

Mapa de susceptibilidad por movimiento en masa en la localidad Rafael Uribe Uribe de Bogotá Colombia, mediante Máquinas de Vectores de Soporte

Landslide Susceptibility Map in the locality “Rafael Uribe Uribe” in Bogotá, Colombia, by using Support Vector Machine

DOI: <https://dx.doi.org/10.17981/ingecuc.21.2.2025.11>

Artículo de Investigación Científica.
Fecha de Recepción: 14/04/2024, Fecha de Aceptación: 02/09/2025

Miguel Alfonso Solano-Padilla 

Universidad Nacional Autónoma de México, (México), México
miguel.solano.padilla@comunidad.unam.mx

Carlos Arturo Peña-Rincón 

Universidad Sergio Arboleda, Bogotá, Colombia
carlos.pena@usa.edu.co

To cite this paper

M. Solano-Padilla & C. Peña-Rincón “Landslide Susceptibility Map in the locality “Rafael Uribe Uribe” in Bogotá, Colombia, by using Support Vector Machine,” *INGE CUC*, vol. 21, no. 2, 2025. DOI: <https://dx.doi.org/10.17981/ingecuc.21.2.2025.11>

Resumen

Introducción: La susceptibilidad a la ocurrencia de deslizamientos fue reconocida como un factor clave para contribuir a la elaboración de escenarios de riesgo y al fortalecimiento del conocimiento sobre el riesgo de desastres, lo que respaldó la gestión del uso del suelo en áreas urbanas. Un reto de las metodologías es contribuir al fortalecimiento de los sistemas de alerta temprana.

Objetivo: Se integraron los datos de las variables condicionantes para producir el mapa de susceptibilidad a deslizamientos en la localidad de ‘Rafael Uribe Uribe’, en la ciudad de Bogotá, mediante la aplicación de la técnica de Máquinas de Vectores de Soporte (SVM).

Metodología: Se utilizó un registro histórico de 430 eventos ocurridos entre 2008 y 2015, junto con variables relacionadas con la topografía, uso del suelo, distancia a vías, geología y precipitación del área de estudio. Con esta información se construyó una base de datos con 12 variables condicionantes. Los datos fueron divididos aleatoriamente: el 75% se empleó para generar el modelo mediante la técnica de Máquinas de Vectores de Soporte (SVM) con validación cruzada K-fold y búsqueda en malla (grid search), mientras que el restante 25 % se destinó a la validación del modelo.

Resultados: Se obtuvo el mapa de susceptibilidad de la localidad con un nivel de precisión del 98 % utilizando el núcleo RBF y del 97 % para el núcleo lineal, validado mediante la curva Característica Operativa del Receptor (ROC). La técnica de Máquinas de Vectores de Soporte (SVM) fue aplicada exitosamente para clasificar las zonas del área de estudio en categorías de estabilidad e inestabilidad.

Conclusiones: Los resultados respaldaron la aplicabilidad del Aprendizaje de Máquina en el análisis de susceptibilidad, en concordancia con los avances promovidos por la comunidad científica internacional.

Palabras clave: Aprendizaje de Máquina, Máquinas de Vectores de Soporte, Sistemas de Información Geográfica (SIG), mapa de susceptibilidad, variables condicionantes; Núcleo.

Abstract

Introduction: The susceptibility to landslide occurrence was recognized as a key factor for contributing to risk scenario development and for enhancing disaster risk knowledge, which supported land-use management in urban areas. A challenge of methodologies is to contribute to strengthening early warning systems.

Objective: Data from conditioning variables were integrated to produce the landslide susceptibility map for the locality of ‘Rafael Uribe Uribe’ in Bogotá, through the application of the Support Vector Machine (SVM) technique.

Methodology: A historical record of 430 events from 2008 to 2015 was used, along with variables related to topography, land use, distance to roads, geology, and precipitation in the study area. Based on this information, a database of 12 conditioning variables was constructed. The data were randomly divided: 75% were used to generate the model through the Support Vector Machine (SVM) technique with K-fold cross-validation and a grid search, while the remaining 25 % were used for model validation.

Results: The susceptibility map of the locality was obtained with an accuracy of 98 % using the RBF kernel and 97 % with the linear kernel, validated through the Receiver Operating Characteristic (ROC) curve. The Support Vector Machine (SVM) technique was successfully applied to classify the study area into stable and unstable zones.

Conclusions: The results supported the applicability of Machine Learning for susceptibility analysis, in alignment with the approaches proposed by the scientific community.

Keywords: Machine Learning, Support Vector Machines; Geographic Information Systems (GIS); susceptibility map; conditioning variables, Kernel.



INTRODUCTION

Landslides are natural phenomena characterized by the displacement of soil, debris, or rock under the influence of gravity [1], [2]. These events occur more frequently in mountainous or high-slope areas, where the geomorphological and climatic conditions favor terrain instability. Its occurrence generates significant social and economic impacts, including human casualties and damage to road corridors, urban settlements, and industrial infrastructure [3]. According to global records, nineteen triggering factors were identified as responsible for 4862 landslides that occurred between 2004 and 2016, excluding events associated with seismic activity. Analyses indicate that precipitation represented the predominant causal factor, intervening in approximately 79% of the documented cases [4]. Similarly, the database called Global Landslide Database (GLS-DAT) reported human casualties from landslides in 128 countries, including Colombia, during the period from 1995 to 2014, with data indicating an average annual rate of 8877 human losses, which evinces the global magnitude of their impact on the population [5].

Precipitation is considered an important factor among the 19 reported factors that contribute to the dynamics of landslides. Additionally, climate change in Colombia and the “La Niña” (*The Girl*) phenomenon are incorporated into disaster risk scenarios, as they contribute to the increase in precipitation [6], [7].

In the city of Bogotá, various types of hazards and disaster risks have been identified, notably floods, seismic activity, and landslides [8]. Therefore, it is important to monitor those areas susceptible to landslides through methodologies that integrate data made available by suitable entities such as the District Institute for Risk Management and Climate Change (IDEGER, by its Spanish initials), the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM, by its Spanish initials), and the Colombian Geological Survey (SGC, by its Spanish initials), among other institutions. This allows improving the early warning systems and disaster risk management, for supporting decision-making in the urban development planning of the city in aspects such as transportation, land-use regulation, population well-being, and the coherence of the urban territory [9].

Currently, the IDEGER has the Bogotá Alert System (SAB) which uses hydro-meteorological information obtained from a network of meteorological stations. Based on the data log, the system reports by using an alert level map that classifies the hazard into three categories. Nevertheless, it has been evidenced that the incorporation of statistical and machine learning methodologies that integrate spatial information, such as Support Vector Machines (SVM), allows for improved determination of landslide susceptibility and strengthens the predictive capability of the alert system [10].

Landslide susceptibility analysis is the probability of landslide occurrence in an area, which depends on local terrain conditions such as its geomorphology, land cover, land use, and geology, and together with triggering factors like precipitation and seismic activity that materialize this hazard [11], [12]. Different methodologies have been developed for its evaluation, as described in [13], [14]. In [15], it is proposed that all methodologies can be grouped into 6 categories, as follows: (i) Classical statistics (logistic regression, discriminant analysis, linear regression), (ii) Index-based methods (heuristic analysis), (iii) Machine learning (fuzzy logic system, SVM, decision trees), (iv) Neural networks, (v) Multi-criteria decision analysis, and (vi) Other statistics.

Statistical learning refers to a set of tools for understanding data; these tools can be classified as supervised and unsupervised [16]. The SVM method constitutes a supervised technique widely used to predict the occurrence of landslides based on the analysis of spatial distribution patterns. Various studies have demonstrated that this approach achieves precision levels between 75% and 95% in the prediction of areas susceptible to such events [10]. Specifically, SVM considers different kernels [17], which have been applied to the study of large mountainous areas considering databases with information on topography, geology, soils, forests, and land use, by using Geographic Information Systems (GIS) to generate the landslide susceptibility map [18], [19], [20].

Other studies have included a landslide inventory map applied to hydrological basins [21], [22], [23], in urban areas with the precipitation triggering factor [24]. Additionally, SVM has been used with other machine learning techniques [25], and SVM has been compared with

decision trees and neuro-fuzzy models, concluding that the performance of these results for the study area is satisfactory [26]. Furthermore, SVM has gained significant importance and interest within the scientific community, particularly from the International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE) [27].

In Colombia, the city of Bogotá, located in the Eastern mountain range, contributed 25.3% to the country's Gross Domestic Product (GDP) in 2021 [28]. However, this participation may be compromised in the future due to the occurrence of landslides in human settlements near quarry exploitation zones, on landfills or slopes without appropriate technical intervention, as well as riparian zones. In Bogotá, 4169 landslides were recorded between 2002 and 2012, registered in the localities of Usaquen, Chapinero, Santa Fe, San Cristobal, Rafael Uribe Uribe, Usme, Ciudad Bolívar, and Suba [29], [30], [31]. This paper opted for the Rafael Uribe Uribe locality as the study area to determine the landslide susceptibility map using the SVM technique.

