

METHODOLOGY FOR ESTIMATING LANDSLIDES SUSCEPTIBILITY USING ARTIFICIAL NEURAL NETWORKS

E. N. Muñoz¹, O. D. Sánchez¹, F. L. Hernandez¹

¹ Remote Sensing Research Group, Universidad del Valle, Cali, Colombia
(evelin.munoz, oscar.llorente, francisco.hernandez)@correounivalle.edu.co

KEY WORDS: Mass Removal Phenomena, Artificial Neural Network, Interferometry, Sentinel 1B, GIS, Remote Sensing, DEM

ABSTRACT:

In this study, the susceptibility to landslides at Sevilla township, Valle del Cauca, located at southwest of Colombia was evaluated. The conditioning factors that involve the generation of landslides were evaluated using Geographic Information Systems (GIS) and Remote Sensing (RS) techniques. For the estimating susceptibility, an Artificial Neural Network (ANN) was implemented by applying the “Backpropagation” method to extract the synoptic weights of the conditioning variables (slopes, flow length, curvature, geology, fracture density, and land cover) on an automatic way with a data training module. The data for the analysis of the conditioning factors were carried out through a Digital Elevation Model (DEM) obtained through Radar Interferometry techniques, with Sentinel-1B satellite images for the year 2018. The results showed that Sevilla’s township has areas with high susceptibility, high slopes, and that it’s crossed by an active geological fault which implies that the earth’s dynamics will condition the terrain stability.

1. INTRODUCTION

The susceptibility of a land is known as soil’s tendency to generate phenomena on removal, depending on its intrinsic properties, that is, it shows the predisposition of the environment and the elements that make up the landscape such as, geomorphology, geology, coverage, etc.

During the years 1996 to 2002 at Sevilla Township there were 9 landslide events, which caused a total of 37 people dead, 104 injured and 141 affected population (Ingeominas, 2002). However, the only ones affected were not only the inhabitants of its urban capital, but also those of its surroundings, which have suffered great losses due to the interruptions of the social and economic dynamics that this phenomenon generates. An example of this occurred on December 24, 2012. A landslide of about 130,000 cubic meters of land, partially destroyed the road at Cristales sector, which led to subsidence on the road up to 3 meters. More than 7,000 peasants from 24 adjacent villages were isolated to the rest of the regions and this affected dramatically the economy of these sectors. (El País, 2013).

According to investigations carried out by the Colombian Geological Service (SGC), there are different methods for mapping landslides susceptibility, which are linked to factors such as: level of study detail, suggested scales for those levels, among others. The methods for obtaining a final susceptibility map are classified as semi-implicit and explicit (empirical, semi-analytical and analytical) (Ingeominas, 2002).

The explicit semi-analytical method, also known as heuristic, has been used by SGC, (2017), Wachal & Hudak, (2013), IDEAM, (2012), Parise, (2001) Sützen & Doyuran, (2004a, 2004b) and Van Westen, (1993), who classify the conditioning factors through the use of weighted scores and a combination of mathematical equations. The weights are assigned by specialists with experience in the study field who determine the values for each variable in order to indicate the contribution of each variable into terrain susceptibility.

The use of GIS is a key point for mapping landslides (Guzzetti, and Paola Carrara, 1999), as it allows different calculations

using large datasets. This tool is commonly used because it facilitates data management and mapping susceptibilities produced by each selected variable. For the calculation of the contribution of each variable to susceptibility, a statistical data management method known as a univariate method is used, and then, through a multivariate analysis, all considered variables are related (Chacón, et al., 2006; Dai, Lee, & Ngai, 2002; Parise, 2001; Sützen & Doyuran, 2004a, 2004b; Van Westen, 1993).

