remote sensing

MDPI

Article

Landslide Susceptibility Analysis on the Vicinity of Bogotá-Villavicencio Road (Eastern Cordillera of the Colombian Andes)

María Camila Herrera-Coy [1], Laura Paola Calderón [2], Iván Leonardo Herrera-Pérez [1,3], Paul Esteban Bravo-López [1,4], Christian Conoscenti [2], Jorge Delgado [1], Mario Sánchez-Gómez [5,6] and Tomás Fernández [1,6,*]

1 Department of Cartographic, Geodetic and Photogrammetric Engineering, University of Jaén, 23071 Jaén, Spain; mchc0009@red.ujaen.es (M.C.H.-C.); ilhp0001@red.ujaen.es (I.L.H.-P.); pebl0001@red.ujaen.es (P.E.B.-L.); jdelgado@ujaen.es (J.D.)
2 Department of Earth and Marine Sciences (DiSTeM), University of Palermo, 90123 Palermo, Italy; laurapaola.calderoncucunuba@unipa.it (L.P.C.); christian.conoscenti@unipa.it (C.C.)
3 Department of Geographic and Environmental Engineering, University of Applied and Environmental Sciences (U.D.C.A.), Bogotá 111166, Colombia
4 Institute for Studies of Sectional Regime of Ecuador (IERSE), University of Azuay, Cuenca 010107, Ecuador
5 Department of Geology, University of Jaén, 23071 Jaén, Spain; msgomez@ujaen.es
6 Natural Hazards Lab of the Centre for Advanced Studies in Earth Sciences, Energy and Environment (CEACTEMA), University of Jaén, 23071 Jaén, Spain
* Correspondence: tfernan@ujaen.es; Tel.: +34-53-212843

Abstract: Landslide occurrence in Colombia is very frequent due to its geographical location in the Andean mountain range, with a very pronounced orography, a significant geological complexity and an outstanding climatic variability. More specifically, the study area around the Bogotá-Villavicencio road in the central sector of the Eastern Cordillera is one of the regions with the highest concentration of phenomena, which makes its study a priority. An inventory and detailed analysis of 2506 landslides has been carried out, in which five basic typologies have been differentiated: avalanches, debris flows, slides, earth flows and creeping areas. Debris avalanches and debris flows occur mainly in metamorphic materials (phyllites, schists and quartz-sandstones), areas with sparse vegetation, steep slopes and lower sections of hillslopes; meanwhile, slides, earth flows and creep occur in Cretaceous lutites, crop/grass lands, medium and low slopes and lower-middle sections of the hillslopes. Based on this analysis, landslide susceptibility models have been made for the different typologies and with different methods (matrix, discriminant analysis, random forest and neural networks) and input factors. The results are generally quite good, with average AUC-ROC values above 0.7–0.8, and the machine learning methods are the most appropriate, especially random forest, with a selected number of factors (between 6 and 8). The degree of fit (DF) usually shows relative errors lower than 5% and success higher than 90%. Finally, an integrated landslide susceptibility map (LSM) has been made for shallower and deeper types of movements. All the LSM show a clear zonation as a consequence of the geological control of the susceptibility.

Keywords: landslide; susceptibility analysis; modelling; Bogotá-Villavicencio road; Eastern Cordillera; Colombian Andes

Citation: Herrera-Coy, M.C.; Calderón, L.P.; Herrera-Pérez, I.L.; Bravo-López, P.E.; Conoscenti, C.; Delgado, J.; Sánchez-Gómez, M.; Fernández, T. Landslide Susceptibility Analysis on the Vicinity of Bogotá-Villavicencio Road (Eastern Cordillera of the Colombian Andes). *Remote Sens.* **2023**, *15*, 3870. https://doi.org/10.3390/rs15153870

Academic Editor: Domenico Calcaterra

Received: 11 June 2023
Revised: 24 July 2023
Accepted: 31 July 2023
Published: 4 August 2023

1. Introduction

Landslides are considered one of the most important natural hazards worldwide, causing thousands of casualties and costs amounting to billions of euros each year [1–3]. Compared to other risk phenomena such as earthquakes or floods, the effect of landslides is more diffuse and continuous in space and time, so their impact can be underestimated

according to some evaluations [2]. Nevertheless, they cause significant damage to infrastructure, properties and the environment, as well as the interruption of socioeconomic activity [1,4].

In Colombia, the occurrence of natural hazard phenomena, including landslides, earthquakes, and volcanic eruptions, is very frequent due to its geographical location in the Andean mountain range, with a steep orography, a great geological complexity and a significant climatic variability [5–8]. In fact, according to [9], it is one of the most prominent countries in global databases such as the Disaster database (EM-DAT [10]), the Disaster Inventory System (DesInventar [11]), the Global Landslide Catalog (GLC [12]) and the Global Fatal Landslide database (GFLD [13]). Specifically, DesInventar [11] reports 10,559 incidents and 7400 deaths for Colombia.

Thus, landslides in Colombia represent almost half of all natural catastrophic events, far exceeding disasters caused by floods, earthquakes and volcanic eruptions [14], with an average of 47 landslides and 59 deaths each year from 1993 to 2004. According to information available in the Mass Movements Information System (SIMMA) of the Colombian Geological Service (SGC [15]), 135,632 mass movements have been reported in the country since 1900. Due to this, 31,631 people have lost their lives, and 68,792 families have been affected. Combining different national and international databases, García-Delgado et al. [8] collected a total of 2351 fatal landslides that caused almost 40,000 deaths, with some of them in historical times (prior to 1912) and the majority in modern times (1912–2020), with an upward trend in the last 20 years. In another work, Aristizábal and Sánchez [6] compiled about 30,730 landslides that caused 34,198 fatalities and economic losses of more than 600 million dollars in the period of 1900–2018.

According to SIMMA [15], among the most affected regions in absolute terms are the departments of Cundinamarca, Boyacá and Norte de Santander, located in the Eastern Cordillera at the central and northern part of the country, as well as Cauca in the Colombian Massif in the south. Other smaller departments, such as Caldas, Risaralda or Quindío in the Central Cordillera or Atlántico in the Sierra de Santa Marta, also present a considerable density in relative terms [15]. These data coincide with the compilation of García-Delgado et al. [8], in which the highest densities of landslides occur in the departments of Tolima, Caldas, Risaralda, Bogotá, Quindío, Cauca and Cundinamarca. Specifically, the departments of Cundinamarca and Meta in the central sector of the Eastern Cordillera are exposed to medium and high probabilities of occurrences of catastrophic phenomena, particularly landslides, caused, among other reasons, by a high anthropic intervention on the slopes with the consequent deterioration of the hydrographic basins and their stability conditions. Thus, the Subdirectorate of Geoambiental Engineering of Ingeominas (now Subdirectorate of Geo-Hazards of the SGC) prioritized six regions with a higher concentration of phenomena: the Guavio river basin, the area around the Bogotá-Villavicencio road, the eastern slope of the Negro River, the Sumapaz river basin, the middle basin of the Bogotá River and the municipality of San Cayetano [16]. This work focuses on the vicinity of the road from Bogotá to Villavicencio, specifically on its section towards the Orinoco river basin.

One of the most effective measures for risk prevention and mitigation is the evaluation of both the hazard of the phenomenon and of the exposure and vulnerability of the elements at risk [17]. In Colombia, some studies have been carried out that evaluate risk assessment and reduction [7,18–20], but there are many more that evaluate hazard or susceptibility. For hazard, there are numerous deterministic and probabilistic methods, the latter of which are generally the most applied for extensive areas due to the lack of precise and exhaustive data in such areas. Within the deterministic methods, different hydrological models [21,22] or stability analysis methods such as infinite slope [23], r.slope.stability of GRASS [24], FOSM [25–27], SLIP [26,27], PEM point estimates [25] or deformation analysis [28] have been applied. These models have been used in high-impact landslides such as the Mocoa debris flow in the southern part of the country [21,29–31], hillslopes around Medellin [22]

and San Eduardo in Boyacá [28] or watersheds such as La Arenosa and La Liboriana [24–27] in the department of Antioquia in the Western Cordillera.

Meanwhile, within probabilistic methods, most are extensively applied in the estimation of spatial probability or susceptibility, which are based on the statistical analysis of correlation between determinant factors and landslides, according to the susceptibility definition established by Brabb [32]. To develop susceptibility models using statistical techniques, numerous methods are available, which, according to Reichenbach et al. [33], can be grouped into models based on indices [34–36] and bivariate statistics [37]; multicriteria evaluation [34]; multivariate statistics [4,38,39]; machine learning, ML [40] and artificial neural networks (ANN) [41,42]. The last two groups, sometimes with the ANN integrated in the more general group of ML, have advanced regarding classical statistical methods due to their greater versatility and better performance in nonlinear systems such as models developed from factors of different nature [43]. These methods allow the integration of a great number of factors that are not usually analyzed and selected since the algorithms directly perform the fit of the models [33]. In our opinion, this lack of factor control and selection leads to a loss of knowledge in the elaboration of the models and sometimes to an overfitting of them.

Different studies have been conducted in the Colombian Andes, using all types of methods from index-based methods such as frequency ratio or weight of evidence [44–48]; statistical methods such as logistic regression [31,46,48–51]; machine learning methods such as random forest [46], graded boosted regression trees, GBRT [46] or multivariate adaptive regression [31] and simple [46,52] or convolutional neural networks [53]. The methods have been applied to different areas such as Capitanejo in the NE of the country [52], Medellín and the department of Antioquia to the NW [47–49], Caldas [44], Cauca [45,50], Boyacá [52], Bogotá and Cundinamarca [46,51] and Mocoa in the south [31]. Most of these studies, not only in Colombia but also throughout the world, do not take into account the landslide typology, which produces less precision and noise in the models.

Finally, it should be mentioned that those methods for determining rainfall thresholds and, where appropriate, establishing early warning systems, exist throughout the whole country [54] or in different departments such as Bolívar, Antioquia and Caldas to the northwest [50,55–58] or in Bogotá [59]. Regarding precipitation, the influence of deep convective systems [60] and the impact of climate change on the generation of landslides and other risks [61] have also been evaluated in these previous works.

The main objective of this work is to present a detailed inventory and susceptibility models of the study area. The inventory will allow the understanding of the different processes that occur in this mountainous area and their characteristics, while the subsequent factor analysis by typologies will allow the factor selection and the determination of the conditions under which they originate, as a previous step to modeling landslide susceptibility. Thus, an important limitation of most current studies, such as the lack of knowledge of the different landslide typologies, factors and conditions, can be overcome. From this knowledge, susceptibility models for the different typologies have been developed using several methods and introducing an increasing number of factors, which will allow a control of the models' behavior in relation to overfitting and noise. Examples of the main groups of methods that have been used are the matrix method (index), linear discriminant analysis (multivariate statistics), random forest (machine learning) and a perceptron ANN (neural networks). The results have been compared in order to provide consistent results in determining the landslide hazard in the region and ensuring robust models that can be applied in other areas. Finally, once the susceptibility models from the different typologies are obtained and selected, they have been integrated into susceptibility maps for shallower and deeper processes in order to assess the hazard in the study area.

2. Materials and Methods

2.1. Study Area

The study area is located in the central sector of the Eastern Cordillera of the Colombian Andes (Figure 1a). This is a mountain chain that extends for about 1000 km from southern Colombia (Colombian Massif, where the cordillera divides into its three main branches, Western, Central and Eastern) to near the border with Venezuela in the north (Sierra Nevada del Cocuy), where the highest altitudes of around 5400 m are reached. To the west of the Cordillera lies the Cundinamarca Plateau, where the capital of Bogotá (Figure 1b) and many other towns are located; from there, it descends to the Magdalena River, which flows between the Eastern and Central Cordilleras; to the east, the basins of the Orinoco and Amazon rivers extend. In this central sector of the Eastern Cordillera, altitudes range from a few hundred meters in the river basins to 4000 m at the summits. The Eastern Cordillera of Colombia is an intracontinental mountain belt 100 to 200 km wide [62] with a SW-NE trend. The materials correspond mainly to marine deposits but also transitional to continental, ranging from the Cretaceous to Paleocene in age [63,64]. On both sides of the Cordillera, in the Magdalena River valley and the Amazon and Orinoco basins, there are Tertiary sedimentary or volcanoclastic deposits. Above all of them appear Quaternary materials: alluvial, colluvial and paludal fillings. The structure consists of thrusts and folds, which, in some cases, bring to the surface metamorphic materials of the Paleozoic substrate (Ordovician and Devonian) without reaching the Proterozoic crystalline basement.

More specifically, the study area has an extension of approximately 746 km^2 on the vicinity of the road between Bogotá and Villavicencio (Figure 1b,c), a city located 75 km southeast of the country capital (120 km by this road, also known as "Vía al Llano" or Route 40). The area extends through the municipalities of Cáqueza, Fosca, Quetame and Guayabetal in the Oriente province of the department of Cundinamarca and the municipality of Villavicencio in the department of Meta. The municipalities' total number of inhabitants and percentage of rural population are Cáqueza with 15,594 inhabitants and 58%; Fosca, 5578 inhabitants and 75%; Quetame with 4929 inhabitants and 77%; Guayabetal with 5809 inhabitants and 70% and Villavicencio with 451,212 inhabitants, of which only the 7% are rural population [65]. The road between Bogotá and Villavicencio connects the entire area together other minor roads.

The elevations range from 600 m in Villavicencio up to 3500 m in the mountain range to the east of Quetame. The average annual rainfall varies from 500 mm in the western sector to over 3000 mm in the Villavicencio sector. The slopes are generally quite steep, with almost 64% in the range of 20–45°. Hydrographically, it corresponds to the Negro river basin, a tributary of the Guayuriba River that in turn flows into the Orinoco river basin.

From a geological point of view, materials from the Paleozoic substrate and Cretaceous sedimentary series [66] outcrop in this area, with the former in the lower part of the basin and the latter in the higher part (Figure 1d). Within the Paleozoic, there are two sets of materials: first, metamorphic rocks of low grade, phyllites, schists and quartzites of Ordovician age; over them, discordantly, there are quartz sandstones and shales of Devonian-Carboniferous age. In the Cretaceous series, there is a small outcrop of conglomerates and transitional environment sands at the base of the series, which pass to lutites of a marine environment, which are predominant in the area. These series end in Paleogene, and then Miocene sedimentary and volcanoclastic deposits fill the basins formed within the Cordillera and at the east over the Paleozoic and Precambrian basement. The structure is of thrusts and folds with NNE-SSW main direction, which allow the outcrop of the underlying Paleozoic formations to the Cretaceous series in the lower part of the area. On top of all these sets, there are Quaternary materials consisting of terraces, colluvial deposits and current alluvial deposits in the riverbeds.

In the municipalities of the area, landslide activity is very high, according to SIMMA [15]: Cáqueza has had 169 incidences; Quetame, 22 incidences; Guayabetal, 9 incidences and Villavicencio, 94 incidences.