MATERIAL AND METHODOLOGY

Study Area

The urban administration of Bogotá is composed of 20 localities, among which is Rafael Uribe Uribe (Fig 1), with the following Zonal Planning Units: *San Jose, Quiroga, Marco Fidel Suarez, Marruecos, Diana Turbay, and Parque entre Nubes*, which present a landslide hazard.

From the disaster risk management, reduction measures have been implemented in the locality, but landslide events persist, mainly during periods of precipitation. Additionally, the unplanned urbanization process as well as the activities intrinsic to the community in the locality have led to progressive environmental deterioration, which has altered the stability conditions of the hillsides, conditioning the occurrence of landslides [31], [32].

Locality 18 has a population of 353,761 inhabitants (4.43% of the total population of Bogotá). It covers a total area of 1,383.4 hectares, comprising 115 neighborhoods containing 1,412 blocks, 191 of which show urban development without any planning in high-risk zones regarding the landslide hazard, placing them in high-hazard zones, configured by informal settlements. On the other hand, Locality 18 is the third locality in Bogotá with the highest population exposed to landslides for the year 2017 (26,638 people), preceded by the localities of Usaquen with 40,948 people and Ciudad Bolívar with 49,295 [31], [32].

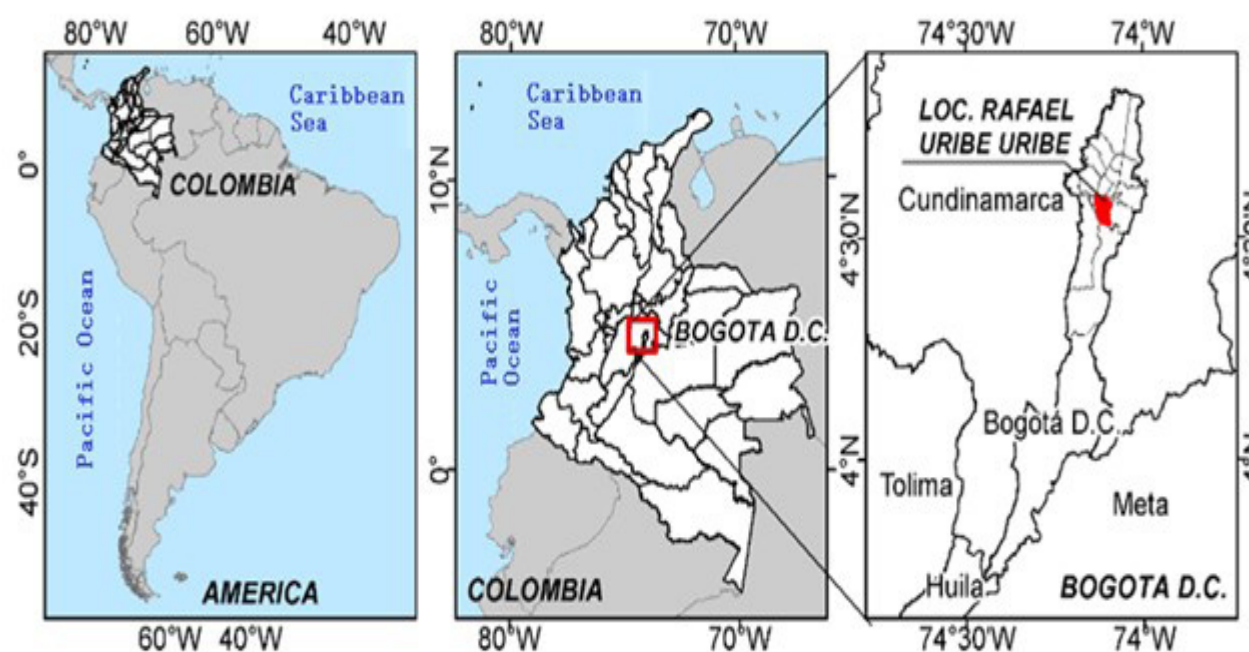


Fig 1. Location of the study area, Locality 18.
Source: Authors.

Data

The database comprises a record of 470 historical landslide events, categorized as soil slides, mudflows, and rockfalls. These events, documented in technical sheets that describe the location and characteristics of each event for the period 2005-2008, were collected by the IDIGER, the local mayor's office, and the fire department [31]. A subsequent analysis of the information showed a strong correlation between the occurrence of these events and precipitation records [31].

The evaluation of landslide susceptibility is based on the data-driven method. Therefore, data curation was performed, and initially, 452 events were successfully georeferenced by using Google Earth Pro. The data are grouped into three categories:

1. Found Data: Event georeferenced by using Google Earth Pro, where the location coincided with the urban nomenclature information recorded in the technical sheet. Total: 247.
2. Approximated Data: Event whose location was approximated by using Google Earth Pro, based on the urban nomenclature information recorded in the technical sheet. Total: 205.
3. Not Found Data: Event whose reported location in the technical sheet did not allow for a precise determination of its urban nomenclature. Total: 18.

Based on the 452 georeferenced events, the conditional variable of historical events representing the Rafael Uribe Uribe locality in the 2008–2015 time period was generated. Of the 452 events recorded in this time period, 430 landslide events were identified as georeferenced within the area of Locality 18. These events were identified as unstable zones, that is to say, places where landslides have historically occurred.

On the other hand, by using the landslide hazard map at a 1:5,000 scale in [31], [32], and by defining a criterion of hillsides with a gradient of less than 2 degrees, areas classified with low landslide susceptibility were determined. These areas were denominated as stable zones, and a random sample composed of 430 locations was obtained. Finally, the stable and unstable zones were distributed as shown in Fig 2.

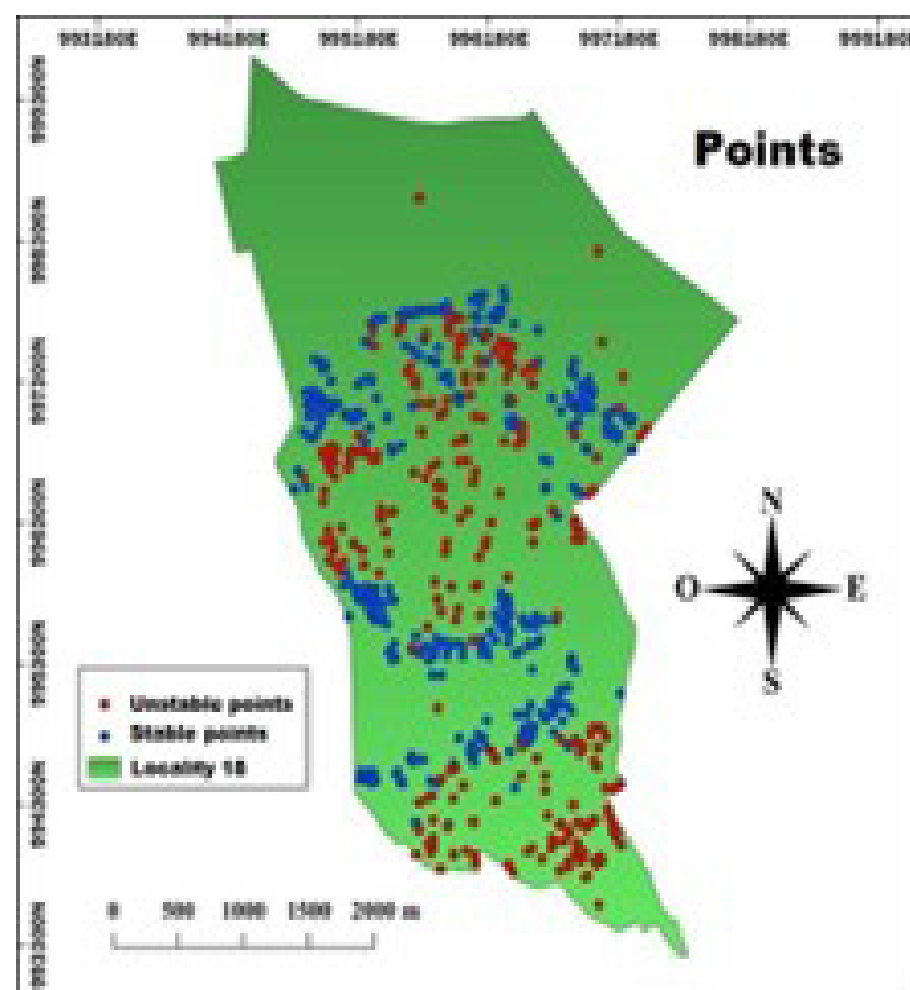


Fig 2. The points represent stable and unstable zones in the locality Rafael Uribe Uribe. Source: Authors.