Kawabata & Bandibas (2009) and Pradhan & Lee (2010b, 2010a), propose a multivariate statistical technique using an ANN to calculate the landslides susceptibility produced by the set of variables considered. In this method, neurons are organized as sequential layers, each consisting of one or more neurons: the input layer, the hidden layer and the output layer. It is a feedback network where each route has an assigned weight that dictates the scale of the relationships between neurons. Weights can take positive or negative values, and their distribution within an ANN determines their performance by using information processing.

Machine Learning tools such as Artificial Neural Networks (ANN) are important because they allow obtaining a prediction of susceptibility in a given area taking into account the selected variables (Kawabata & Bandibas, 2009; Pradhan & Lee, 2010b, 2010a). The quality of the result obtained depends on different aspects such as: the inputs entered, the type of variables to be used: continuous or discrete, the pixel size, the polygons of events for training and testing, among others.

Bui et al, (2013) used Support Vector Machines (SVM) for the calculation of susceptibility in Hoa Binh province, Vietnam. SVM constructs an n-dimensional hyperplane which separates the classes represented as sample points in space, maximizing the distance from the nearest samples and the hyperplane that separates them. With the above, a map with two classifications (with susceptibility - without susceptibility) is obtained as the final forecast.

At present, landslides remain latent at Sevilla township, Valle del Cauca, therefore, organizations such as the Regional Autonomous Corporation of the Valle del Cauca (CVC) and the Seismological and Geophysical Observatory of the Colombian Southwest (OSSO) periodically follow-up the zone through technical visits to the reported affected areas (CVC & OSSO, 2011); In addition, the Colombian Geological Service (SGC) has studies of this nature, however, they are not updated; leaving aside the possibility of carrying out precise and periodic model that allows prevention and mitigation plans for anticipating a new event.

This article presents the methodology applied to estimate the susceptibility to landslides in the municipality of Sevilla, Valle del Cauca. An ANN was implemented, applying the “Backpropagation” method to extract the synoptic weights of the conditioning variables. The slope, flow length and curvature variables were obtained from a DEM generated by Radar interferometry (INSAR). Random Forest (RF) supervised classification was performed to obtain coverage types.

2. STUDY AREA AND MATERIALS

Sevilla is a municipality located on the western flank of the central mountain range of Valle del Cauca - Colombia, composed of mountains of asymmetric and irregular slopes. It has slopes between 0 to 30% at central area, and between 30 to 100% at the north side (SGC, 2013). It is inhabited by approximately 45,152 people, who are in risk due to conditioning and detonating factors that cause landslides.

Landsat 8 OLI Optical satellite images and Sentinel 1B radar images for the second half of year 2018, were used for the study. The data was downloaded through NASA's Alaska Vertex platform (<https://www.asf.alaska.edu>). Information regarding geology and fracture density in shapefile format at 1:25 000 scale were provided by CVC

Landslide inventory (coordinates, date of occurrence, and sliding type) was obtained through the geovisor belonging to the SGC web portal (<http://simma.sgc.gov.co>).

3. METHODOLOGY

3.1. Digital Elevation Model Generation

The DEM was performed using Sentinel 1B and 1A images of descending order and polarization "VV", corresponding to 2018, applying the synthetic aperture radar interferometry technique (InSAR). For processing, two images were selected per each semester with a short time difference between them. The oldest image was assigned as “master” image and the most recent as “slave” (Ivan, Singleton, Horák, & Inspektor, 2017).

The images were processed through the SNAP software of the European Space Agency (ESA), using the following steps: co-registration, interferogram formation, deburst, Goldstein phase filter, unwrapping phase, elevation phase and Doppler terrain correction (Pepe & Calò, 2017). SNAPHU was implemented in Linux operating system for image development, which was subsequently used for an elevation phase generation (Geymen, 2014).

3.2. Variables parametrization process

The variables curvature, flow length, and slope, were extracted from the DEM using the Arc toolbox tools of the ArcGIS 10.3© software. On the other hand, the geology and fracture

density variables were obtained according to the “.shp” file provided by the CVC; finally, the ground cover was acquired from processing Landsat 8 images. Figure 1 shows the methodological scheme for estimating the susceptibility by using an ANN.