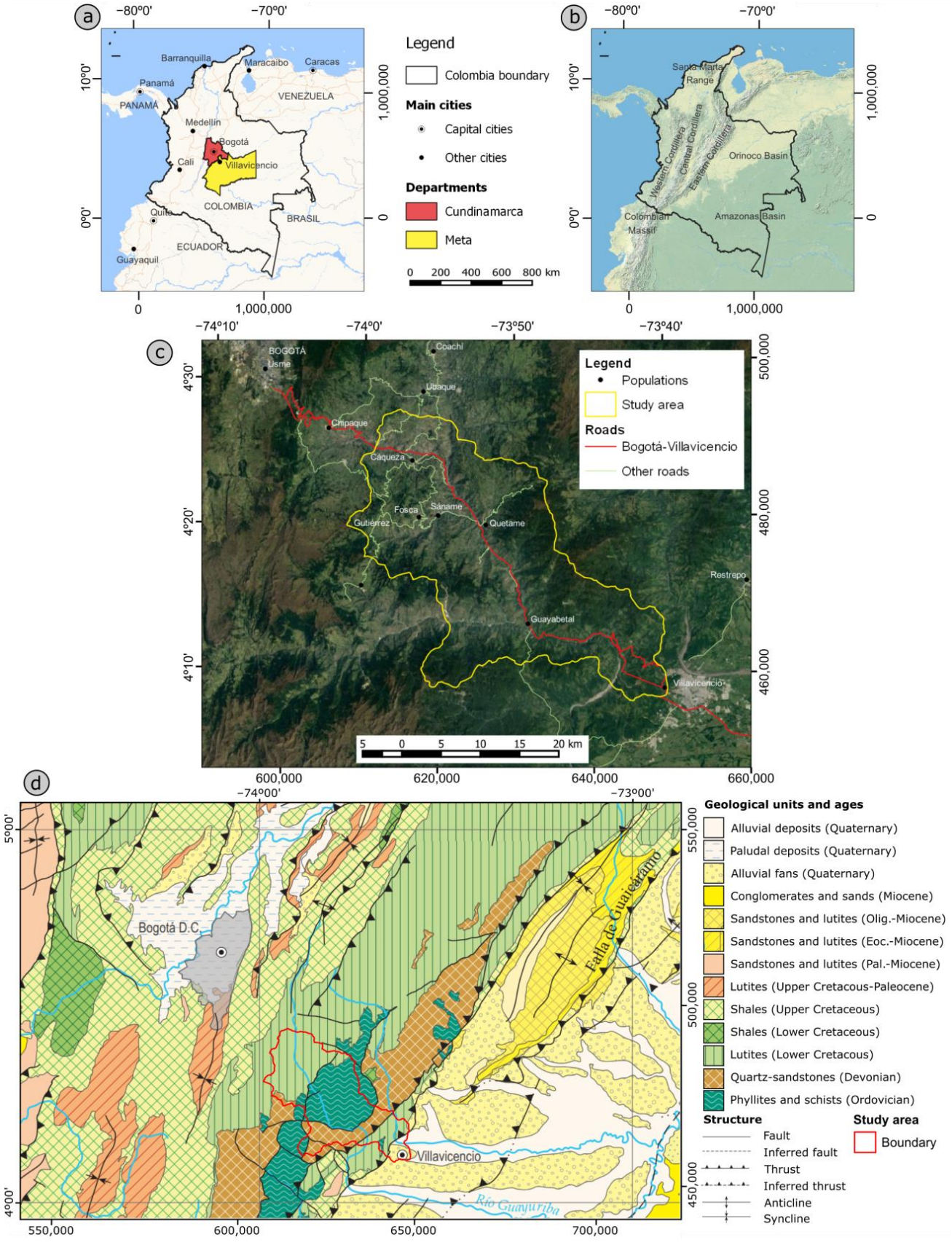

Figure 1. Location and geology of study area: (**a**) Location of Cundinamarca and Meta Departments in Colombia (own elaboration on AutoNavi Base Maps); (**b**) Colombian Andes and Eastern Cordillera

(own elaboration on Esri Physical Map); (**c**): Study area and main populations (own elaboration on Google Satellite); (**d**) Geological setting adapted from the Geological Map of Colombia [64]. Coordinates are in WGS84 (lat/long at left and top margins) and in WGS84-UTM 18 (projected, in right and down margins).

2.2. Materials

The sources of information and software used are listed in Table 1. For the inventory, background images from Google Maps—Google Earth (GE-GM) and Bing Maps (BM) were used, corresponding to Airbus (Pleiades), Maxar and Copernicus (Sentinel-2). In GE, images of different dates and resolutions could be observed.

Regarding the factor layers, the digital elevation model (DEM) obtained by InSAR from ALOS PALSAR images of 2011 [67] was used, with a spatial resolution of 12.5 m, downloaded from the Alaska Satellite Facility [68]. The geology comes from the Geological Atlas of Colombia [69], available as vector information, from which the lithological units have been extracted. A Sentinel image from 2020 [70] was used to calculate the NDVI index and obtain the classification of land cover. Finally, the map of average precipitation available in IDEAM [71] was used.

Table 1. Sources of information and software used in this study.

Information	Resources	Software
Digital elevation model (12.5 m resolution)	JAXA/METI ALOS PALSAR, 2011 [67,68]	Google Earth 7.3.6.9345 [72]
Background images	GM, GE, BM (Airbus, Maxar, Copernicus)	QGIS versión 3.18.3 [73]
Geology: Geological Atlas of Colombia	Layer files (shp): Servicio Geológico Colombiano, 2015 [69]	SAGA versión: 7.9.1 [74]
Sentinel-2 image	Copernicus, 2020 [70]	Rstudio 2022.02.2 [75]
Precipitation in Colombia	Raster files (tif): IDEAM, 2015 [71]	

Regarding software, Google Earth Pro 7.3.6.9345 [72] was used for image visualization, and QGIS 3.18.3 [73] and SAGA GIS 7.9.1 [74] for data processing and analysis. Additionally, R studio 2022.02.2 statistical software [75] was used for generating multivariate statistical and machine learning susceptibility models.

2.3. Methodology

The methodology is summarized in the flowchart of Figure 2. It includes first the elaboration of a detailed landslide inventory in the study area. Second, it includes the analysis of landslide determinant factors by typologies in order to the factor selection and understanding the conditions of their occurrence. Third, it includes the elaboration of susceptibility models (LSM) for each landslide typology using different methods, as well as their validation, and finally, it includes the integration of LSM in synthesis maps.

2.3.1. Landslide Inventory

The landslide inventory was carried out using photointerpretation and digitization from the GM-GE and BM background images. In addition, the database of the Colombian Mass Movements Information System, SIMMA [15], was used as support. The digitization of the identified movements was carried out by connecting to these images through WMS from the open-source software QGIS, although the photointerpretation was helped by the pseudo-3D views of the images in GE (Airbus/Pleiades, Maxar and Copernicus/Sentinel-2).

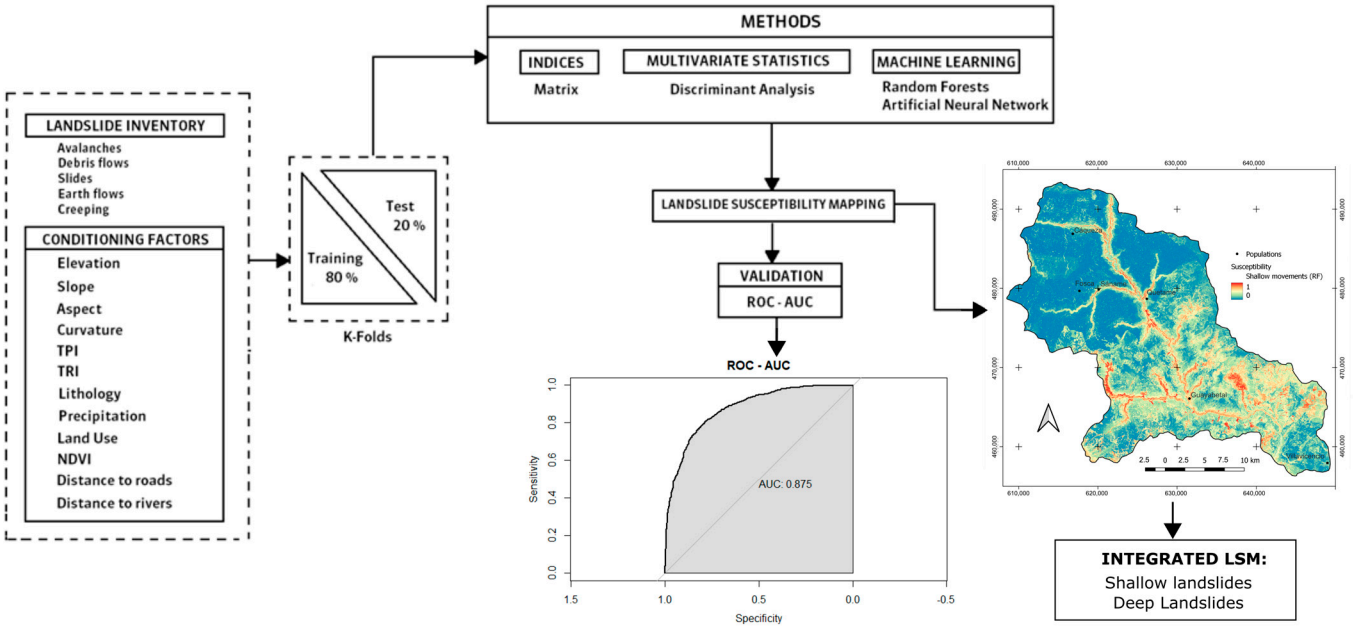

Figure 2. Flowchart followed in this study.

Once digitized, a database was created, which, according to Chacon et al., 2006 [76] and Guzzetti et al., 2012 [77], should include the spatial location, temporal dating and thematic attributes of the landslides. Specifically, this database or inventory includes as attributes the typology [78] and the activity [79], which were supported by the multi-temporal GE images, and their area, calculated with the attribute calculation tools of QGIS. From this database, an analysis has been carried out that will allow us to know the frequency and total extension, the activity and the average area of each landslide typology.

2.3.2. Analysis of Determinant Factors

For the analysis of determinant factors and the elaboration of LSM with GIS, it is necessary to have the factor layers. In this case, factor layers and maps have been obtained from different official geographic information sources in Colombia (Table 2). Figure 3 shows the factor maps both the quantitative (DEM derivatives, precipitation, NDVI and distance to roads and rivers) and the qualitative (lithology and land cover).

Table 2. Factors used in the analysis and their corresponding sources of information.

Factor	Origin
Elevation Slope Aspect Curvature Topographical Position Index (TPI) Terrain Roughness Index (TRI)	Derived from DEM of 12.5 m resolution from JAXA-ALOS PALSAR [67,68]
Lithology	Geological Atlas of Colombia, SGC [69]
Precipitation	Raster files (tif): IDEAM, 2015 [71]
Land Cover Normalized Difference Vegetation Index (NDVI)	Sentinel-2 image, Copernicus 2020 [70]
Distance to roads	Roads digitized on GE-GM image
Distance to rivers	Rivers digitized on GE-GM image

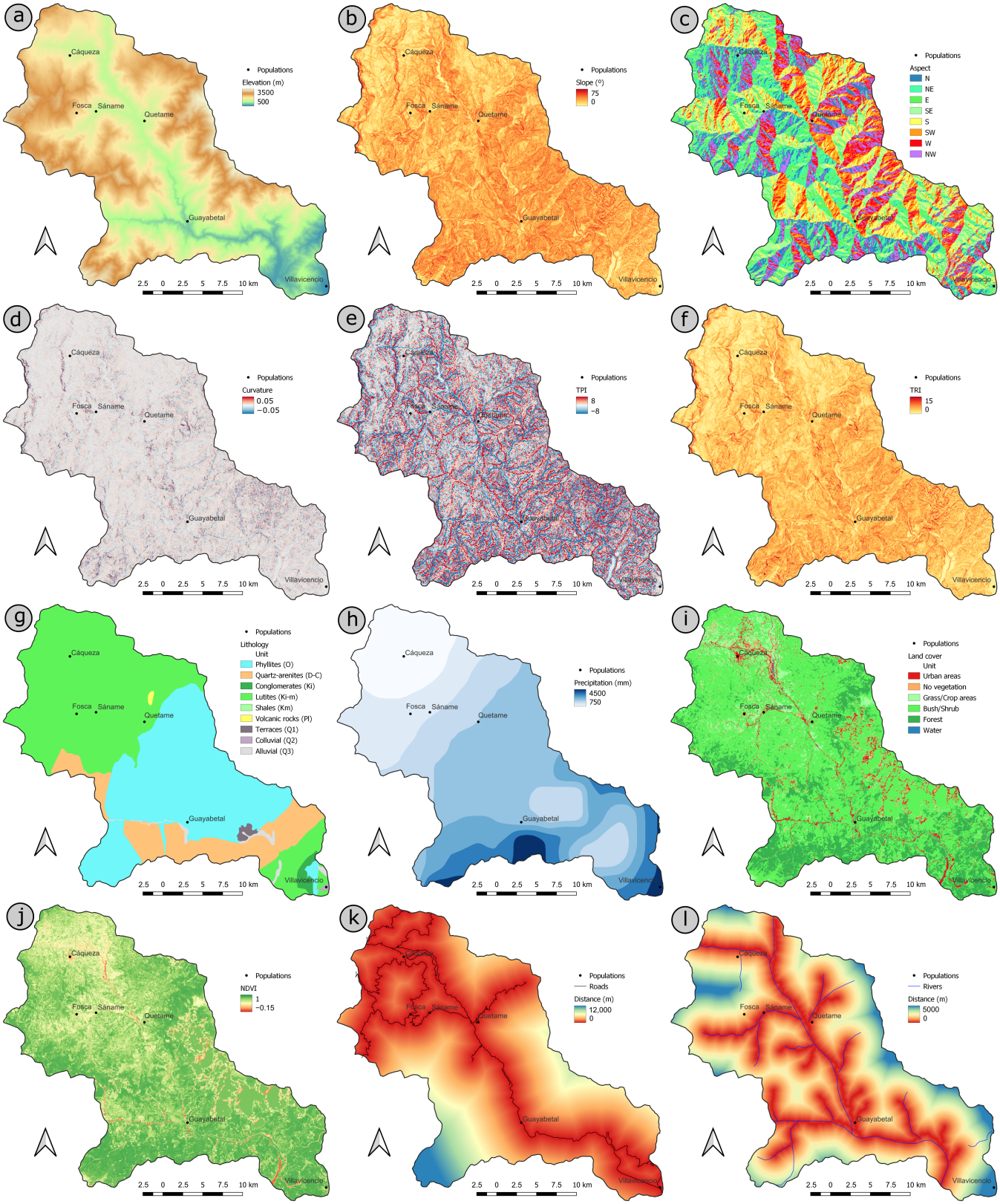

Figure 3. Factor maps considered in the study area; (**a**): Elevation; (**b**): Slope; (**c**): Aspect; (**d**): Curvature; (**e**): TPI; (**f**): TRI; (**g**): Lithology; (**h**): Precipitation; (**i**): Land Cover; (**j**): NDVI; (**k**): Distance to roads; (**l**): Distance to rivers.

From the digital elevation model (DEM), derivative models such as slope, aspect or orientation, terrain curvature, topographic position index (TPI) and terrain roughness index (TRI) have been obtained using QGIS analysis functions. Additionally, lithological units extracted from the Geologic Atlas of Colombia [69] were used, which were rasterized; then, a quantitative value was assigned to each unit based on material resistance [80], as shown in Table 3 (lower to hard rocks and higher to soft rocks). The NDVI and land use were obtained from a Sentinel-2 image using the corresponding formula [81] and supervised classification (maximum probability), respectively. In the case of land cover, a similar scheme to lithology was followed, assigning a value to each unit based on the vegetation cover or other considerations [82], also shown in Table 3. Precipitation data were obtained directly from a raster layer of precipitation intervals [71]. Finally, the distance to rivers and roads was obtained through vector digitization on the GM/GE background image and subsequent distance calculation using the corresponding QGIS function.