In this study, the following conditioning variables were considered: Digital Elevation Model (DEM), hillside gradient, slope aspect, plan and curvature profile, distance to drainages, Topographic Wetness Index (TWI), soil loss factor (PS, by its Spanish initials), distance to roads, distance to geological faults, geology, and annual precipitation map. These variables have a direct or indirect relationship with the evaluation of landslide susceptibility.

The DEM was obtained from the contour line information of Locality 18, by using the Inverse Distance Weighting (IDW) interpolation methodology. Subsequently, local depressions were eliminated by means of the Fill Sinks Algorithm [33], resulting in the DEM variable with a resolution of 14 m. Fig 3.

The slope variable was obtained from the DEM by means of the Slope algorithm in ArcGIS [34] and reclassified into the following seven categories: (1) 0–2°, (2) 2–4°, (3) 4–7°, (4) 7–14°, (5) 14–27°, (6) 27–37°, (7) >37°, See Fig 3.

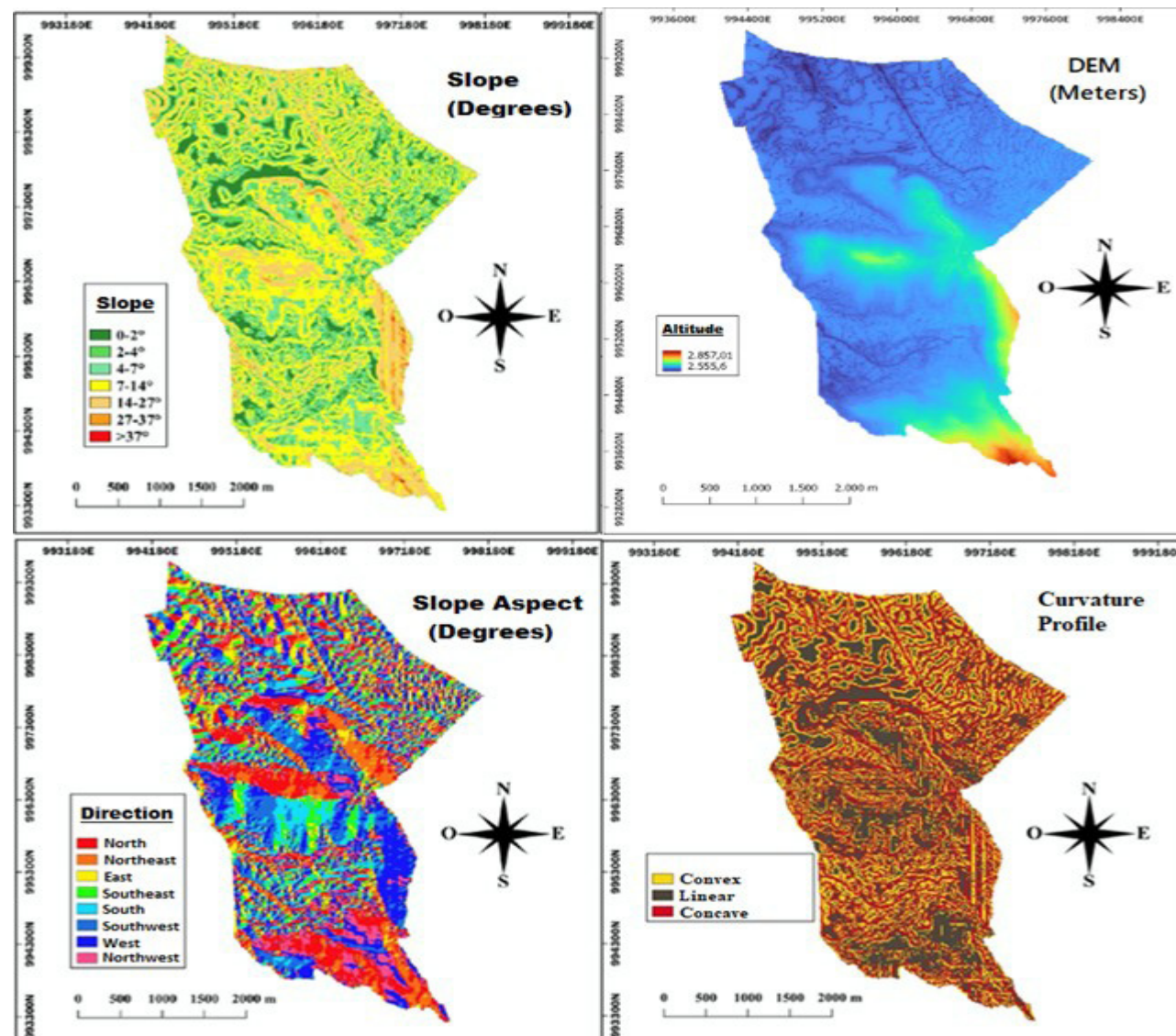


Fig 3. Digital Elevation Model, Slope, Slope Aspect, and Curvature Profile
Source: Authors.

Additionally, information derived from three morphometric variables was integrated into Figure 3: slope aspect, curvature (profile and plan), and distance to the drainage network.

The slope aspect, which identifies the direction toward which the terrain surface is oriented, was generated by using the Aspect algorithm in the ArcGIS software. Subsequently, the resulting variable was reclassified into nine categories: Flat (0) < 0°, North (1) 337.5–360° or 0–22.5°, Northeast (2) 22.5–67.5°, East (3) 67.5–112.5°, Southeast (4) 112.5–157.5°, South (5) 157.5–202.5°, Southwest (6) 202.5–247.5°, West (7) 247.5–292.5° and Northwest (8) 292.5–337.5°.

The terrain curvature (profile and plan) was obtained from the Digital Elevation Model (DEM) by using the Curvature algorithm, which considers the concave, convex, or flat shape of the surface, as well as the convergence or divergence of flow [34]. The results were reclassified into three categories: (-1) < 0 (concave), (0) = 0 (flat), and (1) > 0 (convex), as shown in Fig. 3 and Fig 4.

Finally, the distance to the drainage network was calculated by using the official hydrographic variable, information supplied by IDEAM, by using the Euclidean Distance Algorithm [34]. This procedure generated raster information where each pixel is a representation of the minimum distance from its center to the nearest water channel, as shown in Fig 4.

Landslide susceptibility map in the locality “Rafael Uribe Uribe” in Bogotá, Colombia, by using support vector machine

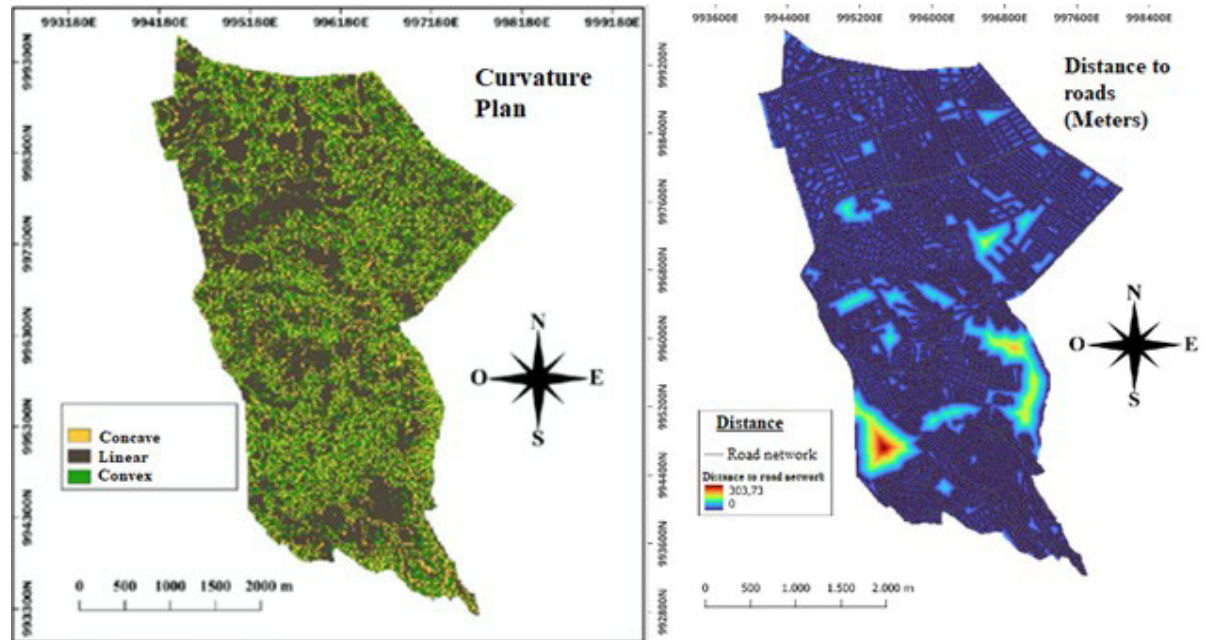


Fig 4. Vector variable of the curvature plan and distance to the drainage network
Source: Authors

In Fig 5., the following information was considered: Topographic Wetness Index (TWI): This is a topographic humidity index used to describe the degree of water accumulation for an area, (1).

$$TWI = \ln\left(\frac{A_s}{\tan(\beta)}\right) \quad (1)$$

Where A_s is the contributing area and β is the slope angle in radians. [35], [36]. The contributing area variable represents the accumulated flow, which is the accumulated weight of all cells that flow towards each cell with a descending slope in the output raster [37].