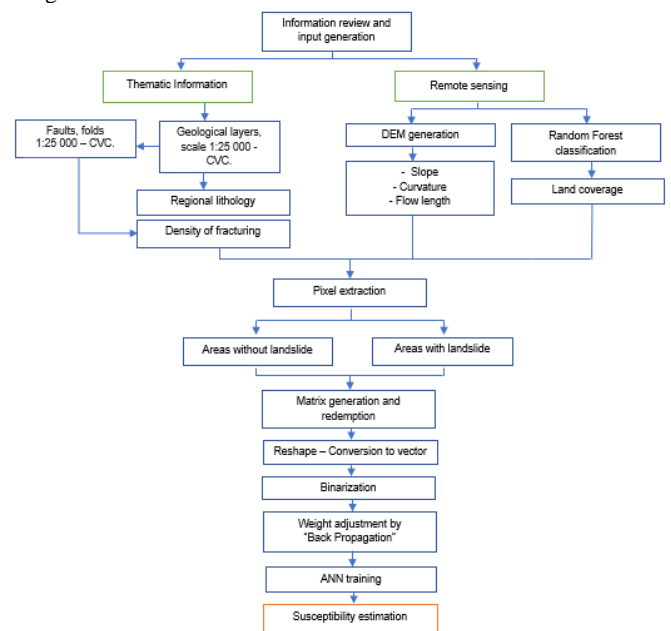


Figure 1. Methodological scheme.

For the land slope calculation was applied (1) (ArcGis, 2019).

$$\gamma = \tan^{-1} \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2} \quad (1)$$

Where: γ is slope in degrees; dz/dx is the surface change rate in horizontal direction; dz/dy is the surface change rate in vertical direction.

Its value was measured in degrees from 0° to 90° and is closely related to the mass movement's appearance, since it is the main geometric factor that appears in the stability analyses.

The curvature corresponds to the value of the second derivative of the input surface cell by cell, as indicated in (2) (ArcGis, 2019).

$$Cv = f''(IS) \quad (2)$$

Where: Cv is curvature; $f''(IS)$ is the second derivative of the input surface.

It is composed by two aspects: concavity and convexity of the land. The value 0 indicates that the surface is flat, positive values that the surface is convex towards the cell and the negative ones indicate concavity towards the cell.

Flow length was generated by calculating the direction and flow accumulation. The format used was “raster”, which means that pixel by pixel was operated taking into account the existing length upstream of the cell. The flow direction was determined by the direction of the steepest descent, or maximum fall, from each cell. It was calculated by applying (3) (ArcGIS, 2019).

$$md = \frac{dz}{x} * 100 \quad (3)$$

Where, md is maximum drop: dz change in height value; x is the distance calculated among cell centers.

The flow length, meanwhile, corresponds to a global function responsible for calculating the length of the longest flow path within a given basin, using equation (4) (ArcGis, 2019).

$$Fl = f(fdr, dm, wr) \quad (4)$$

Where, Fl is flow length; fdr is flow direction raster; dm is direction measurement along the flow path; wr is weight raster.

In the weighting of rocks quality, texture attributes / factory composition, age, origin and mechanical resistance are considered as qualifying attributes and / or grouping factors. Based on the stratigraphic chrono units, igneous rocks are classified according to their texture / factory and strength attributes. Sedimentary rocks are classified according to typical resistance ranges, and metamorphic rocks according to texture and mechanical strength (IDEAM, 2012). Table 1 shows the resistance ranges proposed by Hooke (2000).