Table 3. Values assigned to lithological and land cover units. In lithological units, the values are assigned to each unit based on material resistance (lower to hard rocks and higher to soft rocks). In land cover units, values are assigned to each unit based mainly on the vegetation cover.

Lithology		Land Cover	
Unit	**Value**	**Unit**	**Value**
Phyllites-Schists	0.3	Urban	0.6
Quartzarenites	0.4	Scarce vegetation	0.8
Conglomerates	0.8	Grass-Crops	0.5
Lutites	1.0	Bush-Shrubs	0.4
Shales	0.9	Forest	0.2
Volcanic	0.2	Water	0
Terraces	0.5		
Alluvial fans	0.6		
Alluvial deposits	0.7		

The factorial analysis consisted of cross-tabulating the maps of determinant factors and the landslide inventory, both global and/or by typologies. Then, the Kolmogorov–Smirnov (K–S) coefficient was calculated to compare the distributions of factors in areas affected and not affected by movements, thus estimating the correlation between factors and landslides. At the same time, an analysis was conducted among the factors themselves by determining the Pearson linear correlation coefficient in order to estimate the collinearity between them. These analyses allowed for the selection of factors involved in the models and the identification of conditions for the occurrence of the different landslide typologies. The analysis has been made taking into account the landslide activity, although in most typologies, the differences are not significant, so only the general results are shown in the next section.

2.3.3. Susceptibility Models

Susceptibility models have been developed using different methodologies [33]: matrix method (indices), lineal discriminant analysis, LDA (classical multivariate statistics), random forest, RF (machine learning) and a simple artificial neural network (ANN) (Figure 4).

In the matrix method, the procedure starts by combining the raster layers of factors in order to obtain the units of unique condition. For qualitative factors, discrete values are used (Table 2), and for quantitative factors, continuous values are classified in intervals. Next, a cross-tabulation is performed between the raster layer of unique condition and the binary raster layers of presence/absence of landslides by typologies. From this table, the percentage area of each combination of factors affected by landslides can be obtained [35,36,83,84]. This is the susceptibility index, which is obtained by the expression:

$$I_i = \frac{z_i}{s_i} \times 100 \quad (1)$$

where I_i is the susceptibility index; z_i and s_i are, respectively, the area of each factor combination affected by landslides and the total area of the combination.

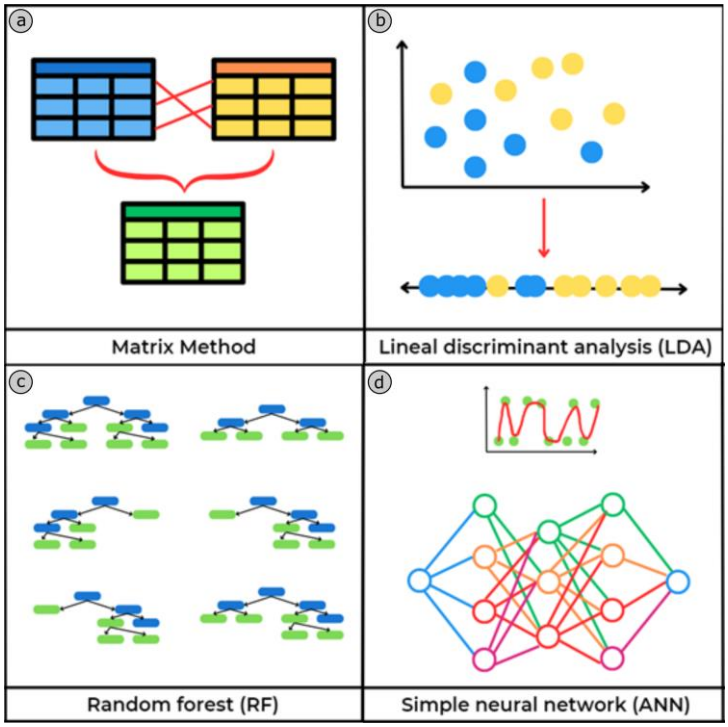

Figure 4. LSM methods: (**a**) Matrix; (**b**) Linear discriminant analysis; (**c**) Random forest; (**d**) Artificial neural network.

This index is used as a classification template, and thus, the susceptibility map is obtained by reclassifying the unique condition raster layer with it [35,36,83,84].

Meanwhile, the remaining methods follow a different scheme that starts with obtaining a random sampling of points. Based on previous works [85–89], a total of 5000 points have been obtained with GIS tools (QGIS) in the stable zone (absence of landslides) and 5000 points in each of the differentiated landslide typologies (presence). A table was then created by extracting the values of the different factors' layers at each point or pixel. Therefore, continuous (for quantitative factors) or discrete (for qualitative factors) values are used. The integrated tables of landslides' presence/absence were introduced in R studio statistical software [75] with the aim of obtaining susceptibility models and maps (LSM).

Linear discriminant analysis is one of the multivariate statistical models most used in slope instability or landslide susceptibility. These models assume that the factors that caused landslides in a given area are the same ones that will cause landslides in the future. The general linear models take the form [4,90]:

$$L = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + ... + B_nX_n + \varepsilon \quad (2)$$

where L is the presence/absence or area percentage of landslides in each mapping unit; X's are input predictor variables or factors in each mapping unit; B's are coefficients estimated from the data through statistical techniques and ε represents the model error.

In discriminant analysis, the probability or susceptibility of landslides in a given area (a pixel in our case) is calculated by adjusting the linear discriminant function to data inputs and then minimizing the model error [38,39]. These data inputs are, on one hand, the presence/absence of landslide in the area and, on the other hand, the values of determinant factors considered in the same area (pixels in our case).

Random Forest (RF) is a nonlinear supervised method used for data classification and regression. It is considered an ensemble method consisting of a combination of

deep decision trees so that each tree depends on values taken from a vector sampled randomly for growing [40,46,91,92]. Each decision tree grows splitting the input data (in our case, factors values) recursively so that each division contains more or less homogeneous states of the target variable (in our case, landslide susceptibility) [93]. In RF, each tree is trained on a subset of the data set and returns a result (in our case, landslide presence/absence). Therefore, the result of each decision tree is considered a vote, and thus, the final result is the one with the most votes or, in our case, the highest probability of landslide occurrence [91,94]. Some relevant RF characteristics are its predictive accuracy, low tendency to overfitting, relatively low computational cost and its ability to work with high dimensional data [40,92,95].

An artificial neural network (ANN) is a set of interconnected nodes or neurons useful for modeling problems with a complex relationship between analysis factors, so it is ideal for dynamic and nonlinear phenomena such as landslide occurrence [41,42]. The ANN architecture consists of a set of inputs (determinant factors); a set of intermediate layers (hidden layers) that perform the processing and an output layer with the prediction result [96]. Neural networks generally refer to supervised classification algorithms, which compare a given output with a predicted output, adapting the necessary parameters based on this comparison [97]. There are several neural network algorithms such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs); however, one of the most widely used is multilayer perceptron (MLP), which has been applied in several studies [41,42,98,99]. A perceptron is an individual neuron that allows us to classify a set of inputs into one or two categories by means of a step function, which returns 1 if the weighted sum of inputs exceeds a threshold or otherwise returns 0 [96]:

$$\begin{aligned} z &= b + \sum w_i x_i \\ y &= \begin{cases} 1 \; if \; z \geq 0 \\ 0 \; if \; z < 0 \end{cases} \end{aligned} \tag{3}$$

where y is the label or output variable (to predict); x_i is the feature or input variable; w_i and b are the weights and the bias, both parameters that the model has to learn during the training process.

Another important feature is the activation function, which allows an ANN to work with nonlinear problems [96] and can be of linear, sigmoid or logistic types, hyperbolic tangent or rectified linear unit (ReLU).

The MLP algorithm consists of a set of perceptrons organized in layers connected by synapses that are assigned a weight. Connection weights, hidden layers and the output layer were initialized and then updated using the backpropagation algorithm [42]. In our case, the MLP implemented had only one hidden layer with 3 or 4 neurons.

In practice, modeling using discriminant analysis methods [4,38,39,90] and random forest [40,46,91,92,94] involved partitioning the sample into training (80%) and validation (20%) sub-samples, giving a 80/20 ratio as some works recommend [100], although other proportions have been tested, such as 70/30 and 60/40. Next, a k-folds procedure (5 folds) was applied with the training sub-sample in order to fit the models while avoiding skewed partitions [101], in turn with an 80/20 ratio in training/testing. This procedure allowed us to obtain the corresponding susceptibility models and maps (LSM) and their validation through the area under the Receiver Operating Characteristic curve (AUC-ROC) [102]. In the case of random forest, model refinement methods are applied based on hyperparameter control, mainly number of trees (ntree, nodesize and tuneGrid) [40,91,95].

In the neural network models [41,42,46,52,95–99,103–107], the same landslide presence/absence samples as in the previous methods were used, from which a partition as also made into training (80%) and validation (20%) sub-samples. With the training sample, a one-layer ANN with 3 or 4 hidden neurons was adjusted using the rprop algorithm for backpropagation [95]. The R script returned the corresponding LSM, and the AUC-ROC value was estimated with the validation sample.

Every method described allows the elaboration of LSM for each typology independently, but from the results of factor analysis, synthesis maps have been made, grouping the types of landslides that present similar conditions. In this way, an integrated LSM was obtained for avalanches and debris flows and another for slides, earth flows and creeping processes.

2.3.4. Models Validation

A key aspect in the use of all these methods and models is validation, which enables them to be used as predictive models for estimating hazard and proposing prevention and mitigation measures. The susceptibility models developed can be validated through random, spatial and temporal partitioning of the inventory [108]. In this study, a random partition validation was mainly used, based on obtaining training and validation samples [33]. As mentioned before, in LDA, RF and ANN methods, a sample of 2000 points (20%) was used to validate the models fitted with the training sample (80%). However, to validate the results of the matrix method, which is based on a different scheme, a new random sampling of 2000 points (1000 in each of the landslide typologies and 1000 in the stable area) was performed in QGIS. In all cases, the susceptibility values were extracted from LSM by means of QGIS tools. The value tables were imported into R studio [75], where the theoretical values (susceptibility) were compared with the actual values (presence/absence) to calculate the AUC-ROC values.

AUC-ROC values were calculated from the ROC curves, which were built representing some values derived from the confusion matrix [109] such as True Positive Rate (TPR) or sensitivity in Y axis, versus the False Positive Rate (FPR) or 1—specificity in X-axis for different thresholds of the predicted values (for instance, for intervals of 0.1). The expressions for *TPR* and *FPR* are:

$$\begin{aligned} TPR\ (sensitivity) &= \frac{TP}{TP+FN} \\ FPR = \frac{FP}{FP+TN} &= 1 - specificity \end{aligned} \tag{4}$$

where: *TP* are the true positives; *TN* are the true negatives; *FP* are the false negatives and *FN* are the false negatives.

In addition to the validation with the AUC-ROC values, derived from the confusion matrix, another independent validation method has been applied. The degree of fit (DF), calculated in previous works from slope units [39] or landslide polygon areas [35,36,84] has been adapted to random point samples. Thus, the LSM obtained with the procedure described before were classified into five levels by means of the quantile method. Then, the additional point sample obtained for validation of the LSM of the matrix method (1000 points in stable area and 1000 points in each landslide typology) were used to extract the susceptibility levels in them (very low, low, moderate, high, very high). The DF of each susceptibility level was calculated as:

$$DF_i = \frac{pl_i/pt_i}{\sum pl_i/pt_i} \tag{5}$$

where pl_i is the number of points in landslide areas in each susceptibility level and pt_i is the total number of points in each susceptibility level (approximately 200 points). Sum of pl_i is 1000, and sum of pt_i is 2000.

The sum of DF in very low to low susceptibility levels was considered the relative error of LSM, while the sum of DF in high to very high levels was the relative success [36].

Finally, a temporal validation was conducted, obtaining the training samples from the landslides catalogued as latent and relict (4000 points in each typology), while the validation samples were extracted from the active landslides (1000 points in each typology). Additionally, from the 5000-point sample obtained for the previous validation strategy, 4000 points were added to complete the training sample, and 1000 points were added to the validation sample, giving again a training/validation ratio of 80/20. The next steps are

the same as in the previous strategy, thus calculating the AUC-ROC values. This validation has been applied only for LDA, RF and ANN methods.

3. Results

3.1. Landslide Inventory

The landslide inventory of Villavicencio-Bogotá (Figure 5) shows a total of 2506 landslides, representing 8.13% of the study area (Table 4). Five basic typologies have been differentiated according to Varnes [77] and Hungr et al. [110]: avalanches or collapses; debris flows; slides and earth flows, often as complex movements but classified according to the dominant type and soil creep, with a reduced speed and slow activity over time.

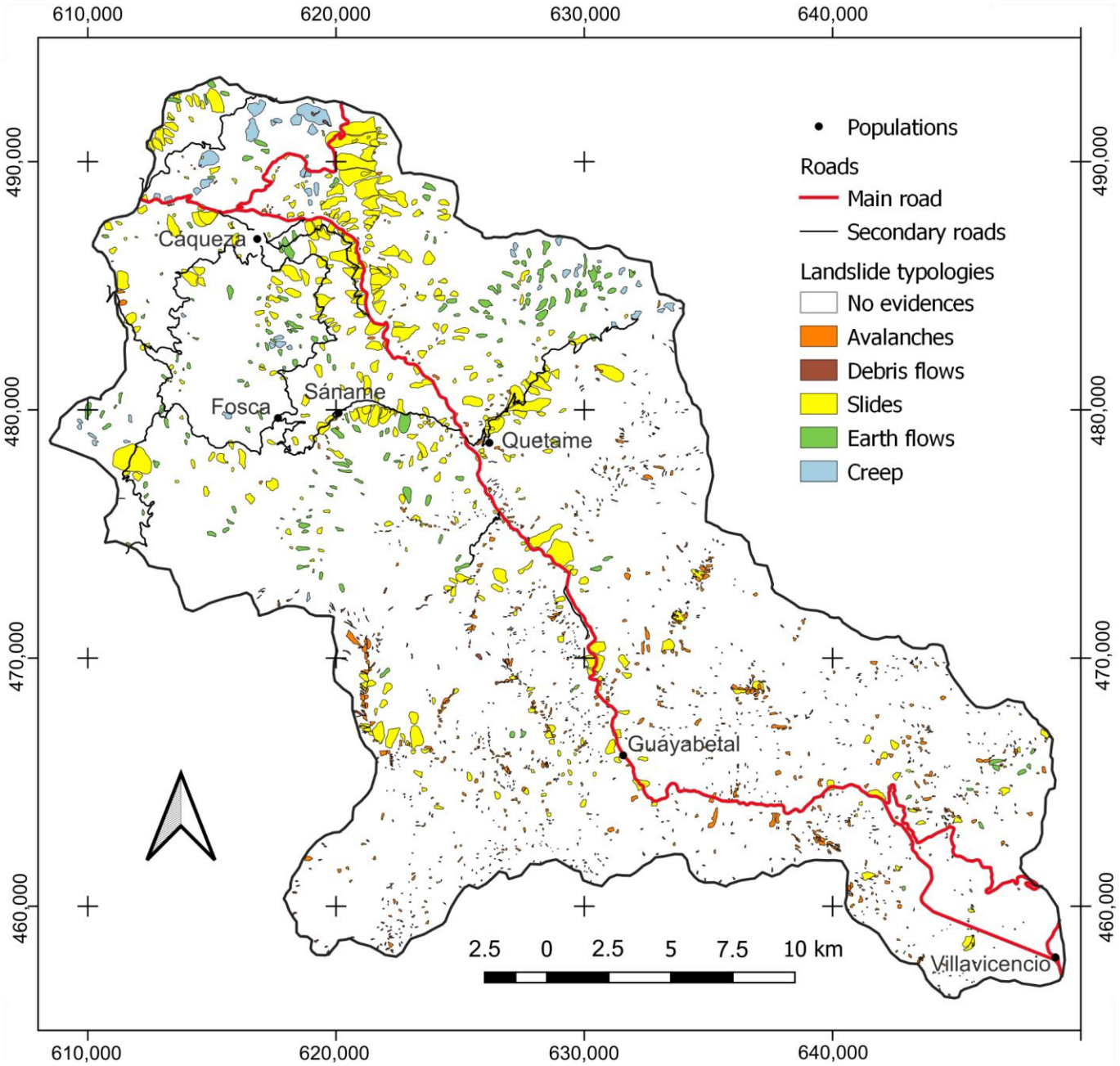

Figure 5. Landslide inventory in the study area. Five landslide typologies are distinguished.

