountain, in ridges and back-slopes, in loamy and sandy loam soils with a humid climate, and at an altitude between 300 and 1200 m asl.

16.3.5.3 Accuracy of the Landslide Classification by the RF Method

The model assessment helped to evaluate the classifier performance for other data. Table 16.9 summarizes the accuracy metrics derived from the confusion matrixes, which compares the reference values with the predicted values.

The overall accuracy was 72% and indicated the percentage of landslide type correctly classified. When the accuracy of each class was evaluated, the slides had a better performance with 75% and 88.2% of user's and producer's accuracy, respectively, while falls had low accuracy. The error of commission of slide classification

Table 16.9 Evaluation of the model classification performance

Type of movement	Class accuracy	
	User	Producer
Falls	60.0%	37.5%
Slides	75.0%	88.2%
Overall accuracy	72.0%	

was 25%, and the omission error was 11.8% for the classification model, differentiating slides and non-slides for the study area.

According to Korup and Stolle (2014), predictive methods based on machine learning analysis achieve an overall success rate of 75–95%, these authors proposed doing more research on the selection of models, the model overfitting, and the effect of slope failure at a regional scale to improve predictions. In our case, another factor that influences the success of the predictions was the inventory of landslides, the relationship with the distance to the roads is sometimes due to that landslides were much more commonly recorded near roads (Stanley et al. 2020). The objective of our study was to evaluate predictive methods with data from the soil survey, geomorphometric parameters calculated from DEM and climate data available on the internet, however geology data were not included, and it could have effect on predictions. On the other hand, the rainfall data used had a spatial resolution of 900 m, which is exceptionally low compared to other data and therefore had no significant effect on the RF prediction.

The landslide classification map based on the results of the RF analysis (Fig. 16.8) shows the areas that meet the conditions required for the occurrence of the main landslide types and, therefore, are more likely to present this phenomenon. The classification showed that most of the area is susceptible to slides, and in less proportion the area can be affected by falls.

16.4 Conclusions

The probability of landslide events occurrence, estimated with LR, had an AUC of 0.91, and the more important predicting factors were road distance, soil silt, sand and clay content, elevation, soil drainage, TRI, landscape, soil depth and TWI. Landslide occurrence is favour by natural factors, while anthropic activities like the construction of roads is the trigger that initiates the process of occurrence of landslides. The susceptibility of the study area to the occurrence of landslides type based on RF analysis had an overall accuracy of 72% with elevation, soil silt content, slope, TRI, landscape unit, soil clay and sand content, and road distance were the more important predictors.

The integration of the DEM as a data source with the results of the soil surveys using LR and RF made it possible to generate information with acceptable reliability and level of detail for the susceptibility of mountain areas to landslides as a first