PS Factor: The PS parameter is an erosion-related parameter that represents the soil mass loss per unit area. This is calculated as: $PS = \frac{A_s}{\ln(\beta)}$, where A_s is the contributing area and β is the slope angle in radians.

Distance to Roads: This variable was obtained from the road network information by means of the Euclidean Distance Algorithm from [34].

Distance to Faults: This variable was derived from the geological faults variable (information requested from the Colombian Geological Survey), and by means of the Euclidean Distance Algorithm from [34].

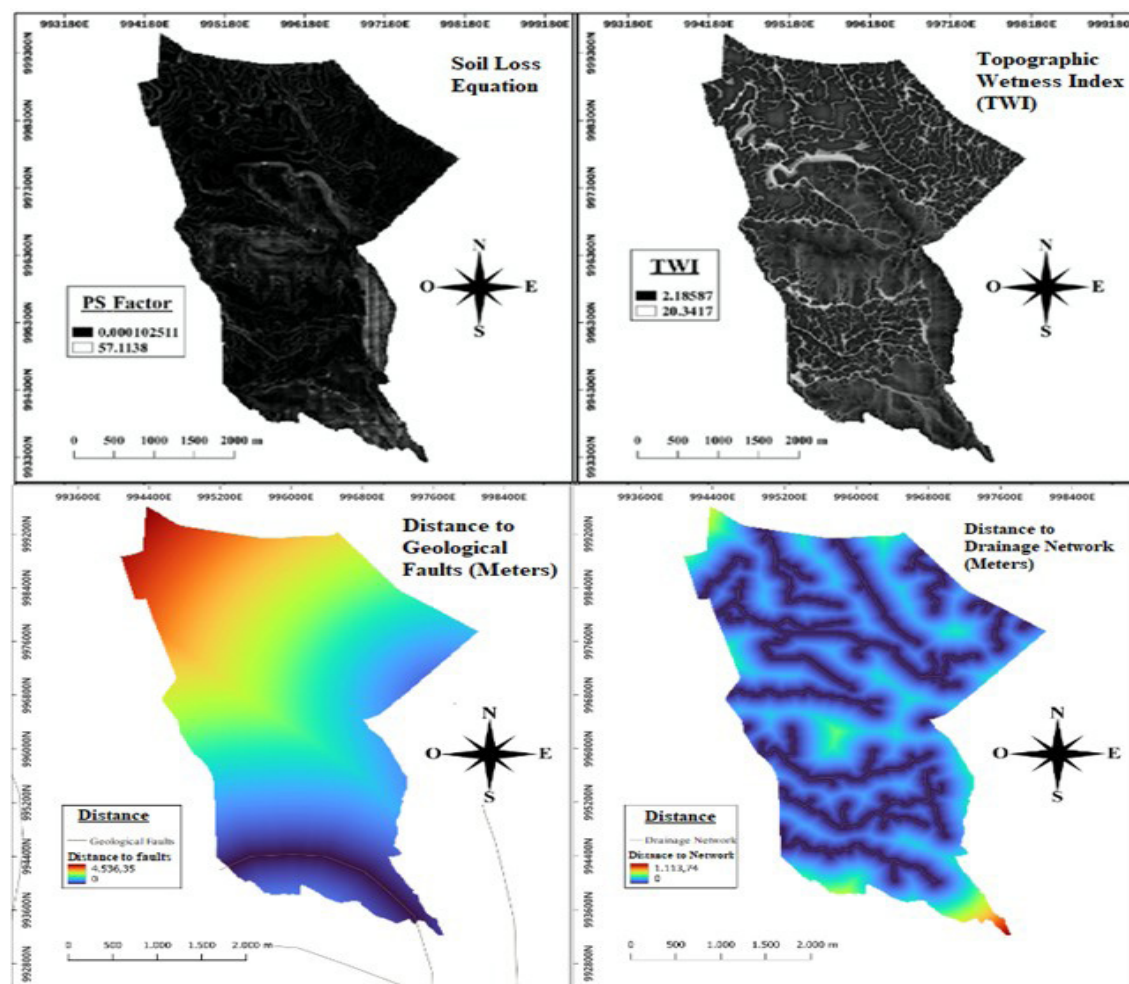


Fig 5: The vector-type information constructed for soil loss, TWI, distance to geological faults, and the drainage network in the locality Rafael Uribe Uribe
Source: Authors.

The precipitation variable (Fig 6) was prepared based on records obtained from meteorological stations located in the surroundings of Locality 18, information supplied by the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM). The data corresponds to the annual precipitation averages recorded during the 2008–2015 period. The spatial interpolation of the scattered data was performed by using the multi-level B-Spline interpolation module [37], with the purpose of generating a continuous surface of precipitation distribution. As a result, the following categories were established: (1) 0–50, (2) 50–100, (3) 100–150, (4) 150–200, (5) 200–300, (6) 300–400, (7) 400–600, (8) 600–800, (9) 800–1000 y (10) >1000 mm/year.

The geological variable (Fig 6) was constructed based on sheet 246 from the Colombian Geological Survey, with the purpose of representing the main lithological units of the study area. The following formations were identified: Pgbos (Bogota Formation), Pgr (Regadera Formation), Pgu (Usme Formation), Qal (Alluvial deposits), Qcc (Cone Complex), and Qtb (Terraces).

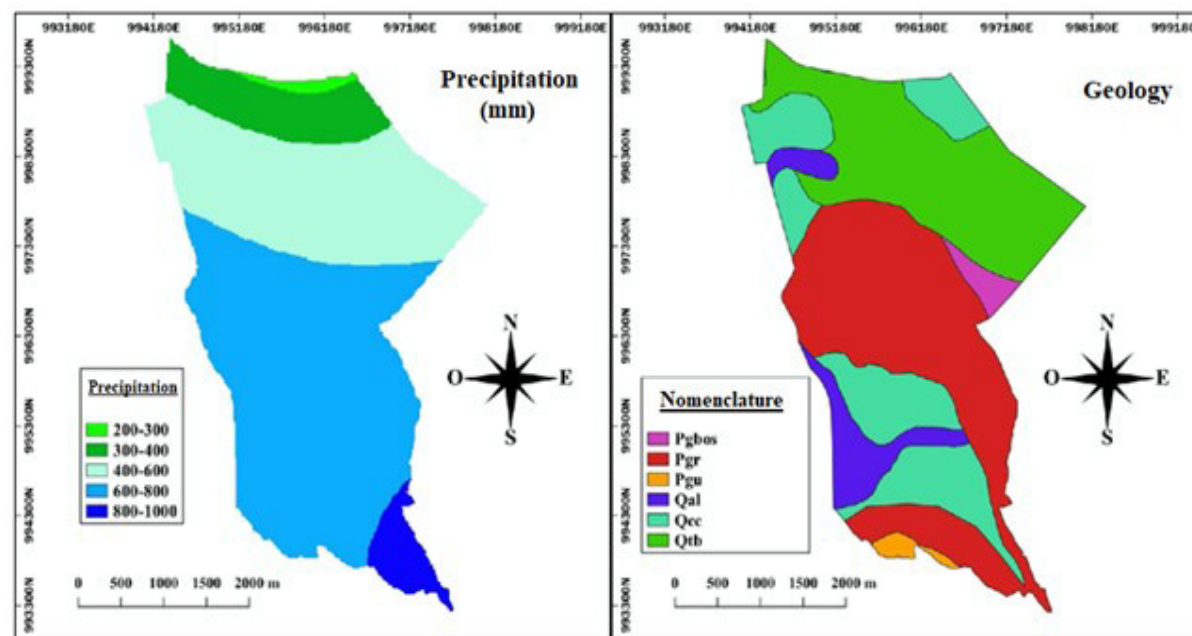


Fig 6. Vector variable of average annual precipitation and Geological variable in the locality Rafael Uribe Uribe. Source: Authors.

Finally, by using the point sampling tool [33], the DEM variable was transformed into a variable of sampling points while preserving its attributes. Then, the conditioning variables were intersected to consolidate a database composed of 860 zones in the study area, -including stable and unstable zones-, where each pixel contains information for each of the 12 conditioning variables. (Fig 7).