Table 4. Analysis of landslide inventory. Areas are in m^2.

Typologies	Number		Total Area		Area Ind. (m^2)	Active		Latent		Relict	
	No	%	A (m^2)	%		No	%	No	%	No	%
Avalanches	979	39	7.95	13	8123	760	78	199	20	20	2
Debris flows	866	35	2.29	4	2649	594	69	271	31	1	0
Slides	437	17	39.00	64	89,261	59	14	122	28	256	59
Earth flows	179	7	7.47	12	41,747	4	2	70	39	105	59
Creep	45	2	3.99	7	88,803	6	13	36	80	3	7
All landslides	2506	-	60.71	8.13 [1]	24,231	1423	57	698	28	385	15

[1] This percentage corresponds to the area occupied by all landslides regarding the total study area.

The most predominant movements are avalanches (39%) and debris flows (35%), followed by slides (17%), earth flows (7%) and creep processes (2%). However, in terms of area, landslides occupy the largest extent (64%), with the percentages of avalanches and debris flows decreasing to 13% and 4%, respectively and the percentages of earth flows and creep increasing to 12% and 7%, respectively. These data are consistent with the fact that the average size (area) of slides and creeping processes is almost 90,000 m^2, whereas that of earth flows is approximately 42,000 m^2 and avalanches and debris flows have average areas of only 8100 and 2650 m^2, respectively.

From the activity perspective, most landslides are catalogued as active (57%), while remaining are considered latent (28%) and relict (15%). However, this distribution is different by type (Table 3), with avalanches and debris flows being predominantly active (78% and 69%, respectively), whereas slides and earth flows are mostly relict (59%) and creeping processes are latent (80%).

3.2. Analysis of Determinant Factors

The factors used in this study (shown in Figure 3) were analyzed; thus, their distribution in the area and their correlation with landslides are shown in Table 5 and Figure 6.

Table 5. Distribution of factor classes and cross-correlation with the landslides (all the landslides and for landslide typologies). The distribution of factor classes is expressed in % of area of every class respect to the total area of the study area. The density of landslides is expressed in % of landslide area for each class respect to the total area of this class. The correlation is expressed as the Kolmogorov-Smirnov (K–S) coefficient. The predominant classes of each factor, the classes with the highest density of landslides and K–S coefficients considered as significant are shown in bold.

Factors	Classes	All Landslides	Avalanches	Debris Flows	Slides	Earth Flows	Creep
Elevation (m)							
500–1000	5.90%	4.00%	1.90%	0.11%	1.61%	0.37%	0.00%
1000–1500	15.53%	11.51%	**3.76%**	0.72%	6.73%	0.30%	0.00%
1500–1800	16.08%	**15.40%**	1.61%	0.50%	**12.05%**	0.88%	0.37%
1800–2000	11.86%	10.35%	0.97%	0.41%	7.55%	1.07%	0.34%
2000–2400	**23.10%**	8.85%	0.80%	0.46%	4.58%	1.71%	**1.29%**
2400–2600	9.75%	6.98%	0.60%	0.48%	3.17%	**2.30%**	0.44%
2600–2800	6.83%	5.85%	0.62%	0.64%	1.75%	1.87%	0.97%
2800–3600	10.96%	2.87%	0.29%	**0.91%**	0.32%	0.73%	0.62%
K–S		**0.18**	**0.32**	0.14	**0.29**	**0.25**	**0.34**
Slope (°)							
0–5	1.94%	6.96%	0.62%	0.29%	3.46%	0.76%	**1.83%**
5–10	6.38%	7.53%	0.58%	0.15%	4.00%	1.26%	**1.53%**
10–20	22.69%	9.02%	0.79%	0.21%	5.12%	**1.73%**	1.17%

Table 5. *Cont.*

Factors	Classes	All Landslides	Avalanches	Debris Flows	Slides	Earth Flows	Creep
20–30	**31.08%**	9.91%	1.22%	0.42%	**6.34%**	1.46%	0.47%
30–45	**32.62%**	8.89%	1.87%	0.86%	5.43%	0.64%	0.09%
45–90	5.29%	9.93%	**3.24%**	**1.37%**	5.08%	0.23%	0.00%
K–S		0.03	**0.19**	**0.27**	0.05	**0.20**	**0.38**
Aspect							
N	11.25%	7.72%	1.34%	0.42%	4.77%	0.98%	0.21%
NE	11.88%	9.11%	1.49%	0.58%	5.71%	0.92%	0.40%
E	12.13%	10.37%	1.50%	0.80%	6.40%	1.08%	0.58%
SE	14.15%	9.47%	1.51%	0.83%	5.41%	0.99%	0.73%
S	13.23%	9.07%	1.52%	0.51%	4.33%	1.47%	**1.24%**
SW	13.17%	10.11%	1.50%	0.44%	5.85%	1.55%	0.77%
W	12.36%	9.37%	1.20%	0.37%	6.46%	1.06%	0.28%
NW	11.82%	7.90%	1.02%	0.37%	5.06%	1.21%	0.24%
K–S		0.04	0.05	0.14	0.06	0.08	**0.24**
Curvature							
−1–−0.02	6.26%	10.45%	**3.20%**	0.98%	5.30%	0.77%	0.21%
−0.02–−0.01	18.06%	10.27%	1.62%	0.60%	6.10%	1.34%	0.62%
−0.01–0.01	51.22%	9.25%	1.16%	0.47%	5.65%	1.29%	0.68%
0.1–0.2	18.28%	8.04%	1.09%	0.47%	4.97%	0.99%	0.52%
0.02–1	6.19%	7.30%	1.67%	0.76%	4.17%	0.51%	0.19%
K–S		0.04	0.11	0.07	0.03	0.06	0.06
TPI							
−100–−6	16.94%	12.29%	**3.24%**	0.82%	6.65%	1.22%	0.37%
−6–−2.5	16.52%	11.41%	1.47%	0.57%	6.87%	**1.70%**	**0.80%**
−2.5–0	16.47%	9.75%	1.01%	0.47%	5.98%	1.47%	**0.81%**
0–2.5	16.02%	8.58%	0.87%	0.44%	5.28%	1.25%	0.75%
2.5–6	16.29%	7.29%	0.88%	0.46%	4.58%	0.90%	0.47%
6–100	17.76%	5.83%	0.84%	0.51%	3.69%	0.49%	0.29%
K–S		0.12	**0.24**	0.09	0.10	**0.15**	0.12
TRI							
0–2	16.02%	7.93%	0.58%	0.17%	4.31%	1.36%	**1.52%**
2–3	16.68%	9.36%	0.87%	0.23%	5.40%	**1.83%**	1.03%
3–4	18.12%	9.96%	1.14%	0.36%	6.34%	1.57%	0.54%
4–6	17.04%	9.72%	1.45%	0.57%	6.34%	1.10%	0.27%
5–6	13.65%	8.90%	1.74%	0.76%	5.59%	0.70%	0.10%
>6	18.50%	8.99%	**2.49%**	**1.15%**	4.94%	0.38%	0.03%
K–S		0.03	**0.19**	**0.27**	0.06	**0.19**	**0.38**
Lithology							
Phylites-Schists	**38.43%**	5.77%	1.85%	0.89%	2.90%	0.14%	0.00%
Quartzarenites	**13.62%**	5.62%	**2.84%**	**1.27%**	1.28%	0.22%	0.01%
Shales	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lutites	**45.48%**	**13.32%**	0.56%	0.06%	**9.07%**	**2.37%**	**1.26%**
Conglomerates	0.89%	4.22%	0.50%	0.20%	3.53%	0.00%	0.00%
Volcanic	0.13%	4.77%	0.85%	0.00%	1.92%	2.00%	0.00%
Alluvial fans	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Alluvial deposit	0.93%	4.41%	1.88%	0.00%	2.53%	0.00%	0.00%
Terraces	0.42%	6.69%	2.86%	0.44%	3.39%	0.00%	0.00%
K–S		**0.23**	**0.28**	**0.42**	**0.31**	**0.48**	**0.55**

Table 5. *Cont.*

Factors	Classes	All Landslides	Avalanches	Debris Flows	Slides	Earth Flows	Creep
Precipitation (mm)							
500–1000	2.43%	**20.56%**	0.38%	0.06%	**15.44%**	1.74%	**2.93%**
1000–1500	12.31%	9.80%	0.29%	0.01%	7.58%	1.63%	0.29%
1500–2000	7.74%	9.10%	1.51%	0.60%	3.58%	**3.07%**	0.35%
2000–2500	**32.25%**	7.91%	**2.07%**	0.85%	4.45%	0.38%	0.17%
2500–3000	14.37%	3.47%	**1.88%**	0.52%	0.90%	0.17%	0.00%
3000–4000	13.50%	3.40%	1.39%	0.86%	0.95%	0.20%	0.00%
4000–5000	17.40%	2.64%	1.12%	**1.12%**	0.40%	0.00%	0.00%
K–S		**0.21**	**0.22**	**0.27**	**0.33**	**0.42**	**0.59**
Land cover							
Urban	5.00%	18.68%	7.06%	1.09%	9.58%	0.53%	0.43%
No vegetation	1.94%	17.16%	**5.70%**	**1.12%**	8.46%	0.58%	**1.30%**
Grass	9.83%	16.15%	1.70%	0.50%	**11.41%**	**1.21%**	**1.33%**
Bush-Shrubs	62.01%	10.20%	1.07%	0.48%	6.27%	**1.68%**	0.71%
Forest	21.03%	2.40%	0.39%	0.33%	1.38%	0.27%	0.03%
Water	0.20%	15.44%	8.45%	0.58%	6.32%	0.00%	0.09%
K–S		**0.17**	**0.29**	0.09	**0.17**	**0.17**	**0.20**
NDVI							
−0.5–0.1	5.08%	4.40%	1.81%	**1.09%**	1.48%	0.01%	0.02%
0.1–0.25	19.68%	2.40%	0.39%	0.34%	1.38%	0.26%	0.02%
0.25–0.4	**58.66%**	10.06%	1.11%	0.49%	6.14%	**1.64%**	0.69%
0.4–0.6	11.07%	**16.61%**	2.47%	0.60%	**11.23%**	1.13%	**1.19%**
0.6–1	5.52%	13.24%	**5.35%**	**1.25%**	5.51%	0.46%	0.66%
K–S		**0.19**	**0.25**	0.08	**0.20**	**0.21**	**0.24**
Distance to roads (m)							
0–100	4.91%	**17.53%**	1.84%	0.02%	**13.88%**	1.17%	0.62%
100–250	6.49%	15.93%	1.31%	0.11%	12.38%	1.34%	0.79%
250–500	9.08%	13.50%	1.12%	0.22%	10.57%	0.83%	0.76%
500–1000	14.66%	11.81%	0.92%	0.30%	8.28%	1.07%	**1.24%**
>1000	64.85%	6.61%	1.50%	**0.73%**	2.79%	1.21%	0.38%
K–S		**0.20**	0.05	**0.22**	**0.34**	0.03	**0.23**
Distance to rivers (m)							
0–100	3.95%	17.43%	**6.33%**	0.15%	10.57%	0.23%	0.15%
100–250	5.85%	17.44%	2.86%	0.46%	**13.45%**	0.45%	0.21%
250–500	9.00%	15.25%	1.93%	0.67%	11.55%	0.98%	0.11%
500–1000	16.32%	11.54%	1.31%	0.52%	7.98%	1.56%	0.18%
>1000	64.88%	6.47%	0.89%	0.57%	3.00%	**1.21%**	**0.80%**
K–S		**0.21**	**0.24**	0.04	**0.31**	0.08	**0.21**

The factors and their correlation with landslides are described below:

Elevation shows a wide range between 500 and 3600 m, although the majority (68%) is well distributed between 1500 and 2800 m. Landslides have their highest density in the range of 1500 to 1800 m. By typologies, avalanches are mainly concentrated in the range of 1000 to 1500 m, debris flows above 2800 m, slides between 1500 and 1800 m, earth flows between 2400 and 2600 m and creeping processes between 2000 and 2400 m. They show a significant correlation in practically all the typologies, especially in avalanches and creeping processes (more than 0.3).