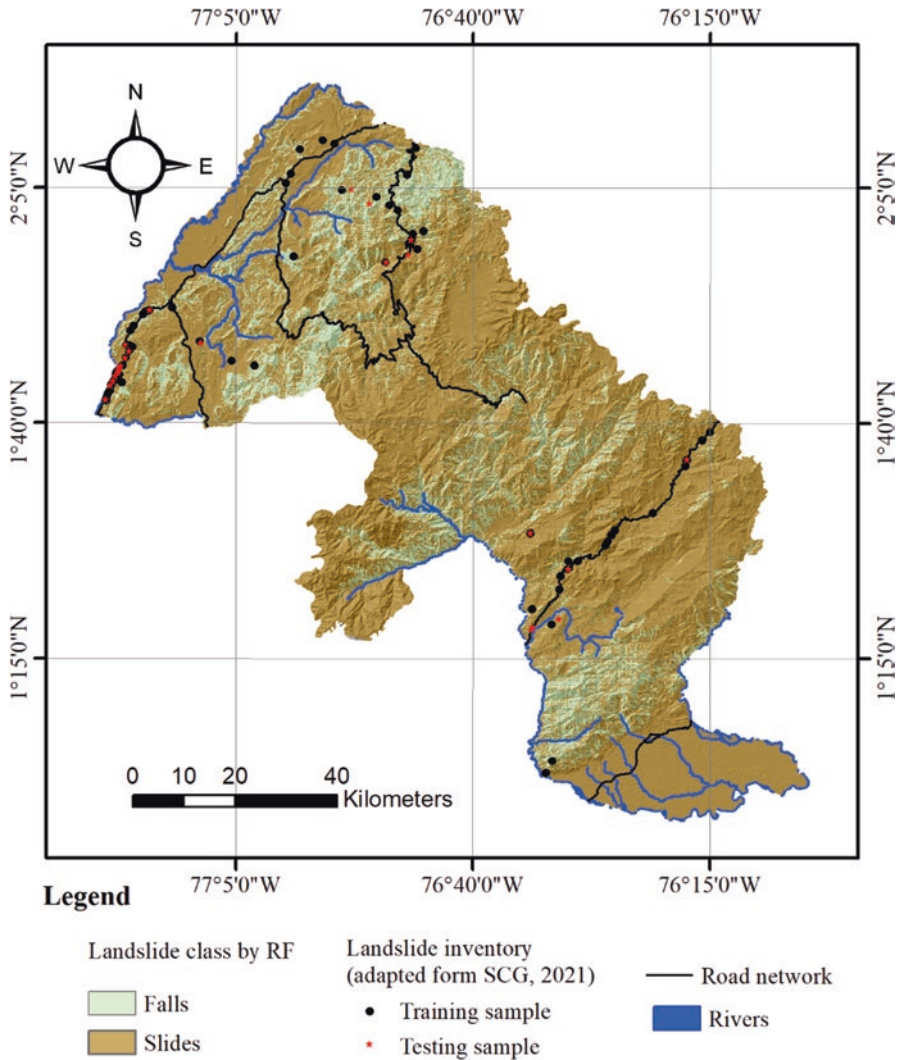


Fig. 16.8 RF classification of landslide type susceptibility areas

approximation for subsequent risks and hazard analyses. This is important considering that the required data is available for all of Colombia for applying more complex predictive models, where data availability and quality are limiting.

The distance to the roads was the factor that had the greatest incidence in the presence of landslides and in its distribution pattern. Consequently, it is also a factor that determines the probability of occurrence of landslides. Most of the study area has medium probability of occurrence but if roads are built it can change to high or very high probability.

Acknowledgments The authors want to thank the National University of Colombia and the Cauca University for the academic support for this research.

References

- Achour Y, Pourghasemi HR (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci Front* 11(3):871–883. <https://doi.org/10.1016/j.gsf.2019.10.001>
- Böhner J, Selige T (2006) Spatial prediction of soil attributes using terrain analysis and climate regionalisation
- Dahoua L, Yakovitch SV, Hadji R, Farid Z (2017) Landslide susceptibility mapping using analytic hierarchy process method in BBA-Bouira Region, case study of East-West Highway, NE Algeria. *Euro-Mediterranean Conference for Environmental Integration*, 1837–1840
- Department of Regional Development and Environment (1991) Primer on natural hazard management in integrated regional development planning. In Organization of American States, Department of Regional Development and Environment Executive Secretariat for Economic and Social Affairs Organization of American. <https://www.oas.org/dsd/publications/Unit/oea66e/begin.htm>
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37(12):4302–4315. <https://doi.org/10.1002/joc.5086>
- Gruber S, Huggel C, Pike R (2009) Modelling mass movements and landslide susceptibility. *Dev Soil Sci* 33(C):527–550. [https://doi.org/10.1016/S0166-2481\(08\)00023-8](https://doi.org/10.1016/S0166-2481(08)00023-8)
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143(1):107–122
- Guo Z, Shi Y, Huang F, Fan X, Huang J (2021) Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geosci Front* 12(6):101249. <https://doi.org/10.1016/j.gsf.2021.101249>
- Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72(1–4):272–299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *Catena* 165:520–529. <https://doi.org/10.1016/j.catena.2018.03.003>
- Hubach E (1982) El Cauca. Las unidades geográficas y geológicas del Departamento y los recursos del suelo y del subsuelo. In *Publicaciones Geológicas Especiales Del Ingeominas*, pp. 23–37
- Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. In *Landslides* 11(2):167–194. <https://doi.org/10.1007/s10346-013-0436-y>
- Instituto Geográfico Agustín Codazzi (2009) Estudio General de Suelos y Zonificación de Tierras del departamento del Cauca. Instituto Geográfico Agustín Codazzi
- Jaafari A, Najafi A, Rezaeian J, Sattarian A, Ghajar I (2015) Planning road networks in landslide-prone areas: a case study from the northern forests of Iran. *Land Use Policy* 47:198–208. <https://doi.org/10.1016/j.landusepol.2015.04.010>
- Jacobsen K (2019) DEMANAL Program System BLUH. In Institute of Photogrammetry and Geoinformation, Leibniz University, Hannover
- Korup O, Stolle A (2014) Landslide prediction from machine learning. *Geol Today* 30(1):26–33
- Lagomarsino D, Tofani V, Segoni S, Catani F, Casagli N (2017) A tool for classification and regression using random forest methodology: applications to landslide susceptibility mapping and soil thickness modeling. *Environ Model Assess* 22(3):201–214
- Lee S, Ryu JH, Min K, Won JS (2003) Landslide susceptibility analysis using GIS and artificial neural network. *Earth Surf Process Landf* 28(12):1361–1376. <https://doi.org/10.1002/esp.593>