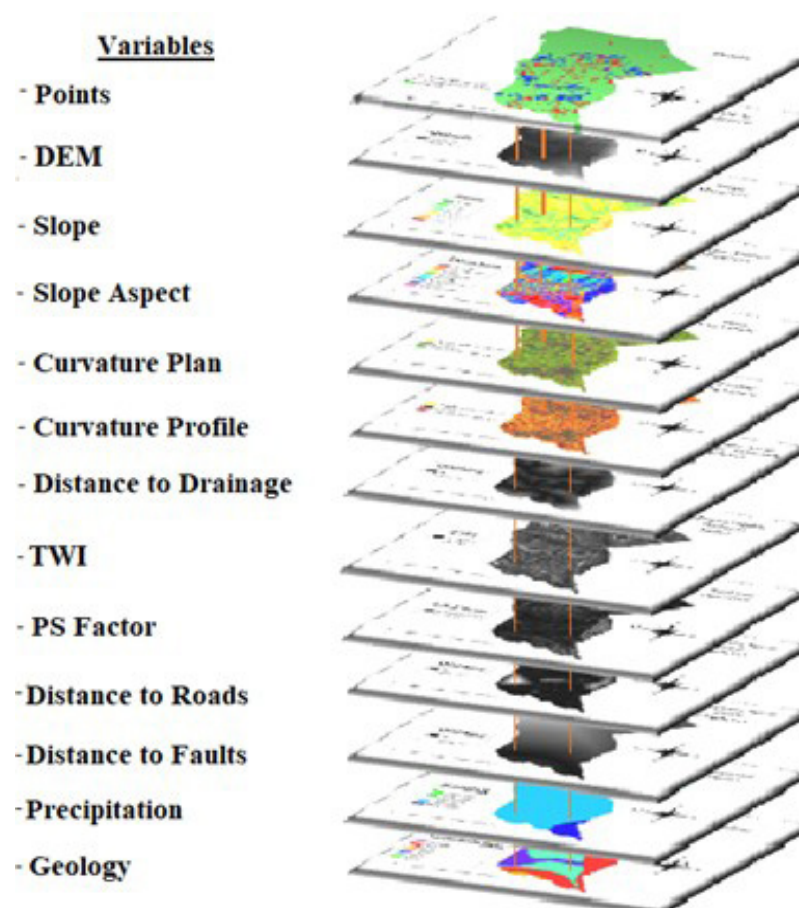


Fig 7. Grouping of the Conditioning Variables Information. Source: Authors.

Support Vector Machines

In 1995, Vladimir Vapnik, along with Corinna Cortes, published the book “Machine Learning” with a chapter called “Support-Vector Networks” in which the idea of a new learning machine for binary classification problems is introduced. The machine was implemented with the following idea: a series of training data, previously classified into two classes, which are non-linearly mapped to a higher-dimensional space. In this space, a linear decision surface called the decision function is constructed, which separates the data according to its class; this allows future, unclassified data to be classified based on this condition [38], [39]; however, different cases need to be considered.

SVM Kernel: Non-linear Case

In this case, the data are mapped into a higher-dimensional vector space called the feature space. (Fig 8).

Definition: Let $X = \{x_1, x_2, \dots, x_m\} \subset \mathbb{R}^n$ be an input space and let be the feature space. A feature mapping is defined, given in (2):

$$\blacksquare \begin{pmatrix} \phi : X \rightarrow F \\ x_i \mapsto \phi(x_i) \end{pmatrix} \quad (2)$$

Where $\phi(x_i)$ represents a vector in the feature space that contains the non-linear attributes [40],[41],[42].

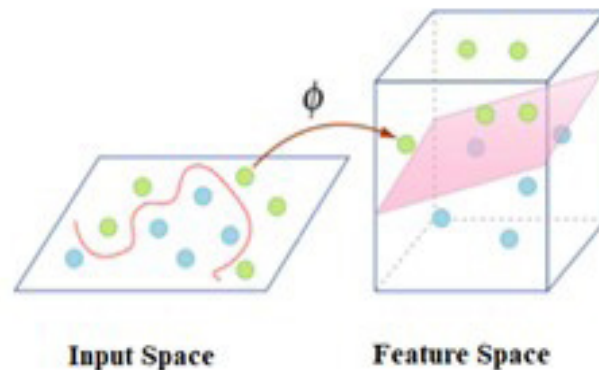


Fig 8. Feature Mapping from X to F [43].

Now, if we apply , the training set in the feature space (3) is obtained:

$$T_\phi = \{(\phi(x_1), y_1), (\phi(x_2), y_2), \dots, (\phi(x_m), y_m))\} \quad (3)$$

For which a linear classification function (T) is desired to be found by means of an optimization problem (4):

$$\blacksquare \begin{pmatrix} \text{Objective Function: } \min_{w', b', \xi_i} \left\{ \frac{\|w'\|^2}{2} + C \sum_{i=1}^m (\xi_i)^k \right\} \\ \text{Linear Constraints: } y_i * (\vec{\phi}(x_i) \cdot w' - b') \geq 1 - \xi_i, \forall (\phi(x_i), y_i) \in T \\ \xi_i \geq 0, i = 1, \dots, m \end{pmatrix} \quad (4)$$

Where is the weight vector, is the bias, and is the error. And along with the dual optimization problem in the feature space (5):

Objective Function:

$$\max_{\alpha} \left\{ L_{\text{dual}} = \sum_{i=1}^m \alpha_i - \frac{1}{2} * \sum_{i=1}^m \sum_{j=1}^m \alpha_i * \alpha_j * y_i * y_j * (\vec{\phi}(x_i) \cdot \vec{\phi}(x_j)) \right\} \quad \text{Linear constraints:} \\ 0 \leq \alpha_i \leq C, i = 1, \dots, m \text{ y } \sum_{i=1}^m \alpha_i * y_i = 0 \quad (5)$$

The dual optimization problem is the same problem as the non-separable case. However, this depends on the dot product of the vectors belonging to the feature space $(\vec{\phi}(x_i) \cdot \vec{\phi}(x_j))$,

Thus, we have the definition: let $X = \{x_1, x_2, \dots, x_m\} \subset \mathbb{R}^n$ be an input space. A kernel is a function $k: X \times X \rightarrow \mathbb{R}$ that satisfies (6):

$$k(x_i, x_j) = (\vec{\phi}(x_i) \cdot \vec{\phi}(x_j)), \forall x_i, x_j \in X \quad (6)$$

Where is a feature mapping from X to F.

Additionally, the function is symmetric because the dot product is symmetric $k(x_i, x_j) = k(x_j, x_i) \forall x_i, x_j \in X$, [43]. Among the kernels to be considered for this paper are:

Homogeneous Polynomial Kernel

Let $x, y \in \mathbb{R}^d$ be the homogeneous polynomial kernel (7).

$$k_q(x, y) = (\vec{\phi}(x) \cdot \vec{\phi}(y))^q = (x^T y)^q = (\vec{x} \cdot \vec{y})^q \quad (7)$$

Where q is the degree of the polynomial. This kernel corresponds to a feature space composed of all products of exactly q attributes; the most commonly used case is the linear kernel ($q=1$), [40].

Gaussian Kernel:

Let $x, y \in \mathbb{R}^d$ be the Gaussian kernel, also called the Gaussian Radial Basis Function (RBF); it is defined in (8):

$$k(x, y) = \exp\left\{-\gamma^* \|x - y\|^2\right\} = \exp\left\{-\frac{\|x - y\|^2}{2\sigma^2}\right\} \quad (8)$$

Where γ is a parameter greater than zero and $\sigma > 0$ is the spread parameter. Note that $k(x, x) = 1$, and also that the kernel value is inversely related to the distance between the points x and y . i.e, if $\|x - y\| \rightarrow 0$ $k(x, y) \rightarrow 1$ then $\|x - y\| \rightarrow \infty$ and if $k(x, y) \rightarrow 0$ then $\|x - y\| \rightarrow \infty$, [40].

RESULTS AND ANALYSIS

With the consolidated data obtained by grouping the 12 variables, denominated input data (Fig 7), the geology variable is found to have categorical values. For the SVM, it is important to consider numerical values; therefore, the geological map was converted to a numerical representation by using One-Hot encoding [44] (Fig 9), and the same was done for the slope aspect variable.

Geology	Pgbos	Pgr	Pgu	Qal	Qcc
Pgbos	1	0	0	0	0
Pgr	0	1	0	0	0
Pgu	0	0	1	0	0
Qal	0	0	0	1	0
Qcc	0	0	0	0	1
Qtb	0	0	0	0	0
Pgr	0	1	0	0	0
Qcc	0	0	0	0	1

Fig 9. One-Hot Encoding for the Geology variable, which contains 6 categories.
Source: Authors.