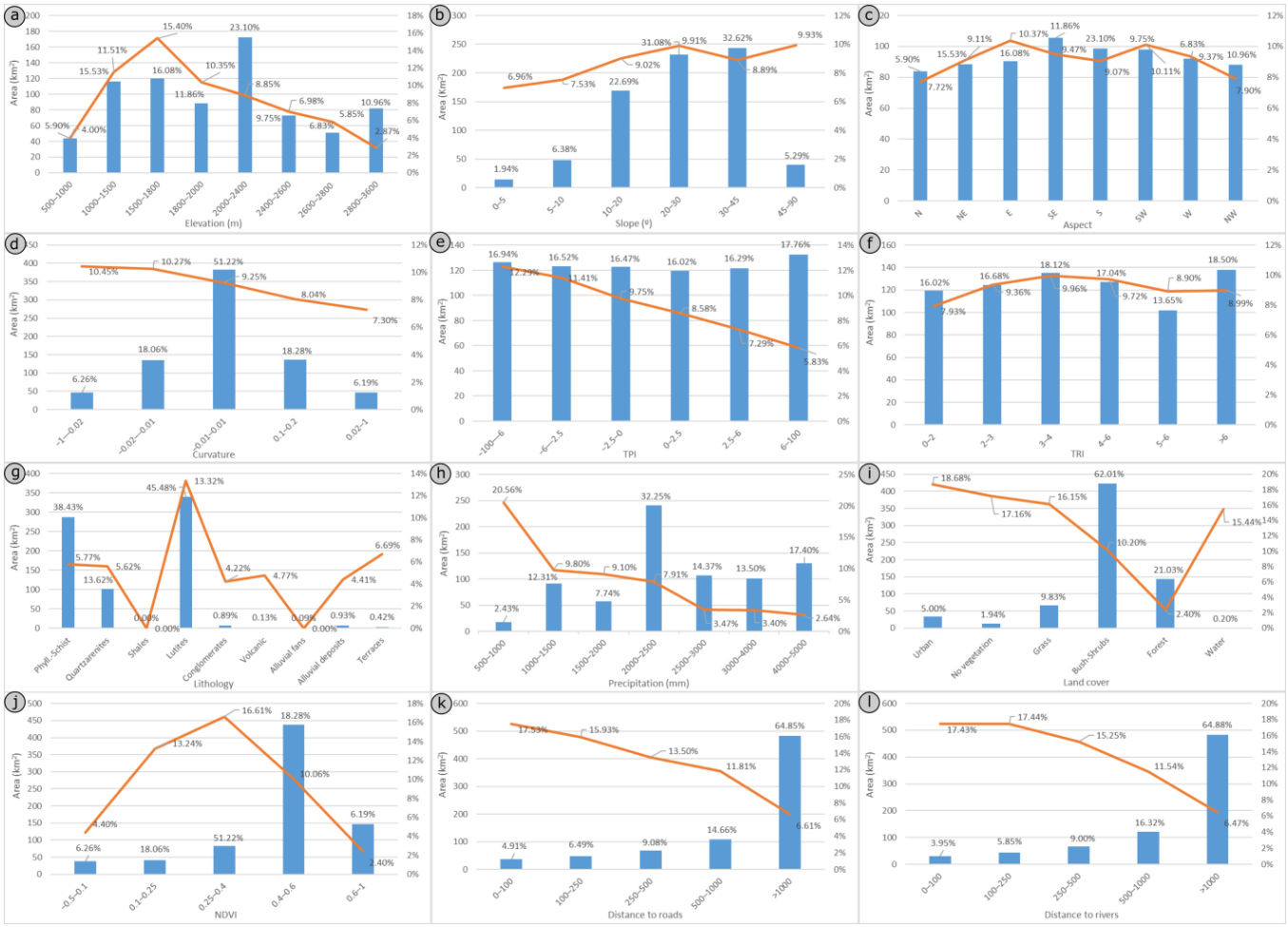

Figure 6. Distribution of classes and landslide density by class in each of the factors considered: (**a**): Elevation; (**b**): Slope; (**c**): Aspect; (**d**): Curvature; (**e**): TPI; (**f**): TRI; (**g**): Lithology; (**h**): Precipitation; (**i**): Land Cover; (**j**): NDVI; (**k**): Distance to roads; (**l**): Distance to rivers. Percentage area of the different classes and intervals are shown as histogram bars in blue; Landslide density in each class is shown as line diagram in red.

Slope is distributed practically throughout the total range from 0 to 90°, but more than 80% is between 10 and 45°. The landslide density is equally well distributed in the different ranges (between 7 and 10%), but by typologies, the distribution is different. Thus, avalanches and debris flows have a higher density as the slope increases, while slides reach their highest density in the range of 20 to 30°, earth flows between 10 and 20° and creeping processes between 0 and 10°. Practically all typologies, except slides, show a significant correlation.

Aspect appears to be fairly well distributed in all main orientations (11–14%). Meanwhile, the density is equally similar for all the movements (8–10%) and even in the analysis by typologies. Only the creeping processes have a higher density on south-facing slopes, so the correlations are generally low in all typologies, except for these processes.

Curvature presents a normal distribution concentrated around values close to 0. However, the distribution of landslides is higher in negative values (concave shapes) than in positive values (convex shapes). This is more evident in some typologies such as avalanches and slides, while in creeping processes, the higher density occurs in values close to 0. Correlations are not significant in any case.

Topographic Position Index (TPI): The classification allows for a balanced distribution (16–17%) of all intervals. However, the landslides' overall density is higher in negative values (10–12%) compared to positive values (6–8%). This asymmetry is very clear in

avalanches, which concentrate in the lower sections of the hillslopes with negative index, and not so much in slides, earth flows and creeping processes that move towards the lower-middle and middle sections of the hillslopes. The correlation is significant in avalanches, landslides and creeping processes.

Terrain Roughness Index (TRI): As in the previous case, the classification produces a fairly uniform distribution of the different classes (14–18%). In this case, the landslide density is relatively uniform in the different classes, although the analysis by types does show important differences. Thus, avalanches and debris flows mainly concentrate in high roughness classes, while landslides and especially creeping processes do so in low roughness classes. The correlation is significant in all types, except for slides.

Lithology: Of the nine differentiated classes, three of them, such as lutites (45%), phyllites and schists (38%) and quartz sandstones (14%), have a significant extension, while the remaining classes barely reach 1%, including quaternary materials. Taking this into account, a higher density is observed in Cretaceous lutites in all the landslide types; but especially, a clear difference is observed between the distribution of avalanches and debris flows with higher density in phyllites and quartz sandstones compared to slides, earth flows and creeping processes, with higher density in Cretaceous lutites and conglomerates. Avalanches and slides also involve quaternary materials (alluvial and colluvial). Meanwhile, correlations are significant in all types and even for the landslides as a whole.

Precipitation has a wide range from 500 to nearly 5000 mm of average annual precipitation, with the interval of 2000 to 2500 m having the greatest extension (32%). Unlike that which might be expected, the landslide density decreases in the intervals of higher precipitation for all the movements, but certain differences are observed between the different typologies. Thus, avalanches and debris flows have a higher density in areas of medium or high precipitation, while slides, earth flows and creep processes have a higher density in areas of lower precipitation. In all cases, correlations are significant.

Land cover: Of the six differentiated classes, there is a predominance of shrub areas (62%) over-forested areas (21%), grasslands and crops (10%), areas with scarce vegetation (2%), urban areas (5%) and water (0.2%). Excluding water and urban areas, which are occasionally affected by movements, the highest landslide density occurs in areas with scarce vegetation and grass/crop areas (19% and 17%, respectively) compared to shrub areas (10%) and forested areas (2.4%). There are certain differences by typologies, as avalanches and debris flows have a higher density in areas with scarce vegetation, while slides, earth flows and creep processes occur mainly in grass/crop areas and even shrub areas. In all cases, forest areas produce the lowest densities. The certain incidence of avalanches and slides in urban or water areas is noteworthy. The correlations are also significant in all cases.

NDVI: The most extended class is the 0.4–0.6 with almost 60%, followed by the 0.6–1 with almost 20% and the 0.25–0.4 with 10%. In terms of landslide density, the highest is achieved in the middle values of the index, decreasing towards the low and high values. By typologies, avalanches and debris flows present higher densities in the low values than in the high ones, except in the very low values generally associated with urban areas; meanwhile, slides, earth flows and creeping processes have the highest densities in the middle values of the index. The correlations are significant in practically all cases.

The distance to roads logically shows an increasing distribution of the area as the distance increases, with almost 65% of the surface area being more than 1 km from the main roads. The landslide density, however, is higher at shorter distances than at longer ones. By typologies, avalanches and earth flows barely present a relationship with roads, while in debris flows and creeping processes, the density increases with distance; only in slides is there an increase in density at shorter distances. The correlations are only significant in slides, debris flows and creeping processes.

Distance to rivers: As in the case of roads, it shows an increasing distribution of the area with distance, and almost 65% of the surface is more than 1 km away from the riverbeds. Similarly, the landslide density increases at shorter distances compared to

longer ones, although this does not happen in the same way for all typologies. Thus, avalanches and slides have higher density at shorter distances, while creep processes have higher density at longer distances, and debris flows and earth flows do not present a clear relationship. Correlations are only significant in avalanches, slides, and creeping processes.

Meanwhile, the results of multicollinearity analysis among the factors are shown in Table 6. As can be seen, strong correlations are only found between slope and TRI (0.97), and land use and NDVI (0.75); moderate correlations appear between TPI and curvature (0.65) and lithology and precipitation (0.59). Finally, weak correlations appear between elevation and distance to roads and distance to rivers and between lithology and slope and roughness.

Table 6. Correlation coefficients (Pearson) between factors. Strong and moderate correlations are shown in bold.

Factors	Elevation	Slope	Aspect	Curvat.	TPI	TRI	Lithol.	Precip.	Land C.	NDVI	D. Roads	D. Rivers
Elevation	1.000											
Slope	0.014	1.000										
Aspect	0.001	0.022	1.000									
Curvature	0.034	0.005	0.000	1.000								
TPI	0.090	0.009	0.000	**0.654**	1.000							
TRI	0.010	**0.971**	0.023	0.003	0.006	1.000						
Lithology	0.006	0.355	0.045	0.003	0.007	0.326	1.000					
Precipitation	0.248	0.231	0.029	0.000	0.001	0.210	**0.585**	1.000				
Land cover	0.114	0.073	0.140	0.000	0.004	0.058	0.022	0.215	1.000			
NDVI	0.107	0.041	0.175	0.005	0.007	0.026	0.043	0.150	**0.754**	1.000		
D.Roads	0.380	0.179	0.026	0.003	0.013	0.169	0.470	0.401	0.044	0.014	1.000	
D.Rivers	0.471	0.086	0.012	0.011	0.036	0.071	0.029	0.119	0.021	0.006	0.185	1.000

Based on these analyses, a factor selection has been made for the elaboration of susceptibility models and maps (Table 7). At the first level, all 12 factors used were considered. At the second level, those factors that showed a clear collinearity (strong and moderate correlation) were discarded, retaining those eight factors considered as independent. At the third level, only those non-collinear factors that showed significant correlation with the different landslide typologies were considered. Finally, the four factors that show total independence between them but correlation with most typologies were maintained: elevation, slope, TPI and lithology.

Table 7. Selected factors for susceptibility models: 1: All factors; 2: Non-collinear factors; 3: Factors with significant correlation with the different landslide typologies; 4: Independent factors between them but correlated with most typologies: Elevation, slope, TPI and lithology factors.

Factors	Elevation	Slope	Aspect	Curvat.	TPI	TRI	Lithol.	Precip.	Land C.	NDVI	D. Roads	D. Rivers
Avalanches	1234	1234	12	1	1234	1	1234	1	123	1	12	123
Debris flows	1234	1234	123	1	124	1	1234	1	123	1	12	12
Slides	1234	124	12	1	124	1	1234	1	123	1	123	123
Earth flows	1234	1234	12	1	1234	1	1234	1	123	1	12	12
Creep	1234	1234	123	1	124	1	1234	1	123	1	12	12

In summary, the conditions under which landslides occur preferentially are lithology of lutites; anthropic land use, areas with scarce vegetation and grass-crop lands, with NDVI between 0.2 and 0.4; elevation range between 1500 and 2000 m; precipitation between 500 and 1000 mm and the lower-concave section of the hillslopes.

If the analysis is considered by typologies:

- Avalanches show a higher density in Paleozoic quartz sandstones and phyllites, areas with scarce vegetation and NDVI between 0.1 and 0.25, altitudes between 1000 and 1500 m, slopes greater than 30°, the lower-concave sections of the hillslopes, areas with high roughness and areas near streams.
- Debris flows occur mainly in phyllites and quartz sandstones in areas with scarce vegetation, elevations above 2800 m, slopes greater than 30°, areas facing the east and southeast and areas with high roughness.
- Slides occur more frequently in Cretaceous lutites and grass-crop areas with NDVI between 0.25 and 0.4, elevations between 1500 and 1800 m, the middle-lower sections of the hillslopes and areas near streams and roads.
- Earth flows are concentrated mainly in lutites in areas with shrub vegetation with NDVI between 0.4 and 0.6, elevations between 2400 and 2800 m, slopes between 10 and 20° and the middle-lower sections of the hillslopes with low roughness.
- Creeping processes occur in lutites and grass-crop areas with NDVI between 0.25 and 0.4, elevations between 2000 and 2400 m, slopes of 0 to 10° and areas facing south with low roughness.

As can be observed, there is a certain similarity between the conditions for the occurrence of avalanches and debris flows on the one hand, and slides, earth flows and creep processes on the other hand. Thus, the first group is mainly associated with the lithology of phyllites, schists and quartz sandstones, areas with scarce vegetation, slopes greater than 30° and lower sections of the hillslopes with high roughness. Meanwhile, the second group is associated with Cretaceous lutites, grass-crop areas with NDVI between 0.25 and 0.4 and slopes of 0 to 20° in middle-lower sections of hillslopes with low roughness.

3.3. Susceptibility Models and Validation

The results of the susceptibility models and maps (LSM) performed with training samples of 8000 points (80%), using various methods and sets of factors, are shown in Figure 7. Meanwhile, Table 8 shows the results of the AUC-ROC obtained with sample validation (20%), Table 9 the DF for the same validation and Table 10 the AUC-ROC values obtained with temporal validation.

It can be observed that in general, for all landslide typologies, techniques and numbers of input factors, the models were well fitted, with the AUC-ROC for the validation samples always above 0.70. Regarding typologies, the creeping processes presented AUC-ROC values generally higher than 0.90; the avalanches and earth flows also reached quite high values (average above 0.84), while slides and debris flows had the lowest values, although still high (average above 0.80).

Regarding methods, the matrix method provides very high fits in general (average close to 0.90), followed by RF (0.88), ANN (0.84) and finally LDA (0.82). Generally, starting from the models obtained with the four basic factors (elevation, slope, TPI and lithology), all statistical and machine learning methods underwent an improvement when one or two factors were introduced that had some correlation with the landslides and low collinearity (reaching AUC-ROC from approximately 0.80 to 0.84). Then, they moderated their growth when non-collinear factors that did not show a high correlation with landslides (AUC-ROC up to 0.86) were introduced and even stabilized when all factors were introduced, including those that showed collinearity (0.87). The matrix method also improved when factors correlated with landslides were introduced (AUC-ROC from 0.80 to 0.90), stabilized with non-correlated factors (0.90) and increased again when all factors, even those showing collinearity, were considered (0.98).

An analysis carried out with other training/validation ratios (70/30 and 60/40) showed similar results. Thus, creeping processes presented the maximum AUC-ROC values (average of 0.93 in both cases), and debris flows presented the minimum values (0.80–0.81). Meanwhile, excluding matrix methods in which these ratios are not considered, RF presented the highest AUC-ROC values (0.87–0.88) and LDA the lowest (0.82).