Table 1. Resistance ranges

Resistance ranges (Mpa)	Grade	Score	Category
Extremely hard >250	R6	1	Very low
Very hard 100-250	R5	1	Very low
Hard 50-100	R4	2	Low
Moderately Hard 25-50	R3	3	Half
Soft 5-25	R2	4	High
Very soft	R1	5	Very high
Extremely soft	R0	5	Very high

To generate the fracture density map at 1:25 000 scale, with a pixel size of 30 m, a buffer or affectation radius of 500 m was defined, based on the structural geology layer referring to folding and faults of the municipality (5).

$$D = \frac{(L_1 * V_1) + (L_2 * V_2)}{A_c} \quad (5)$$

Where, D is fracture density; L_i is the length of fault lines; V_i is weight of tectonic activity; A_c is the circle area.

The procedure consists in calculating the density of lines in the vicinity of each pixel, which is defined by a search radius R , the length of the fault lines that fall within this circle is measured and multiplied by weight tectonics activity. Finally, the total sum of the length of the fault lines is divided by the circle's area, as shown in Figure 2. In cases where some value of V_n is not qualified, the length measured in the circle will be taken. (SGC, 2013).

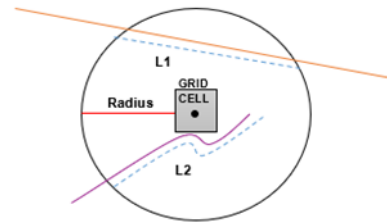


Figure 2. Procedure to calculate fracture density

Land coverage: To obtain coverage types, Landsat 8 satellite images were used for the years of study. The Python programming language was used for applying the Random Forest (RF) supervised classification method (Wang, Li, Jin, & Xiao, 2019). A total of 3000 regression trees were used to obtain the classification map for land usage and coverage.

For training the Random Forest classifier, 300 samples were used. Those corresponds to five categories of land coverage, according to the Colombian adaptation of the European methodology for CORINE Land Cover classification. An additional category corresponds to the presence of clouds in the scene. The classes obtained refer to lands dedicated to agricultural activities, forests and semi-natural areas, bare areas, urban areas and water surfaces. For each of these classes, 50 samples were obtained. The 85% of the data was used to feed the model, and the remaining 15% was kept it for validation of the adjusted models. Finally, the susceptibility classification was carried out with the ranges presented in Table 2.

Table 2. Susceptibility categories according to coverage type

Coverage Level	Score.	Category
Urban zone	4	High
Transient crops	2	Low
Permanent crops	2	Low
Herbaceous and shrub vegetation	3	Moderate
Weed grasses	3	Moderate
Clean pastures	4	High
Gallery and Ripario Forest	1	Very low
Dense forest	1	Very low
Arbustal and dense thicket	1	Very low
Bare and degraded lands	4	High
Inland waters	1	Very low

3.3. Artificial neural network

For the generation of the ANN, the layers for the study variables were loaded and from these, the polygons that contains areas with presence and absence of landslides are carried out by using Sentinel and Google Earth Pro images. The ANN training was performed selecting 800 random data, 400 with presence and 400 with absence of landslides. For extracting the pixel value of each variable, an algorithm was made in Python.

With the delimitation of the zones, a matrix was defined based on: updating matrix size for each variable, conversion to a vector format by means of a “reshape” and binarization process. A column of 1's and 0's was added to areas with presence and absence of landslides, respectively. The six columns of the variables and the additional one of the binary classification was taken to create a single matrix of seven columns.

The ANN model is created with six (6) input neurons, one (1) hidden layer with a total of 13 neurons and two (2) output layers. The “tanh” activation function was used in the input layer and “ReLU” in the hidden layer. For the training body, the first six columns of the matrix were used as the base, and the seventh column with the test. The weights of the variables are adjusted through the “BackPropagation” method using a total of 15% of the data for validation.

Finally, susceptibility estimation was carried out through the ANN training model performed and the complete images of each variable, in matrix format (Figure 3).