On the other hand, the input data are normalized by using $x_{norm} = (x_i - \mu) / \sigma$, where x_i is the datum, μ is the mean, and σ is the standard deviation, in order to prevent various orders of magnitude that may distort the calibration of the SVM model. Subsequent to this procedure, the input dataset was divided into two sets known as the training data (75%) and the test data (25%), with the aim of training and validating the SVM.

According to the theoretical framework, Support Vector Machine (SVM) models depend on a series of hyperparameters associated with the type of kernel used; therefore, it was necessary to define their optimal values through a systematic search process. To this end, a grid search was implemented, considering different parameter configurations, which were evaluated by using 10-fold cross-validation. In this procedure, the data are divided into 10 subsets of equal size; in each iteration, one of them is reserved for validation, and the remaining nine are used for training, repeating the process until all partitions are covered. Upon completion of the iterations, a set of 10 scores is obtained whose average corresponds to the model’s overall performance, thus allowing for the selection of the configuration with higher predictive efficiency. Finally, the main types of kernels commonly used in landslide analysis were evaluated, selecting the linear and Radial Basis Function (RBF) models due to their optimal performance reported in literature [44, 45].

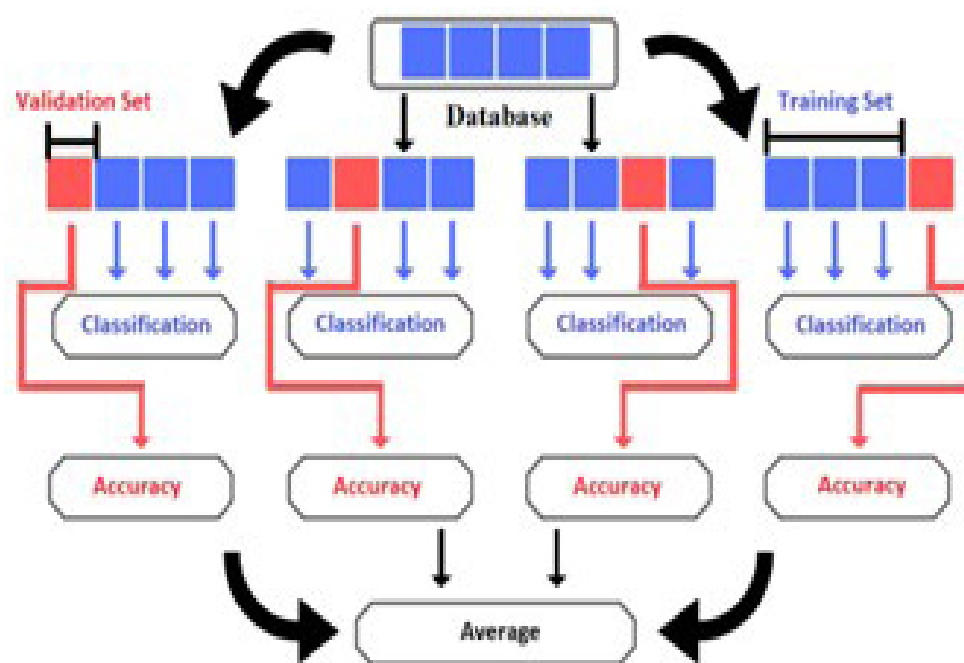


Fig 10. Procedure for 4-fold validation.
Source: Authors.

The parameters selected to perform the grid search were the following: $C \in [1, 500]$ with steps of 1, and for the RBF $\gamma \in [2^{-17}, 0.5000076293945312]$ with steps of 0.001. The following results were obtained. Table 1:

TABLE 1. RESULTS OF THE PARAMETERS FOR EACH KERNEL

Kernel		γ	Accuracy
Linear	2	-	0.98
RBF	257	0.00200762939453125	0.97

Source: Authors.

Once the parameters were defined, the SVM was implemented. Additionally, the model’s performance was validated and tested through the ROC curve, which presents a True Positive Rate (TPR) that indicates the model’s successes, -that is, the percentage of the landslide area that is correctly classified, (Y-axis)-, and a False Positive Rate (FPR) that indicates that the zones identified as unstable by the model are, in fact, stable zones on the process map (X-axis) [13], [46]. The Area Under the Curve (AUC) is within the interval [0,1] and represents the efficiency of our machine learning [45], [47], [48]. For the SVM model with the RBF kernel, 98% accuracy was obtained, and for the linear kernel, it was 97%. The ROC confirms that the True Positive Rate stabilizes rapidly, providing support for the results obtained in Fig 11.

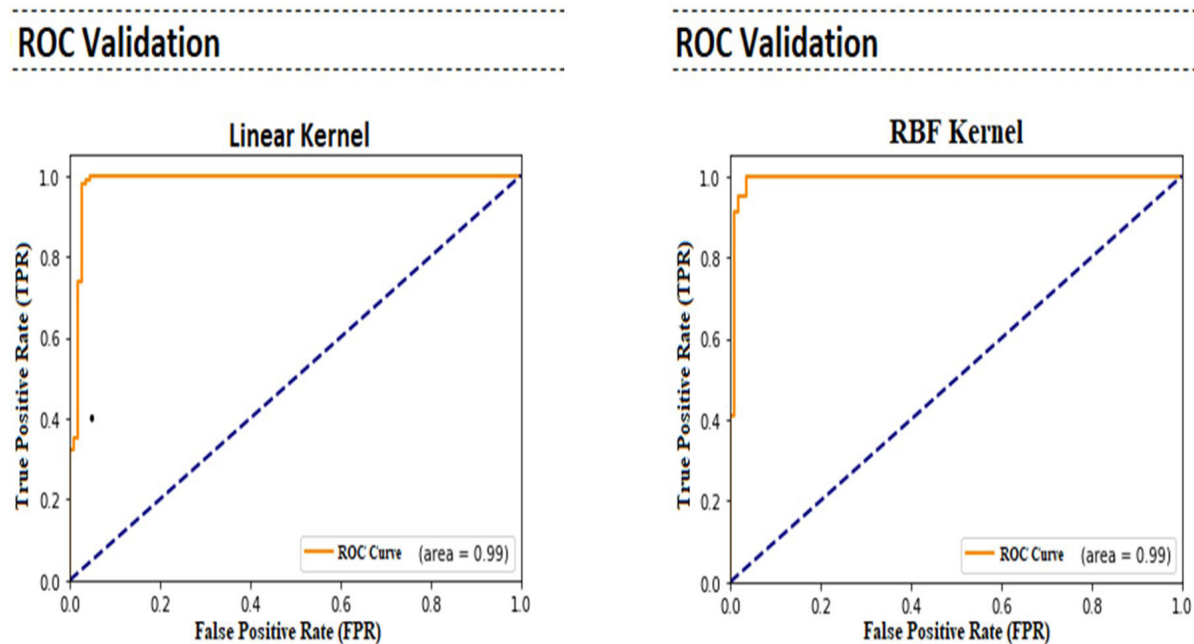


Fig 11. ROC Validation.
Source: Authors.

Once the SVM was implemented, the maps showing the landslide susceptibility in the locality Rafael Uribe Uribe were determined. By using the database (Fig 7) and the SVM models, the areas most prone to being stable and unstable were projected (Fig 12).

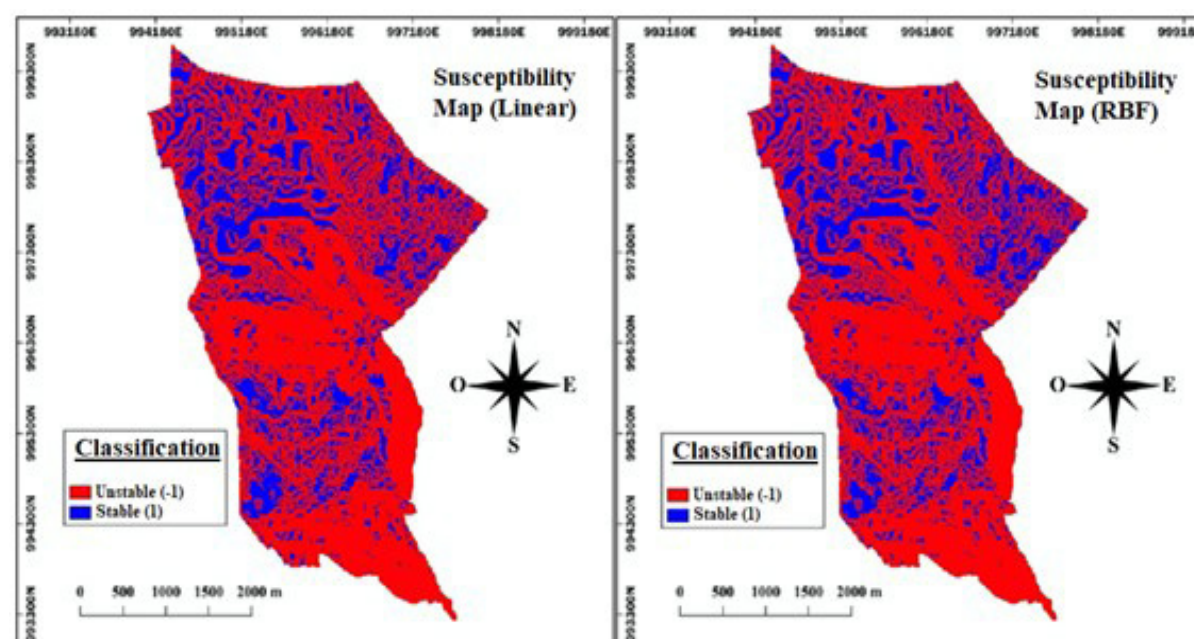


Fig 12. Landslide susceptibility maps for the linear and RBF kernels, classifying the stable and unstable areas.
Source: Authors.