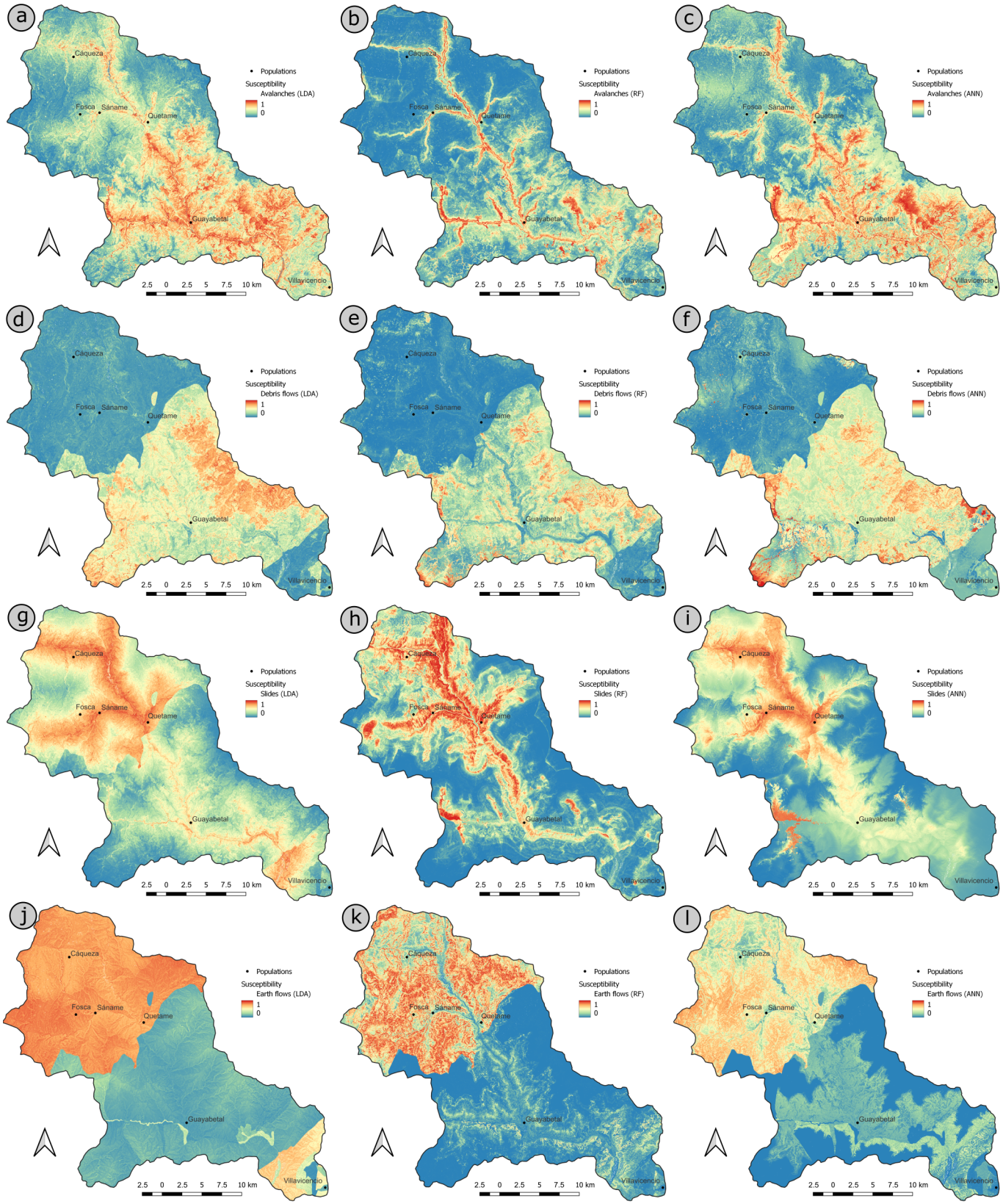

Figure 7. *Cont.*

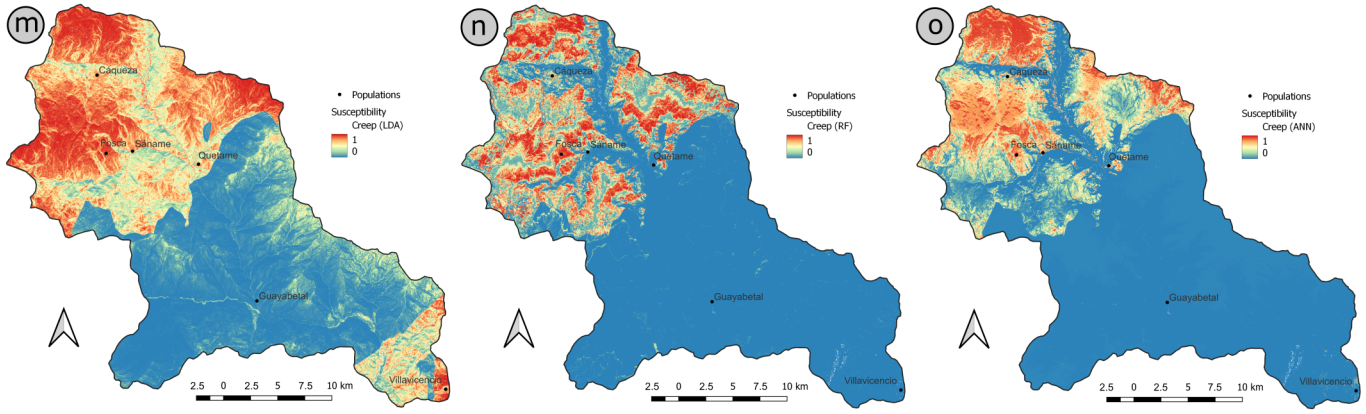

Figure 7. Landslide susceptibility models: Avalanches (**a**): LDA; (**b**): RF; (**c**): ANN; Debris flows: (**d**): LDA; (**e**): RF; (**f**): ANN; Slides: (**g**): LDA; (**h**): RF; (**i**): ANN; Earth flows: (**j**): LDA; (**k**): RF; (**l**): ANN; Creep: (**m**): LDA; (**n**): RF; (**o**): ANN. Color scale from blue-green (lower levels of susceptibility) to orange-red (higher levels of susceptibility.

Table 8. AUC-ROC values of LSM in sample validation for the different typologies, methods and number of factors.

Methods	N Factors	Avalanches	Debris fl.	Slides	Earth Flows	Creep	All mov.
Matrix	4 f	0.804	0.739	0.835	0.842	0.884	0.825
	5–6 f	0.908	0.897	0.881	0.911	0.942	0.908
	8 f	0.914	0.861	0.904	0.924	0.952	0.910
	12 f	0.977	0.978	0.977	0.984	0.987	0.982
LDA	4 f	0.790	0.750	0.733	0.805	0.875	0.791
	5–6 f	0.832	0.781	0.780	0.805	0.901	0.820
	8 f	0.834	0.791	0.782	0.808	0.919	0.827
	12 f	0.848	0.794	0.807	0.815	0.932	0.839
RF	4 f	0.773	0.727	0.776	0.817	0.908	0.800
	5–6 f	0.874	0.870	0.874	0.830	0.916	0.873
	8 f	0.894	0.882	0.891	0.915	0.984	0.913
	12 f	0.885	0.873	0.897	0.939	0.988	0.916
ANN	4 f	0.800	0.715	0.802	0.810	0.867	0.796
	5–6 f	0.857	0.806	0.814	0.822	0.909	0.842
	8 f	0.857	0.799	0.818	0.833	0.934	0.848
	12 f	0.844	0.819	0.841	0.877	0.940	0.845
Average		0.856	0.818	0.838	0.862	0.927	0.860
Methods	Matrix	0.901	0.869	0.899	0.915	0.941	0.906
	LDA	0.826	0.779	0.776	0.808	0.907	0.819
	RF	0.857	0.838	0.860	0.875	0.949	0.876
	ANN	0.840	0.785	0.819	0.844	0.903	0.838
N. Factors	4 f	0.792	0.733	0.787	0.821	0.884	0.803
	5–6 f	0.868	0.839	0.837	0.842	0.917	0.861
	8 f	0.875	0.833	0.849	0.870	0.947	0.875
	12 f	0.889	0.866	0.881	0.904	0.969	0.902

Table 9. Degree of Fit of lower/higher susceptibility levels for the different typologies, methods and number of factors.

Methods	N Factors	Avalanches	Debris fl.	Slides	Earth Flows	Creep	All mov.
Matrix	4 f	5/79	11/58	6/88	3/95	2/98	5/84
	5–6 f	2/96	3/96	1/92	1/99	1/98	2/96
	8 f	3/95	5/87	1/94	1/97	1/97	2/94
	12 f	2/97	1/99	1/99	1/99	1/99	1/99
LDA	4 f	5/82	4/72	9/78	7/93	1/99	5/85
	5–6 f	6/86	4/81	8/84	7/93	1/99	5/89
	8 f	5/88	4/81	7/84	6/93	1/99	4/89
	12 f	4/90	4/81	5/85	6/93	1/99	4/ 90
RF	4 f	4/82	5/74	4/87	1/95	1/99	3/88
	5–6 f	2/93	2/90	1/98	1/97	1/99	1/95
	8 f	1/96	1/95	1/98	1/99	1/99	1/98
	12 f	2/95	1/94	1/97	1/99	0/100	1/97
ANN	4 f	4/82	4/79	5/87	2/93	1/99	3/87
	5–6 f	4/89	3/86	3/92	1/94	1/99	3/92
	8 f	4/89	4/86	4/88	1/94	1/99	3/91
	12 f	4/90	4/88	4/89	1/95	1/99	3/92
Average		4/89	4/84	4/90	3/96	1/99	3/92
Methods	Matrix	3/92	5/85	2/93	1/98	1/97	2/93
	LDA	5/86	4/79	7/83	6/93	1/99	3/89
	RF	2/91	2/88	1/95	1/98	1/99	2/94
	ANN	4/86	4/84	4/89	1/94	1/99	3/91
N. Factors	4 f	5/81	6/71	6/85	4/94	1/98	4/86
	5–6 f	4/91	3/88	3/91	3/96	1/99	3/93
	8 f	3/92	4/87	3/91	2/96	1/99	2/93
	12 f	3/94	2/91	2/94	2/98	1/99	2/95

Table 10. AUC-ROC values of LSM in temporal validation for the different typologies, methods and number of factors.

Methods	N Factors	Avalanches	Debris fl.	Slides	Earth Flows	Creep	All mov.
LDA	4 f	0.803	0.794	0.724	0.781	0.857	0.792
	5–6 f	0.845	0.800	0.716	0.790	0.894	0.809
	8 f	0.845	0.811	0.724	0.784	0.900	0.813
	12 f	0.848	0.786	0.779	0.786	0.921	0.824
RF	4 f	0.745	0.699	0.687	0.755	0.868	0.751
	5–6 f	0.789	0.795	0.735	0.787	0.865	0.794
	8 f	0.794	0.724	0.743	0.806	0.913	0.796
	12 f	0.832	0.748	0.790	0.829	0.923	0.824
ANN	4 f	0.801	0.770	0.705	0.768	0.831	0.775
	5–6 f	0.819	0.793	0.726	0.764	0.847	0.790
	8 f	0.834	0.793	0.734	0.795	0.869	0.805
	12 f	0.926	0.785	0.785	0.808	0.926	0.846
Average		0.823	0.775	0.737	0.788	0.885	0.802
Methods	LDA	0.835	0.798	0.736	0.785	0.893	0.809
	RF	0.790	0.742	0.739	0.794	0.892	0.791
	ANN	0.845	0.785	0.737	0.784	0.868	0.804
N. Factors	4 f	0.783	0.754	0.705	0.768	0.852	0.773
	5–6 f	0.818	0.796	0.726	0.780	0.869	0.798
	8 f	0.824	0.776	0.734	0.795	0.894	0.805
	12 f	0.869	0.773	0.785	0.808	0.923	0.831

The validation made with degree of fit showed that the error/success ratio was very suitable in most cases, with an average of 3/92, with the creeping processes being those that presented the best ratio (1/99) and the debris flows the worst (4/84). By methods, the matrix method and RF presented slightly better average ratios (2/93 and 2/94) than ANN (3/91) and LDA (3/89). Regarding the number of factors included in the models, the aver-age ratio of the models with four factors was the worst (4/86), while the remaining ones reached 2–3/93–95.

The temporal validation shows acceptable results in general, with AUC-ROC values being mostly higher than 0.7 but between 0 and 12 points lower than in the sample validation. The average is 0.80, which is 4.5 points lower than in the sample validation. By typologies, the best results are obtained also in creeping processes (0.89) and the worst in slides (0.74), in which the AUC-ROC values decrease about eight points. By methods, all of them (LDA, RF and ANN) present similar AUC-ROC average values, about 0.80. Regarding the number of factors involved, the models present increasing AUC-ROC values, from 0.77 with 4 factors to 0.83 with 12 factors.

Finally, the integrated LSM of shallower processes (avalanches and debris flows) and deeper processes (slides, earth flows and creep), modeled with the random forest method and eight factors, are shown in Figure 8. In this case, the AUC-ROC values using sample validation and an 80/20 ratio are 0.88 and 0.83, respectively.