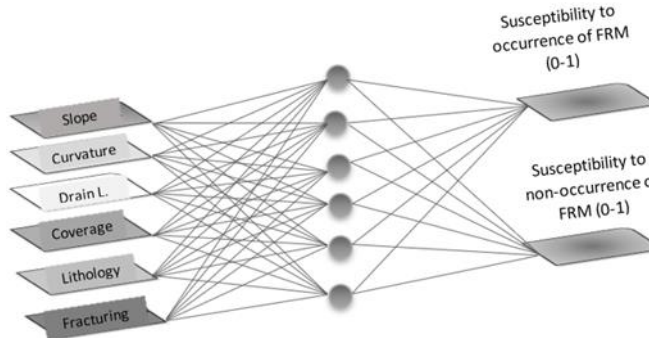


Figure 3. Artificial neural network architecture

4. RESULTS

4.1. Random forest classification

In this project, the Random Forest (RF) method was applied to classify Landsat 8 images. Accuracy of 80%, 90% and 100% was obtained for coverage identification. The confusion matrix (Figure 4) shows that in four of the six categories, the assignment was successful and in two categories there was a percentage equal to or greater than 80%. The cloud class is not taken into account in susceptibility because it does not correspond to values in the field.

RF confusion matrix

Predicted class	Observed class					
	1	2	3	4	5	6
1	100%	0%	0%	0%	0%	0%
2	0%	100%	20%	0%	0%	0%
3	0%	0%	80%	0%	0%	0%
4	0%	0%	0%	90%	0%	0%
5	0%	0%	0%	10%	100%	0%
6	0%	0%	0%	0%	0%	100%

Figure 4. RF confusion matrix: (1) artificial territories, (2) agricultural, (3) forests, (4) bare areas, (5) water surfaces and (6) clouds.

The Digital Elevation Models (DEM) was developed by using the Synthetic Aperture Radar Interferometry, Sentinel 1 images, and SNAP software. A pixel size of approximately 14 m was obtained and an accuracy error of approximately 3 m.

4.2. Susceptibility estimation (2018, second period)

Susceptibility estimation was obtained through multivariate statistics. An artificial neuronal network was used with six (6) input neurons, one (1) hidden layer with a total of 13 neurons and two output layers; the variables slope, curvature, flow length, coverage, lithology and fracturing were used. A training accuracy of 97% was obtained.

The validation of the neural network was performed using the backpropagation method with the 15% of the data that were not included in the model. A network accuracy of 97.26% was obtained. Figure 5 shows the susceptibility model for Sevilla township, Valle del Cauca, obtained through the artificial neural network.

Very low susceptibility: It is found in the majority of the municipality, approximately 74.51% (corresponding to 433.386 ha). In these areas, the network did not find many pixel values corresponding to susceptible lands according to the sum of synoptic weights in the hidden layer. This type of susceptibility corresponds to terrain with flat and abrupt topography, however, the characteristics of the soil and its geology do not generate instability.

Low susceptibility: Belonging to 10.06% of the municipality with a total of 58.480 ha. It is found in areas near fault lines and unstable geology, however, the slopes correspond to values between 0° to 12°.

Moderate susceptibility: Present in 8.25% of the municipality with 47.997 ha. It appears in areas adjacent to fault lines, steep slopes, but with a high level of drainage, therefore, susceptibility is maintained at this level.