The landslide susceptibility maps obtained are supported by the assumption that landslides are more likely to occur in situations similar to landslides in the past. The SVM, being a data-driven method, considers the principle of similarity from the geological theory, which allows to evaluate the spatial probability of future landslides to occur [49]. A similar condition was precipitation, due to its positive correlation with the registered events [31]; therefore, it was included as input information for the model.

The SVM shows better performance if the dataset is relatively small, but with large-scale data, its performance can be compromised [50]. Furthermore, it may have overestimation difficulties for the evaluation of landslide susceptibility [18]. Similarly, considering the available data, the model's accuracy, its robustness, the computational cost, and the clarity of the results, it could be considered that the SVM is a viable option for developing the evaluation of landslide susceptibility, compared to other data-driven techniques [49]. However, it is necessary to keep the susceptibility map updated because it is subject to influencing factors such as earthquakes, anthropic activities, alterations in terrain morphology, and intense precipitation events. It is also worth noting that the SVM does not consider the internal dynamic processes that develop during the landslide, such as hydrological processes.

The SVM, as a binary classifier, depends on the kernel used, but it has the advantage of having a unique solution. The RBF and linear kernels were used because they are considered the most suitable for classifying stable and unstable zones in a particular study area [45]. Nevertheless, future challenges for this paper include incorporating residential land-use

information, and also performing a sensitivity analysis of the input variables to identify thematic maps with a greater impact on the susceptibility map [20]. Moreover, machine learning models inherently contain uncertainty in the input variables, which was not considered here. To address this difficulty, hybrid models that manage to reduce uncertainty in susceptibility maps should be incorporated [51], [52], [24]. Similarly, there is the absence of data, which often results in approximating the locations where the landslide event occurs; even so, the data-driven method is very useful, which is valid for logistic regression, SVM, and random forest models [53]. Lastly, another challenge is to obtain an output in the three susceptibility levels, similar to the report of the SAB early warning system; the foregoing represents new opportunities in this line of research.

CONCLUSIONS

This paper contributes to strengthening knowledge on landslide risk management through the application of the SVM method in urban areas, with an emphasis on the Locality 18 of Bogotá. This approach constitutes a relevant input for territorial planning and the protection of community welfare, by offering an objective tool for the identification of susceptible areas.

The SVM methodology can be considered a potential complement to the Bogota Alert System (SAB, by its Spanish initials) currently operating in the city, by providing an analytical basis that improves the capacity for prediction and response to terrain instability events. Given the advancement rate of machine learning, it is plausible that these methodologies will transform risk assessment and decision-making processes in the future, significantly impacting preventive management and the formulation of public policies [27].

The susceptibility maps generated for Locality 18 through the SVM binary classification achieved success percentages of 98% for the RBF kernel and 97% for the linear kernel, which demonstrates a high level of model accuracy. Nevertheless, the effectiveness of this methodology can be increased by incorporating the sensitivity of the input variables, the analysis of data uncertainty, and the validation of results in the field with the support of experts. These actions would allow consolidating the applicability of the SVM as a robust tool for risk assessment and management in complex urban contexts.

ACKNOWLEDGMENTS

The authors thank all the government institutions for their collaboration, including IDEAM, IDIGER, the Colombian Geological Survey, the District Planning Secretariat, and the National Administrative Department for Statistics (DANE) for the secondary data provided to carry out this paper, and additionally, the Sergio Arboleda University for its support in the development of this research.

CRedit AUTHORSHIP CONTRIBUTION STATEMENT

M. Solano-Padilla: Methodology, Research, Data Curation, Visualization, Formal analysis, Writing- review & editing. C. Peña-Rincon: Conceptualization, Methodology, Research, Data Curation, Formal analysis, Writing-original draft, Writing-review & editing.

REFERENCES

- [1] Hungr, O., Leroueil, S., & Picarelli, L, “The Varnes classification of landslide types, an update”. *Landslides.*, vol. 11, pp. 167-194. 2014. <https://doi.org/10.1007/s10346-013-0436-y>.
- [2] Li, Y., & Mo, P., “A unified landslide classification system for loess slopes: A critical review”. *Geomorphology*, vol. 340, pp. 67-83. 2019. <https://doi.org/10.1016/j.geomorph.2019.04.020>.
- [3] Fan, X., Dufresne, A., Subramanian, S. S., Strom, A., Hermanns, R., Stefanelli, C. T. & Xu, Q., “The formation and impact of landslide dams—State of the art”. *Earth-Science Reviews*, vol. 203, pp. 103116. ,2020 <https://doi.org/10.1016/j.earscirev.2020.103116>.

- [4] Froude, M. J., & Petley, D. N. ,“Global fatal landslide occurrence from 2004 to 2016”. *Natural Hazards and Earth System Sciences*, vol.18, no. 8, pp. 2161-2181.2018. <https://doi.org/10.5194/nhess-18-2161-2018>.
- [5] Haque, U., Da Silva, P. F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A. & Glass, G. E, “The human cost of global warming: Deadly landslides and their triggers (1995–2014)”. *Science of the Total Environment*, vol. 682, pp. 673-684. 2019. <https://doi.org/10.1016/j.scitotenv.2019.03.415>.
- [6] Segoni, S., Piciullo, L., & Gariano, S. L., “A review of the recent literature on rainfall thresholds for landslide occurrence”. *Landslides*, vol. 15, no. (8), pp. 1483-1501.2018. <https://doi.org/10.1007/s10346-018-0966-4>.
- [7] Villegas-González, P. A., Ramos-Cañón, A. M., González-Méndez, M., González-Salazar, R. E., & De Plaza-Solórzano, J. S.,”Territorial vulnerability assessment frame in Colombia: Disaster risk management”. *International journal of disaster risk reduction*, vol. 21, pp. 384-395. 2017. <https://doi.org/10.1016/j.ijdr.2017.01.003>.
- [8] Sarmiento, J. P., Hoberman, G., Ilcheva, M., Asgary, A., Majano, A. M., Poggione, S., & Duran, L. R. , “Private sector and disaster risk reduction: the cases of Bogota, Miami, Kingston, San Jose, Santiago, and Vancouver”. *International Journal of Disaster Risk Reduction*, vol. 14, pp. 225-237,2015.<https://doi.org/10.1016/j.ijdr.2014.09.008>.
- [9] Guzman, L. A., Escobar, F., Peña, J., & Cardona, R., “A cellular automata-based land-use model as an integrated spatial decision support system for urban planning in developing cities: The case of the Bogotá region”. *Land use policy*, vol. 92, pp.104445, 2020. <https://doi.org/10.1016/j.landusepol.2019.104445>.
- [10] Korup, O., & Stolle, A., “Landslide prediction from machine learning”. *Geology today*, vol. 30, no. 1, pp. 26-33,2014. <https://doi.org/10.1111/gto.12034>.
- [11] Scheidegger, A.E. (1998). “Tectonic predesign of mass movements with examples from the Chinese Himalaya”. *Geomorphology*, vol. 26, no. (1-3), pp. 37-46,1988. [https://doi.org/10.1016/S0169-555X\(98\)00050-6](https://doi.org/10.1016/S0169-555X(98)00050-6).
- [12] Colombian Geological Survey. (2017). Methodological Guide for the Zoning of Mass Movement Hazard Scale (“Guía Metodológica para la Zonificación de Amenaza por Movimiento en Masa Escala”) 1:25.000, Available on: <https://doi.org/10.32685/9789585978225>.
- [13] Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., ... & Smith, J. T., “Recommendations for the quantitative analysis of landslide risk”. *Bulletin of engineering geology and the environment*, vol. 73, pp. 209-263, 2014. <https://doi.org/10.1007/s10064-013-0538-8>.
- [14] Peña Rincón, C. A., Precipitation Data with the HSB Model for Forecasting Shallow Landslide Susceptibility (“Datos de Precipitación con el modelo HSB para pronóstico de deslizamiento de suelos superficiales”), *Boletín de Geología*, vol. 39, no. 2, pp. 49-56, 2017. <https://doi.org/10.18273/revbol.v39n2-2017003>.
- [15] Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F., “A review of statistically-based landslide susceptibility models”, *Earth-Science Reviews*, vol. 180, pp. 60-91, 2018. <https://doi.org/10.1016/j.earscirev.2018.03.001>.
- [16] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer. Available on: <https://www.statlearning.com>.
- [17] Feizizadeh, B., Roodposhti, M. S., Blaschke, T., & Aryal, J., “Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping”. *Arabian Journal of Geosciences*, vol. 10, no. 5, pp. 1-13, 2017..<https://doi.org/10.1007/s12517-017-2918-z>.
- [18] Yao, X., Tham, L. G., & Dai, F. C. (2008). Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology*, vol. 101, no. 4, pp. 572-582. 2008. <https://doi.org/10.1016/j.geomorph.2008.02.011>.