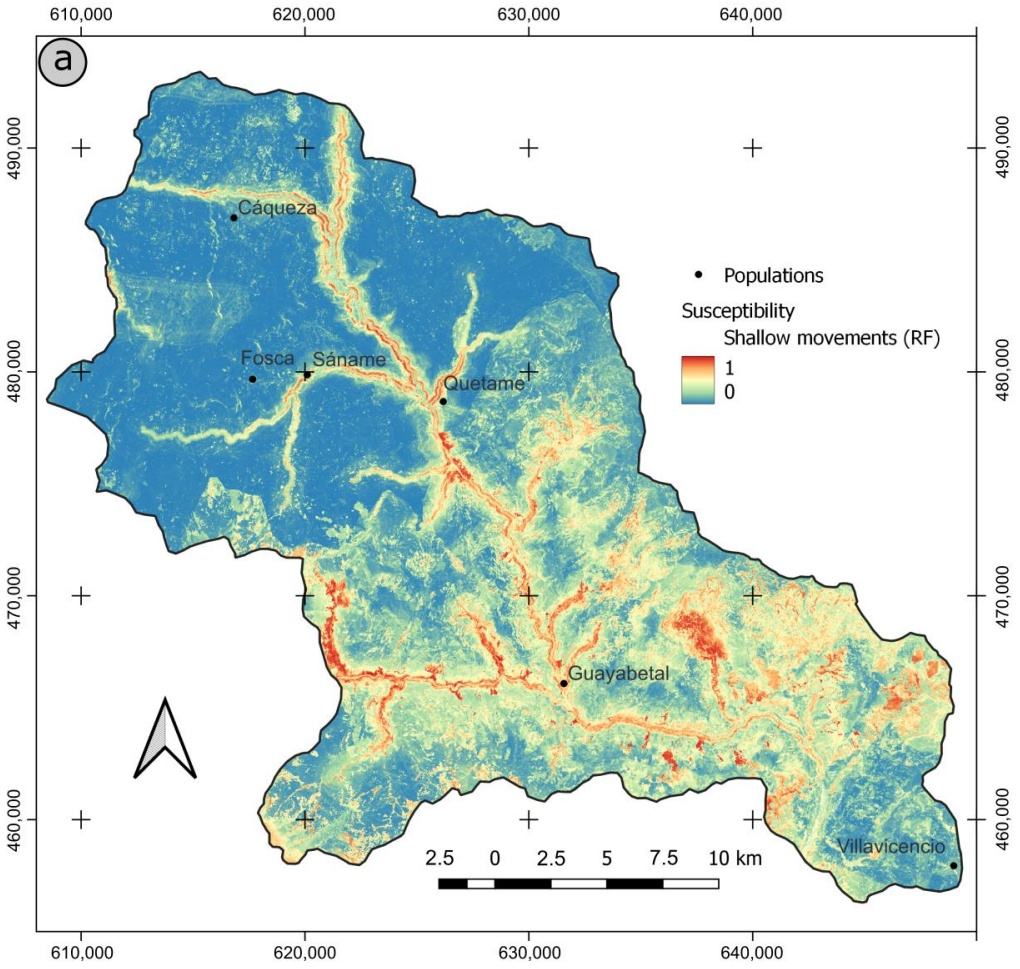

Figure 8. *Cont.*

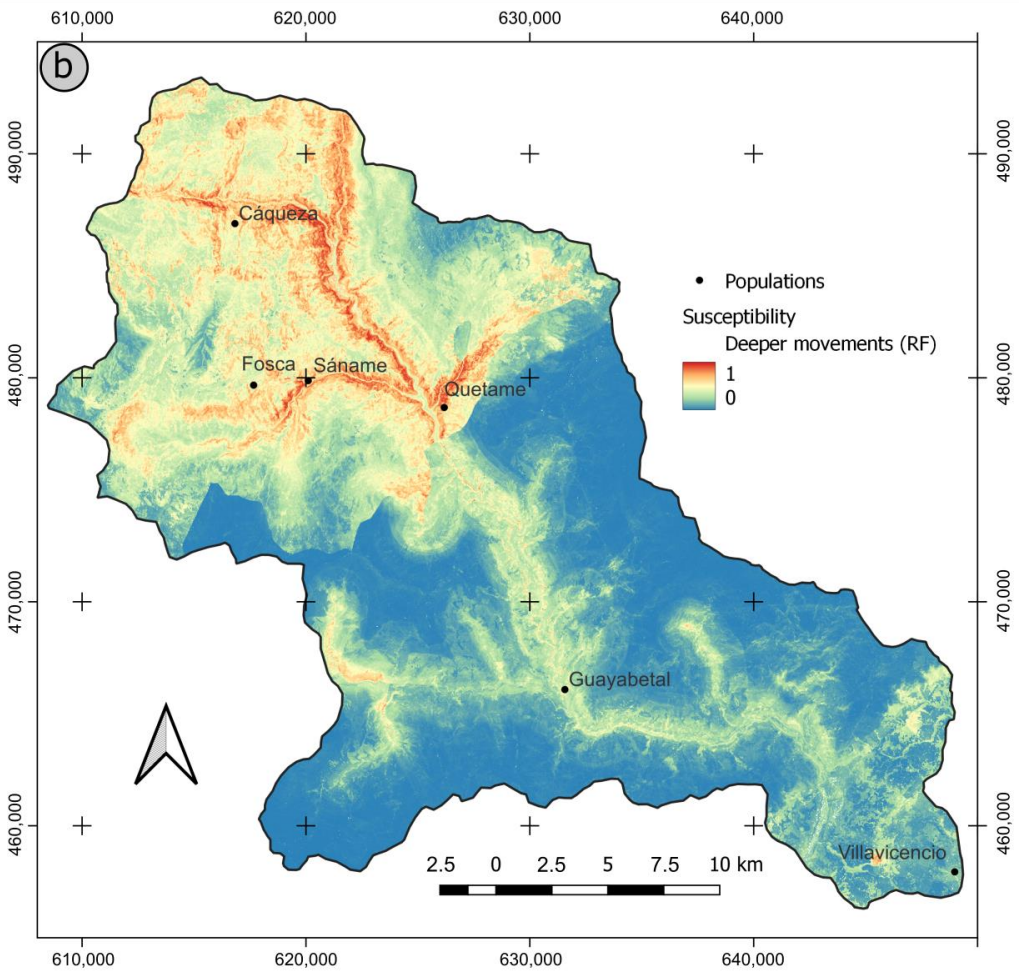

Figure 8. Integrated LSM: (**a**) Shallow movements; (**b**) Deeper movements. Color scale from blue-green (lower levels of susceptibility) to orange-red (higher levels of susceptibility).

4. Discussion

This study was conducted in an area of the central sector of the Eastern Andes mountain range in Colombia, which is characterized by intense landslide activity due to particular geological, topographical and climatic conditions. The geology of the area consists of Cretaceous sedimentary series, mostly composed of lutites and sandstones, structured by thrust faults and folds with a SSW-NNE direction, through which materials from the Paleozoic substrate with a certain degree of metamorphism outcrop [63,64]. The predominant elevation in the area ranges from 1000 to 3000 m, with relatively steep slopes (modal range between 20 and 30°), generally above 1000–1500 mm of precipitation, and land cover made up mainly of shrub-bush, grass-crop and forest areas.

4.1. Lanslides Inventory

A predominance of slide and earth flow type movements was observed when the total area or extension of landslides were considered, although avalanches, debris flows and creeping processes were also present. The individual areas were larger in slides and soil creep (almost 90,000 m^2), whereas that of earth flows was approximately 42,000 m^2 and only 8100 and 2650 m^2, respectively, for avalanches and debris flows. Meanwhile, slides and earth flows occasionally had a complex character, although in this work, they were considered the dominant process, which generally depends on their greater or lesser evolution, respectively. On the other hand, the creeping processes corresponded to undifferentiated flows with slow movement in general. Avalanches, which in other terminologies may be called collapses [36,111], are frequently in transition with both rock falls and debris

slides-debris flows; however, considering the morphology and slopes observed (steep but not sub-vertical) and the materials in which they originate (phyllites, schists and quartzites often superficially weathered), they are classified as debris avalanches. Nevertheless, these avalanches can evolve into debris flows if the hillslope morphology allows it.

The inventory and differentiated typologies generally coincided with other studies of the Eastern Cordillera of the Colombian Andes. Thus, in the study by Calderón et al. [46], translational, rotational and wedge slides were differentiated in the vicinity of the Bogotá-Villavicencio highway. In the study by Valencia and Martínez-Graña [52], debris-flow, debris slides and rock falls were also inventoried in the Capitanejo area (Santander), further north in the Eastern Cordillera. Garcia-Delgado et al. [28] studied deep landslides and gravitational processes in San Eduardo, which is also to the North. Finally, Pradhan et al. [51] and Ramos-Cañon et al. [59] catalogued rock falls, avalanches, rotational and translational landslides, earth-mud flows and debris flows. In other parts of the country, such as the Western and Central Cordilleras, shallow slides, falls, debris flows and mud-flows have been identified [5,24–27,29,45,50,55]; in the southern Colombian Massif, specifically in Mocoa, debris flows, debris avalanches and shallow slides have been found [8,21,30,31].

Regarding the estimated activity based on photointerpretation, smaller movements such as avalanches, debris flows and small slides show higher activity than larger slides, earth flows and creeping processes, which have lower activity. This generally agrees with what happens in other regions of the world where these types of analyses have been addressed [111]. Activity generally depends on precipitation, which is abundant in the region due to the influence of deep convective systems [60] and which will probably have an even greater impact in the coming years [61,112]. This influence of precipitation has been analyzed by applying hydrological models [21,22,29,39] or determining rainfall thresholds [54–59,113]. The influence of other phenomena as triggering factors, such as earthquakes [8,114], active faults [115] or deforestation [116], has been also considered.

4.2. Analysis of Determinant Factors

Regarding factor analysis, the number of factors to be used in the models can be very high, especially in machine learning models [33], where it is common not to perform factor selection and allow the algorithms to fit the models. In this work, 12 variables have been used, which more or less coincide with those used in previous studies on susceptibility modelling, both globally [35,36,38,40–42,90,97,103,104,106,107] and in Colombian Andes [31,44–53]. Among them, the ones derived from the DEM stand out, which are related to the spatial distribution of important parameters such as slope, morphology, soil moisture or flow direction [51,117,118]; those related to geology and the geomechanical characteristics of materials [119] or those related to land use and land cover [81].

Despite what was said above, some authors recommend performing a certain factor selection to optimize the predictive capacity of the models and the performance of computational processes [120], generally based on the multicollinearity between factors or through methods of dimensionality reduction, which allows the selection of the most determinant factors and discarding other ones [121]. In addition, this analysis allows us to determine the conditions of the different landslide typologies [36,74,122]. The analysis carried out in this work shows that the factors that mainly condition the landslides' generation are elevation and lithology; although in the differentiated analysis by typologies, slope must be considered, and in some cases, TPI, aspect, land cover, distance to roads and distance to rivers should be considered as well. Some factors such as TRI, precipitation and NDVI also show some correlation with different landslides typologies but at the same time are strongly correlated to slope, TPI, lithology and land cover, respectively. These factors are similar to those found in previous works in other areas of the Colombian Andes, where a factor selection has been made. Thus, in Calderón et al.'s work [46], landslides were found to be related to the distance to faults, profile curvature, flow length, accumulated flow and land use. Salazar et al. [44] considered elevation, slope, planar curvature, landform shape, distance to faults, geological units, distance to rivers and distance to roads.

Correa et al. [50] considered slope, flow length, TWI, convergence index and soil types. Goyes-Peñafiel et al. [45] considered slope, curvature, TWI, landform shape, geological units and land use.

The conditions in which landslides occur preferably are the lithology of lutites, which is usually lower-resistance material and therefore more prone to landslides; areas with scarce vegetation and grass-crop lands, with NDVI between 0.2 and 0.4, that is, areas where there is a scarce or null vegetation cover protecting the soil from erosion and weathering processes; elevation between 1500 and 2000 m, where various geological, soil and morphological conditions favourable to instability are concentrated; rainfall between 500 and 1000 mm, which has no justification but is a consequence of the correlation of this factor with lithology and is therefore discarded in the analysis and, finally, lower-concave sections of the hillslopes, where hydrological and erosive phenomena promote the landslide generation. These conditions are similar to those found in other parts of the world and specifically in the Colombian Andes. Thus, in Salazar et al. [44] and Valencia and Martínez-Graña [52], landslides were associated with steep slopes (above 25–30°), southern orientation, concave morphologies, high roughness and a certain proximity to rivers but not so much to roads. However, they differed in altitude intervals and especially in the most affected lithology, which in the case of Valencia and Martínez-Graña [52] in the Eastern Cordillera were lutites or shales similar to those in the study area; and in Salazar et al. [44] in the Western Cordillera, with a different geological environment, they were volcanic rocks. Meanwhile, in Grima et al.'s work [116], landslides occurred in a proportion of six times more in deforested areas than in forested areas, while in Renza et al.'s work [53], no clear relationship was found between NDVI and other vegetation indices and landslides.

One of the interesting aspects of this work is the typology-based factor analysis, which has allowed us to observe differences in the conditions of occurrence of the different landslide typologies. From the results, a dichotomy was observed in the conditions in which landslides occur, with a first group of shallower landslides corresponding to avalanches and debris flows and a second group of larger and usually deeper landsides corresponding to slides, earth flows and creep processes. These conditions are shown in the Figure 9 for every determinant factor. Lithology appears as a crucial factor that in turn influences other factors such as slope, elevation or land cover, among others.

Thus, since lithology is the most determinant factor, a higher resolution of the geological map is needed, which allows greater precision in the identification of the conditions in which landslides and their corresponding typologies originate. The same can be stated about other factors such as the land cover, NDVI and even DEM derivatives.

Despite this, it can be observed that avalanches and debris flows present similar occurrence conditions that are different from the remaining typologies. As previously noted, there can be a certain transition between avalanches and debris flows, such that the former may eventually evolve into the latter if the morphological conditions allow for it. Thus, the conditions that lead to a higher landslide density are mainly the lithology of Paleozoic quartz sandstones and phyllites, which are rocks more resistant a priori, enabling the formation of quite steep slopes where these processes originate [36]. Therefore, the slopes on which they occur are generally higher than 30° and even up to 45°, with a high terrain roughness. However, avalanches occur in the lower-concave sections of the hillslopes and close to the channels, which is not the case with debris flows, which extend to higher sections, as the elevation analysis shows (1000–1500 m and over 2800 m, respectively). Furthermore, both types are associated with areas of low vegetation cover and low NDVI values (0.10–0.25).

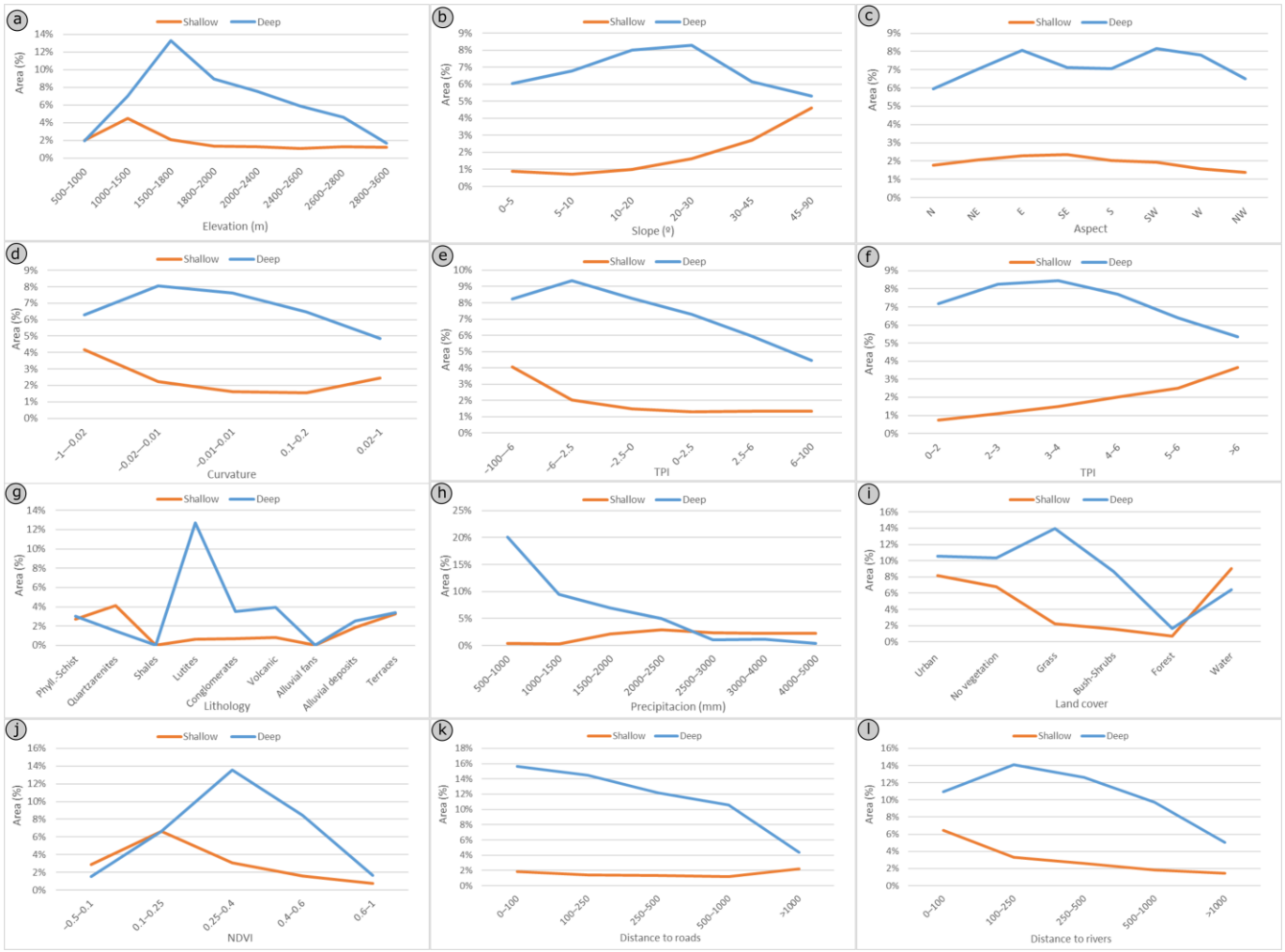

Figure 9. Conditions in which the shallower (orange line) and deeper landslides (blue line) occur. X axis: Classes of the different factors; Y axis: % of area occupied for landslides in each class. (**a**): Elevation; (**b**): Slope; (**c**): Aspect; (**d**): Curvature; (**e**): TPI; (**f**): TRI; (**g**): Lithology; (**h**): Precipitation; (**i**): Land Cover; (**j**): NDVI; (**k**): Distance to roads; (**l**): Distance to rivers.