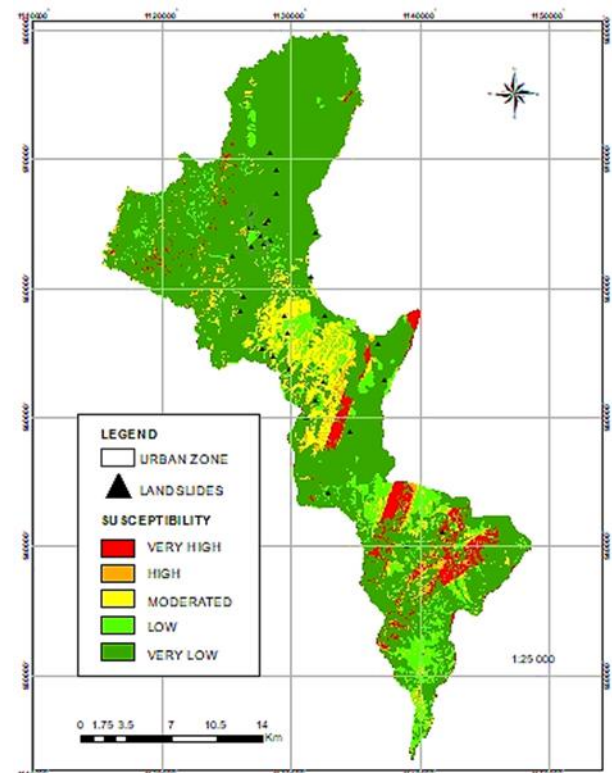


Figure 5. Susceptibility 2018-2 by RNA

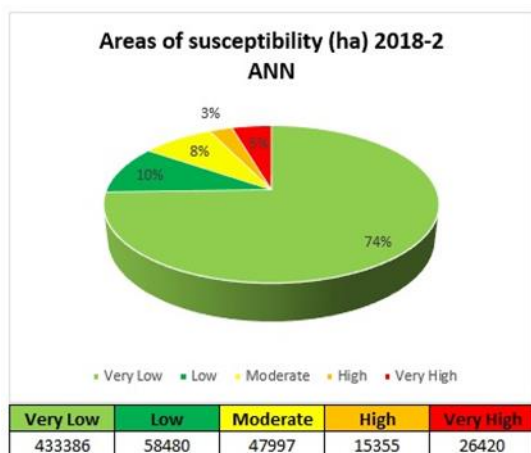


Figure 6. RNA susceptibility level

High susceptibility: It is located in 2.64% of the municipality, corresponding to 15.355 ha. In these regions, mountain ridges with high characteristics of each of the variables involved (soil cover, geology, curvature, fault density, slope and flow length) are taken.

Very high susceptibility: Present in 4.54% of the territory of Sevilla, Valle del Cauca, with a total of 26.420 ha; a hierarchy is evident in the variables of fracture density and geology, since it belongs to the pixels located in the course of the faults present in the sector.

5. CONCLUSIONS

The municipality of Sevilla - Valle del Cauca, is located on steep topography, low stability soils in terms of its physical and chemical composition, and close to the interception of fault lines, which creates ideal scenarios for landslides.

Artificial Neural Networks have a number of advantages over the different methods used to estimate susceptibility. One of them is based on the use of the “Backpropagation” technique that facilitated the procedure to reach total susceptibility, because for its implementation and structure only a hidden layer with a low number of neurons is necessary (13 in this case), the input layer (6 neurons) and finally, the output layer (2 neurons). With the above, there is another advantage, which focuses on the processing capacity, since a large computational capacity is not required for its realization. Finally, unlike deterministic methods, it does not require a set of experts who must assess each of the variables to be used.

Performing the prediction of susceptibility to landslides through the implementation of Artificial Neural Networks (ANN) is more optimal than when performed with SVM techniques because they can have variations and a greater number of classifications who can be used to assess a more detailed terrain qualities while with SVM it could only have a dichotomous result.

According to the susceptibility map obtained, it is evident that the geology is decisive in the stability of the terrain, since, where a fault line is located, fault interceptions or changes in the composition of the terrain, the susceptibility of the terrain grows progressively.

It is possible to obtain a more accurate result in the calculation of susceptibility by means of ANN, using continuous variables, instead of discrete ones, because the continuous

variables are extracted directly from the DEM and therefore contain the own values of the variable, while that the continuous ones are based on the heuristic qualification, being very prone to variations according to the study area.

Studies of this nature are useful for risk management entities and the Basic Land Management Plan (PBOT) as it allows to know the sites suitable for urban distribution and expansion.