- Lee JH, Sameen MI, Pradhan B, Park HJ (2018) Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology* 303:284–298. <https://doi.org/10.1016/j.geomorph.2017.12.007>
- Metz L, Bear-Crozier AN (2014) Landslide susceptibility mapping: A remote sensing-based approach using QGIS 2. 2 (Valmiera) (Vol. 2)
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5(1):3–30
- Mukul M, Srivastava V, Makul M (2015) Analysis of the accuracy of shuttle radar topography Mission (SRTM) height models using international global navigation satellite system service (IGS) network. *J Earth Syst Sci* 124(6):1343–1357. <https://doi.org/10.1007/s12040-015-0597-2>
- O’Callaghan JF, Mark DM (1984) The extraction of drainage networks from digital elevation data. *Comput Vision Graph Image Process* 28(3):323–344
- Olaya V (2009) Basic land-surface parameters. *Dev Soil Sci* 33(C):141–169. [https://doi.org/10.1016/S0166-2481\(08\)00006-8](https://doi.org/10.1016/S0166-2481(08)00006-8)
- Pham BT, Prakash I, Tien Bui D (2018) Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* 303:256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>
- Pike RJ, Evans IS, Hengl T (2009) Geomorphometry: concepts, software, applications. In: Hengl T, Reuter HI (eds) *Developments in soil science*, vol 33. Elsevier, pp 1–28
- Quinn PFB, Beven K, Chevallier P, Planchon O (1991) The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol Process* 5(1):59–79
- Ray RL, Jacobs JM, Cosh MH (2010) Landslide susceptibility mapping using downscaled AMSR-E soil moisture: a case study from Cleveland Corral, California. *US Remote Sens Environ* 114(11):2624–2636. <https://doi.org/10.1016/j.rse.2010.05.033>
- Romstad B, Etzelmüller B (2012) Mean-curvature watersheds: a simple method for segmentation of a digital elevation model into terrain units. *Geomorphology* 139:293–302
- Schoch A, Blöthe JH, Hoffmann T, Schrott L (2018) Multivariate geostatistical modeling of the spatial sediment distribution in a large-scale drainage basin, Upper Rhone, Switzerland. *Elsevier Geomorphology* 303:375–392. <https://doi.org/10.1016/j.geomorph.2017.11.026>
- Servicio Geológico Colombiano (2017) *Las Amenazas por movimientos en masa de Colombia. Colección Guías y Manuales*. SGC. https://srvags.sgc.gov.co/Archivos_Geportal/Manuales/Libro_MNMM.pdf
- Servicio Geológico Colombiano (2021) *Inventario nacional de movimientos en masa*. <https://www2.sgc.gov.co/ProgramasDeInvestigacion/geoamenazas/Paginas/Inventario-nacional-de-movimientos-en-masa-y-SIMMA.aspx>
- Shawn R, Stephen D, DeGloria ER (1999) A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain J Sci* 5:1–4
- Shu H, Hürlimann M, Molowny-Horas R, González M, Pinyol J, Abancó C, Ma J (2019) Relation between land cover and landslide susceptibility in Val d’Aran, Pyrenees (Spain): historical aspects, present situation, and forward prediction. *Sci Total Environ* 693:1–14. <https://doi.org/10.1016/j.scitotenv.2019.07.363>
- Sîrbu F, Drăguț L, Oguchi T, Hayakawa Y, Micu M (2019) Scaling land-surface variables for landslide detection. *Prog Earth Planet Sci* 6(1):1–13
- Stanley TA et al (2020) Building a landslide hazard indicator with machine learning and land surface models. *Environ Model Softw* 129:1–15
- Suárez J (1998) *Deslizamientos y estabilidad de taludes en Zonas Tropicales*. Instituto de Investigaciones sobre Erosión y Deslizamientos, pp 1–10
- Szypuła B (2017) Digital elevation models in geomorphology. *Hydro-Geomorphology-Models Trends InTechOpen* 2017b:81–112
- Tansey KJ, Luckman AJ, Skinner L, Balzter H, Strozzi T, Wagner W (2004) Classification of forest volume resources using ERS tandem coherence and JERS backscatter data. *Int J Remote Sens* 25(4):751–768. <https://doi.org/10.1080/0143116031000149970>

- Vijith H (2019) Modelling terrain erosion susceptibility of logged and regenerated forested region in northern Borneo through the Analytical Hierarchy Process (AHP) and GIS techniques. *Geoenviron Disasters* 6(8):18
- Wang L, Liu H (2006) An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int J Geogr Inf Sci* 20(2):193–213
- Wang Q, Guo Y, Li W, He J, Wu Z (2019) Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. *Geomat Nat Haz Risk* 10(1):820–835. <https://doi.org/10.1080/19475705.2018.1549111>
- Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir region. *Saudi Arab Geosci Frontiers* 12(2):639–655. <https://doi.org/10.1016/j.gsf.2020.05.010>
- Yu X, Zhang K, Song Y, Jiang W, Zhou J (2021) Study on landslide susceptibility mapping based on rock–soil characteristic factors. *Sci Rep* 11(1):1–27. <https://doi.org/10.1038/s41598-021-94936-5>
- Yuvaraj RM, Dolui B (2021) Statistical and machine intelligence-based model for landslide susceptibility mapping of Nilgiri district in India. *Environ Chall* 5(May):100211. <https://doi.org/10.1016/j.envc.2021.100211>
- Zhou X, Wen H, Zhang Y, Xu J, Zhang W (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci Front* 12(5):101211. <https://doi.org/10.1016/j.gsf.2021.101211>