- [19] Pourghasemi, H. R., Jirandeh, A. G., Pradhan, B., Xu, C., & Gokceoglu, C., “Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran”. *Journal of Earth System Science*, vol. 122, no.2, pp. 349-369, 2013,<https://doi.org/10.1007/s12040-013-0282-2>.
- [20] Lee, S., Hong, S. M., & Jung, H. S. (2017). A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability*, vol. 9, no. 1, pp. 48, 2017. <https://doi.org/10.3390/su9010048>.
- [21] Hong, H., Pradhan, B., Jebur, M. N., Bui, D. T., Xu, C., & Akgun, A., Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environmental Earth Sciences*, vol.75, no. 1, pp. 1-14, 2016. <https://doi.org/10.1007/s12665-015-4866-9>.
- [22] Ballabio, C., & Sterlacchini, S., “Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy”. *Mathematical geosciences*, vol. 44, no. 1, pp.47-70, 2012. <https://doi.org/10.1007/s11004-011-9379-9>.
- [23] Kumar, D., Thakur, M., Dubey, C. S., & Shukla, D. P., “Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology*”, vol. 295, pp. 115-125, 2017.<https://doi.org/10.1016/j.geomorph.2017.06.013>.
- [24] Althuwaynee, O. F., Pradhan, B., & Lee, S., “Application of an evidential belief function model in landslide susceptibility mapping”. *Computers & Geosciences*, vol. 44, pp. 120-135, 2012. <https://doi.org/10.1016/j.cageo.2012.03.003>.
- [25] Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., & Kanevski, M., “Machine learning feature selection methods for landslide susceptibility mapping. *Mathematical geosciences*”, vol.46, no.1, pp. 33-57, 2014. <https://doi.org/10.1007/s11004-013-9511-0>.
- [26] Pradhan, B., “A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS”. *Computers & Geosciences*, vol.51, pp. 350-365, 2013.<https://doi.org/10.1016/j.cageo.2012.08.023>.
- [27] Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S., “Machine learning and landslide studies: recent advances and applications”. *Natural Hazards*, vol. 114 ,no. 2, pp. 1197-1245, 2022. <https://doi.org/10.1007/s11069-022-05423-7>.
- [28] DANE (2022), Total GDP by Departments, Sheet:charts 1.Available on: <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales>.
- [29] District Secretariat of Government Bogotá D.C. (2013). Statement of Reasons for Draft Agreement 279, whereby the District System for Prevention and Emergency Care, SDPAE, is transformed into the District System for Risk Management and Climate Change, SDGR–CC, and its instances are updated. Available on: <https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=55279>.
- [30] District Planning Secretariat Bogotá D.C. (2017). Monographs of the Localities of the Capital District; Rafael Uribe Uribe. Available on: <https://acortar.link/t0fZTW>
- [31] Acosta Jiménez, Y., & Peña Rincón, C. A., “Physical Vulnerability Index before the Landsliding Hazard in Houses Located in the Locality of Rafael Uribe Uribe, Bogotá DC”. *Tecciencia*, vol. 14, no. 27, pp. 1-9, 2019. <https://doi.org/10.18180/tecciencia.2019.27.1>
- [32] Hospital “Rafael Uribe Uribe” (HRUU) E:S:E, local diagnosis “Rafael Uribe”. (2012). Available on: http://www.saludcapital.gov.co/DSP/Diagnosticos%20Locales/Locales%20Preliminares/18_Rafael_Uribe.pdf.
- [33] QGIS Team. (2009). QGIS Development Team. Geographic Information System Software. Open-source Geospatial Foundation. Available on: <http://qgis.org>.
- [34] ArcGIS, (2018). ArcGIS geoprocessing tool reference. Available on: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/main/arcgis-pro-tool-reference.htm>.

- [35] Montoya, S. (29 November 2013). Saga GIS Tools in QGIS for Hydrological Analysis (“Herramientas de Saga Gis en QGis para análisis hidrológico”). Available on: <https://gidahatari.com/ih-es/herramientas-saga-gis-en-qgis-para-analisis-hidrologico>.
- [36] Gruber, S.; Peckham, S. “Land-surface parameters and objects in hydrology”. Vol. 33, Elsevier, 2009, pp. 171-194.
- [37] Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. (2015): System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, *Geosci. Model Dev.*, 8, pp. 1991-2007, Available on: doi:10.5194/gmd-8-1991-2015.
- [38] Cortes, C., & Vapnik, V., “Support-vector networks”. *Machine learning*, vol. 20, no.3, pp. 273-297,1995. <https://doi.org/10.1007/BF00994018>.
- [39] Vapnik, V.N, *The Nature of Statistical Learning Theory*. Springer, New York.2000, pp. 225-265.
- [40] Zaki M, Meira Jr W, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press. 2020, pp. 134-161, pp.517-523.
- [41] Hastie, T., Tibshirani, R. and Freidman, J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer-Verlag, New York, 2001, pp. 165-192.
- [42] Cristianini N, Shawe-Taylor J, *An introduction to support vector machine and other kernel-based learning methods*. Cambridge University Press, 2000, pp. 26-51, pp. 93-124.
- [43] Rudin, C. (2012).15.097 Prediction: Machine Learning and Statistics, lecture 13. Kernel. https://ocw.mit.edu/courses/15-097-prediction-machine-learning-and-statistics-spring-2012/resources/mit15_097s12_lec13/.
- [44] Hsu Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin.(2003), A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University. Available on: <http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf>.
- [45] Abraham, M. T., Satyam, N., Jain, P, Pradhan, B., & Alamri, A. (2021). Effect of spatial resolution and data splitting on landslide susceptibility mapping using different machine learning algorithms. *Geomatics, Natural Hazards and Risk*, 12(1), 3381-3408. <https://doi.org/10.1080/19475705.2021.2011791>.
- [46] Aristizábal, E., Morales-García, P., Vásquez-Guarín, M., Ruíz-Vásquez, D., Palacio-Córdoba, J., Ángel-Cárdenas, F. P. & Ordóñez-Carmona, O., Methodologies for the evaluation of landslide hazard as part of the basic hazard studies: case study municipality of Andes, Antioquia, Colombia. (“Metodologías para la evaluación de la amenaza por movimientos en masa como parte de los estudios básico de amenaza: caso de estudio municipio de Andes, Antioquia, Colombia”), *Boletín de Geología*, vol. 44, no.3, pp. 199-217, 2022. <https://doi.org/10.18273/revbol.v44n3-2022009>.
- [47] Refaeilzadeh P., Tang L., Liu H. (2009) Cross-Validation. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_565.
- [48] Shahabi, H., & Hashim, M.,” Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment”. *Scientific reports*, vol. 5, no. 1, pp. 1-15, 2015. <https://doi.org/10.1038/srep09899>.
- [49] Huang, Y., & Zhao, L., “Review on landslide susceptibility mapping using support vector machines”. *Catena*, vol. 165, pp. 520-529, 2018. <https://doi.org/10.1016/j.catena.2018.03.003>.
- [50] Pham, B.T., Tien Bui, D., Prakash, I., Nguyen, L.H., Dholakia, M.B., “A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS”. *Environ. Earth Sci.* vol. 76, no. 10, pp. 1-15, 2017. <https://doi.org/10.1007/s12665-017-6689-3>.

- [51] Dikshit, A., Sarkar, R., Pradhan, B., Acharya, S., & Alamri, A. M., “Spatial landslide risk assessment at Phuentsholing, Bhutan”. *Geosciences*, vol.10, no. 4, pp. 131, 2020. <https://doi.org/10.3390/geosciences10040131>.
- [52] Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., ... & Abderrahmane, B., “Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance”. *Earth-Science Reviews*, vol. 207, pp. 103225, 2020. <https://doi.org/10.1016/j.earscirev.2020.103225>.
- [53] Zhu, A. X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., & Hong, H., “A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods”. *Catena*, vol. 183, pp. 104188, 2019. <https://doi.org/10.1016/j.catena.2019.104188>.