6. REFERENCES

- Chacón, J., Irigaray, C., Fernández, T., & El Hamdouni, R. (2006). Engineering geology maps: Landslides and geographical information systems. *Bulletin of Engineering Geology and the Environment*, 65(4), 341–411. <https://doi.org/10.1007/s10064-006-0064-z>
- CVC & OSSO. (2011). Estudio de amenaza, vulnerabilidad, escenarios de riesgo y diseño de ingeniería y tratamientos de mitigación ante movimientos en masa en los barrios monserate y cafetero.
- Dai, F. ., Lee, C. ., & Ngai, Y. . (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1), 65–87. [https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X)
- El País. (2013). En Sevilla, Valle, las fallas geológicas se tragan vías y casas. Retrieved October 31, 2019, from <https://www.elpais.com.co/valle/en-sevilla-las-fallas-geologicas-se-tragan-vias-y-casas.html>
- Geymen, A. (2014). Digital elevation model (DEM) generation using the SAR interferometry technique. *Arabian Journal of Geosciences*, 7(2), 827–837. <https://doi.org/10.1007/s12517-012-0811-3>
- Guzzetti, F., Carrara, A., & Paola, R. (1999). Use of GIS Technology in the Prediction and Monitoring of Landslide Hazard. *Natural Hazards*, 20(November), 117–135. <https://doi.org/10.1023/A:1008097111310>
- IDEAM. (2012). Metodología Para La Zonificación De Susceptibilidad General Del Terreno a Los Movimientos En Masa. Bogotá, 1(20), 3527110–3527160. Retrieved from www.ideam.gov.co
- Ingeominas. (2002). Catálogo Nacional de Movimientos en Masa, 289. <https://doi.org/10.1016/j.jbankfin.2011.01.023>
- Kawabata, D., & Bandibas, J. (2009). Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology*, 113(1–2), 97–109. <https://doi.org/10.1016/j.geomorph.2009.06.006>
- Pepe, A., & Calò, F. (2017). A review of interferometric synthetic aperture RADAR (InSAR) multi-track approaches for the retrieval of Earth’s Surface displacements. *Applied Sciences* (Switzerland), 7(12). <https://doi.org/10.3390/app7121264>
- Pradhan, B., & Lee, S. (2010a). Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environmental Earth Sciences*, 60(5), 1037–1054. <https://doi.org/10.1007/s12665-009-0245-8>

Pradhan, B., & Lee, S. (2010b). Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides*, 7(1), 13–30. <https://doi.org/10.1007/s10346-009-0183-2>

Parise, M. (2001). Landslide mapping techniques and their use in the assessment of the landslide hazard. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 26(9), 697–703. [https://doi.org/10.1016/S1464-1917\(01\)00069-1](https://doi.org/10.1016/S1464-1917(01)00069-1)

SGC. (2013). Servicio Geologico Colombiano República De Colombia Ministerio De Minas Y Energía Servicio Geologico Colombiano Subdirección De Amenazas Geológicas Y Entorno Ambiental, (2).

Süzen, M. L., & Doyuran, V. (2004a). A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environmental Geology*, 45(5), 665–679. <https://doi.org/10.1007/s00254-003-0917-8>

Süzen, M. L., & Doyuran, V. (2004b). Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchment, Turkey. *Engineering Geology*, 71(3–4), 303–321. [https://doi.org/10.1016/S0013-7952\(03\)00143-1](https://doi.org/10.1016/S0013-7952(03)00143-1)

Van Westen, C. J. (1993). Application of Geographical Information System to landslide zonation. ITC Publication.

Wachal, D. J., & Hudak, P. F. (2013). in Travis County , Texas , USA Mapping landslide susceptibility. *GeoJournal*, 51(3), 245–253. <https://doi.org/10.2307/41147516>

Wang, B., Li, J., Jin, X., & Xiao, H. (2019). Mapping Tea Plantations from Multi-seasonal Landsat-8 OLI Imageries Using a Random Forest Classifier. *Journal of the Indian Society of Remote Sensing*, 47(8), 1315–1329. <https://doi.org/10.1007/s12524-019-01014-5>