Meanwhile, slides, earth flows and creeping processes occur more frequently in Cretaceous lutites, which are rocks with less resistance than the previous ones. Hence, landslides can occur even at lower slopes than with the previous types. Specifically, slides occur at a wide range of slopes but more frequently between 20 and 30°, with earth flows between 10 and 20° and creeping processes on slopes below 10°. Additionally, they often occur in the middle and lower sections of the slopes, with low roughness. Regarding land cover and NDVI, they are more concentrated in grass-crop and shrub areas, with medium NDVI values (0.25–0.6). However, the elevations range from lower for slides to medium-high for earth flows and creeping processes.

4.3. Susceptibility Models and Validation

Regarding susceptibility models and maps (LSM), four methods have been applied according to the groups established in Reichenbach et al. [33]: matrix method (index), discriminant analysis (multivariate statistics), random forest and an artificial neural network (machine learning). In general, all methods showed good results, with AUC-ROC values above 0.70 for the different typologies and number of factors considered in the models (Table 8, Figure 10a). For typologies, creeping processes generally had AUC-ROC values above 0.90, which shows that these processes are associated with very specific conditions

that make the models very fitted. In fact, in this case, there are methods such as RF that have higher ROC AUC values (0.95) than even the matrix method (0.94). Next are avalanches and earth flows with also quite high values (average around 0.86), reaching maximum values in the matrix method (0.90–0.91). Finally, slides and debris flows present the lowest AUC-ROC average values (0.84 and 0.82, respectively), since the conditions for the landslide occurrence are not so clearly defined in these typologies, reaching maximum values also in the matrix method (0.88–0.90).

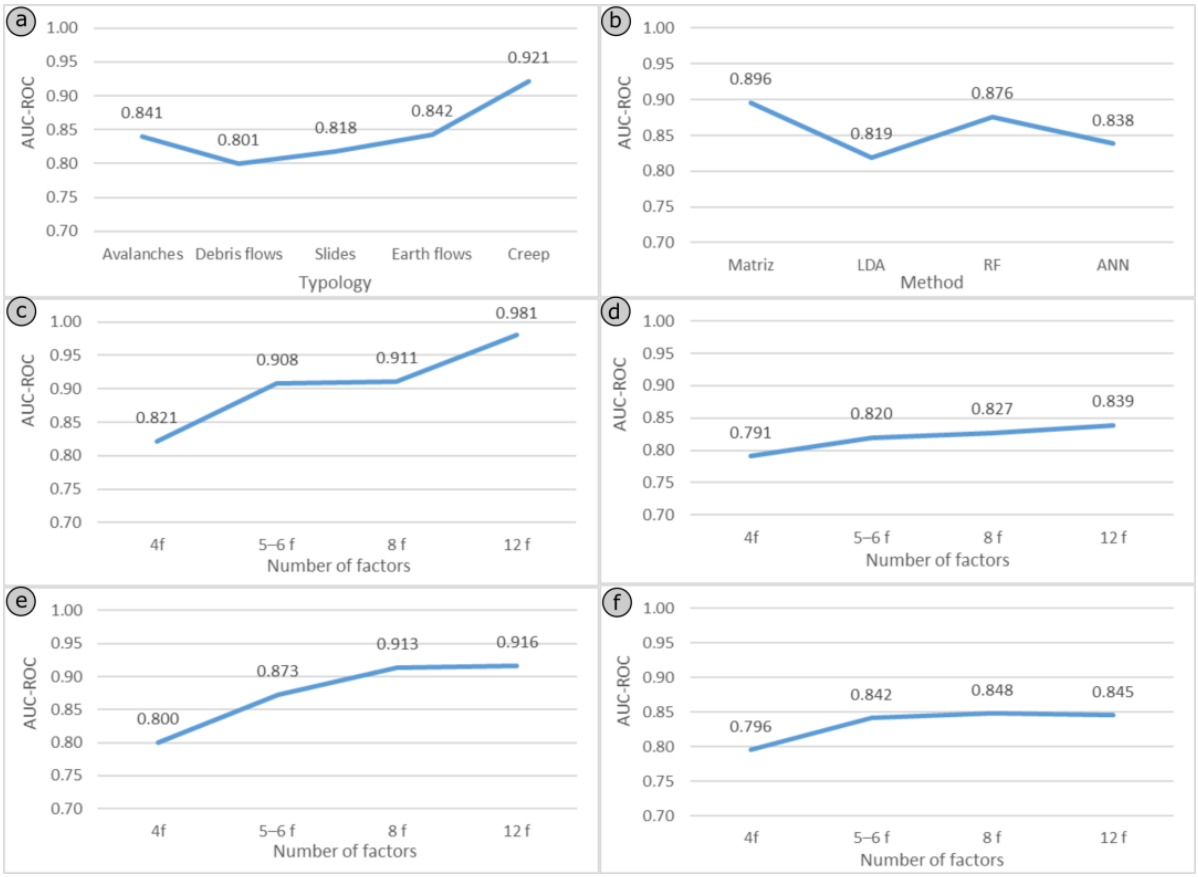

Figure 10. AUC-ROC values: (**a**): By typologies; (**b**): By methods; By number of factors (4 to 12) in the different methods (**c**): Matrix; (**d**): LDA; (**e**): RF; (**f**): ANN.

Linking to the above, the comparison by methods shows that the matrix method provides very high fits in general (average close to 0.90), followed by RF (0.88), ANN (0.84), and finally LDA (about 0.82) (Figure 8b). As can be seen, the matrix method elaborated from unique condition units provides very good fits even with randomly selected testing samples. This agrees with the results obtained in previous works, where other validation techniques such as temporal validation (performed with inventories elaborated after the one used in the model) produce degree of fit [39] of 5–10% in the classes of very low to low susceptibility (errors) and 70–80% in the classes of high to very high susceptibility (success) [35,36,83]. Despite that, matrix and other index-based or bivariate statistical methods have some limitations related to the simplification of the conditioning factors (especially when they are classified as in matrix approach) and the assumption of conditional independence between them [123]. Thus, they fit very well to the specific conditions of an area when inventories are exhaustive, but their performance is reduced when they are transferred to other areas [124] or when the starting inventories are less exhaustive.

The LDA models showed relatively lower fits, although they can be considered acceptable, as they reached average values of AUC-ROC around 0.80 for most landslide typologies and even 0.90 for creeping processes. This is corroborated by previous studies

where 70–80% of the slopes were correctly classified in high susceptibility classes [4,39]. However, multivariate statistical methods maintain some of the limitations of the bivariate methods, such as the simplification and dependence between factors [123]. Moreover, these limitations cannot be overcome due to the linearity of the discriminant functions compared to the greater versatility and better performance of machine learning methods in non-linear systems such as susceptibility models developed from factors of different nature [105,125].

The RF method produced excellent results, with AUC values between 0.84 and 0.88 in most landslide typologies and 0.95 in creeping processes, which agrees with other studies where values close to and above 0.90 were achieved [40,92,126–128]. In the studies carried out Colombian Andes, Calderón et al. [46] applied RF starting from 14 similar factors, with very good results (area under the success rate curve, A_{SRC}, of 93%).

Finally, the application of a perceptron-type ANN with a single hidden layer and three–four neurons provided fits that were somewhat lower to RF but higher to LDA. Thus, except for the lower AUC-ROC value for debris flows (0.79), the remaining typologies had values between 0.82 and 0.84, and even the creeping processes reached a value of 0.90. These values are consistent with the results obtained by numerous authors who have applied neural networks of different types, from MLP to convolutional ones, passed through Radial Basis Funtion (RBF) networks and others. Focusing on perceptron-type networks, Pradham and Lee [41] obtained an AUC of 0.91–0.94; Tien Bui et al. [42], 0.92; Bravo et al. [95], 0.76; Zare et al. [98], 0.88; Pham et al. [99], 0.87; Park et al. [104], 0.81 and Aslam et al. [107], 0.87. In the Colombian Andes, Calderón et al. [46], in an area that encompasses that used in this article, applied an MLP-type ANN with 14 factors (input neurons) and two hidden layers of 16 neurons, which provided an A_{SRC} of 0.88. When factor selection was applied (five factors), the A_{SRC} became 0.86, and after applying factor reduction by means of PCA, the A_{SRC} decreased to 0.81. Meanwhile, Valencia and Martínez-Graña [52], further north in the Santander department, applied a perceptron-type ANN with backpropagation algorithm, starting from 14 factors and a layer of 20 neurons, which provides a very high AUC-ROC value (0.988).

Regarding the number of factors, although all methods show a better fit as the number of factors used in the models increases, not all do so in the same way, which allows us to extract considerations about the behavior of these models and the opportunity to make or not make factor selection (Figure 10c–f). Thus, starting from models with four basic factors common to all typologies and independent of each other (elevation, slope, TPI and lithology), statistical methods (LDA) and machine learning (RF and ANN) undergo a significant increase when introducing one or two factors that have a certain correlation with landslides and low collinearity. Then, the AUC-ROC value increases from approximately 0.80 to 0.845, mainly in RF. However, this growth is attenuated (0.86) when up to eight non-collinear factors are included, although some of them did not show a correlation with landslides, and it practically stabilizes when all factors are introduced, including those that show collinearity (0.87). This behavior, in which the value of the AUC-ROC on the validation sample stabilizes, may be evidence of noise when introducing factors that do not show an influence on landslides and of overfitting or overtraining [129] when introducing redundant factors. In this sense, it is important to mention that the strong correlation between factors also increases the probability of overfitting [130]. This shows the interest of the factor analysis that allows for factor selection.

These observations are partially corroborated by the behavior observed in the matrix method, which does not aim to develop a statistical or learning model but rather fit the susceptibility zoning to the data on factors and movements in a specific area. Thus, the value of AUC increases as correlated factors with landslides are introduced (from 0.80 to 0.90), which is consistent with what was observed in the previous methods, indicating the importance of including as many determinant factors of landslides as possible. Then, the AUC-ROC value stabilizes when introducing factors uncorrelated to landslides (0.90), which confirms that these factors do not improve the model and may introduce noise. However, unlike the other methods, the AUC-ROC value again increases when all factors

are introduced, even those that show collinearity (0.98), which could be a clear indication of overfitting in this method when a large number of factors are introduced. It should be noted that, despite using a random validation sample, this validation sample is not entirely independent of the total mobilized surface used to develop the susceptibility map, since the matrix method uses the entire landslide area to fit the model.

Regarding the other methods and strategies of validation, the analysis carried out with other training/validation ratios showed similar results for AUC-ROC values, which ensured that the ratio used (80/20) is valid for this case and in general [100]. Meanwhile, the degree of fit showed an error/success ratio that was very suitable in most cases. Relative errors were lower than 5% in practically every case, and relative successes reached values usually higher than 90% and even 95%. Moreover, the results by typologies, methods and number of factors involved were in agreement with those obtained with the AUC-ROC. Finally, the temporal validation showed acceptable results in general, with AUC-ROC values mostly higher than 0.7. However, these values were between 0 and 12 points lower than in the sample validation, being an average of 4.5 points lower. By typologies, methods and number of factors, similar tendencies were observed, with the models of creeping processes and the models with a greater number of factors presenting higher AUC-ROC values. However, there were some differences such as the lowest values observed in slides that probably are related to the different size and, thus, the conditions of the active slides (smaller, in areas of higher slopes and lower sections of hillslopes) and non-active (larger, in areas of moderate slopes and lower-middle sections of the hillslopes). The other typologies presented similar conditions between active and non-active landslides, so the AUC-ROC values obtained in temporal validation were higher. The other difference was the similar behavior between the methods analyzed (LDA, RF and ANN since this validation was not applied for the matrix approach). In this case, the global reduction of AUC-ROC values made the values of different methods less distinguishable.

In summary, we can conclude about the importance of choosing the more adequate method as well as the factor selection for LSM. Regarding the method, it seems that machine learning methods, especially random forest, show better performance than statistical methods due to their greater flexibility and non-linear adjustment. Regarding the matrix method, it has less statistical basis and produces significant overfitting, especially when there is no factor selection. Factor analysis and selection appear to be a recommended procedure to avoid overfitting and noise, even in machine learning methods where the importance of factor selection is not as apparent.

Finally, regarding the distribution of landslide susceptibility, a clear zoning can be observed in the maps of the different typologies whatever the method and the number of factors that have been used. Thus, the LSM corresponding to avalanches and debris flows show greater susceptibility in the south-eastern part (lower basin), which is also clearly observed in the integrated map of these typologies; meanwhile, the LSM of slides, earth flows and creep processes have a higher susceptibility in the north-western part (upper basin), as can also be observed in the integrated map. This zoning is the consequence of geological control, as mentioned before in the discussion of factor analysis, since metamorphic rocks outcrop in the lower basin through fold and thrust structures; these rocks are generally more coherent and resistant, developing steep slopes where avalanches and debris flows occur (affecting mainly the superficial layers of weathered rocks). Meanwhile, the Cretaceous shales outcrop predominantly in the upper part of the basin, and these sedimentary rocks, usually less resistant, are more susceptible to slides, earth flows and creeping processes. Logically, this general zoning is discriminated by the remaining factors, such as slopes, TPI, land cover and distance to roads in those maps that show better results (e.g., maps obtained with RF). Moreover, given the importance of the lithological factor, a higher resolution of the geological maps would have allowed a better discrimination of susceptibility.

5. Conclusions

The present study has allowed the characterization of landslides in a sector of almost 750 km^2 in the Eastern Cordillera of the Colombian Andes, by means of the elaboration of inventories using photointerpretation, GIS factors analysis and landslide susceptibility maps (LSM). A total of 2506 landslides were inventoried, occupying approximately 8% of the study area, including avalanches, debris flows, slides, earth flows and creeping processes. Debris flows (39%) and avalanches (35%) were the most abundant in number, while landslides (64%) occupied the largest area due to their larger individual size. Avalanches and debris flows were predominantly active, while most slides and earth flows were relict and creeping processes were latent.

The factors analysis showed that elevation, lithology and land cover were the factors that most influenced the generation of landslides. However, in the differentiated analysis by typologies, slope and, in some cases, TPI, the aspect, distance to roads and distance to rivers had to also be considered. Some factors, such as TRI, precipitation and NDVI, also showed some correlation with different landslide typologies but were strongly related to slope, TPI, lithology and land cover, respectively. This analysis also allowed us to observe differences in the conditions of occurrence of the different typologies. Thus, avalanches and debris flows had similar conditions of occurrence, which were the lithology of quartz-sandstone and phyllites from the Paleozoic era (more resistant rocks) and areas with scarce vegetation and low NDVI values (0.10–0.25); slopes higher than 30° and lower sections of the hillslopes with high roughness. Meanwhile, slides, earth flows and creeping processes occurred mainly in Cretaceous lutites and grass/crop areas; morphologically, they occurred on a wide range of slopes, generally lower than 30° in the middle and lower sections of the hillslopes with low roughness.

Regarding LSM, different types of methods were tested to provide consistent results in determining landslide hazards in the region, including index-based methods (matrix), multivariate statistical methods (discriminant analysis, LDA), machine learning methods (random forest, RF) and a perceptron-type neural network (ANN). In general, all methods produced good results, with the AUC-ROC values always above 0.7 and obtained with the validation sample (20%). For landslide typologies, the best fits occur in cases where the conditions are more specific, such as creeping processes (0.90) and debris flows and avalanches (0.84), and the worst fits occur where they are not so specific, such as slides and debris flows (0.82–0.80). By methods, although the matrix method provides very high fits (average close to 0.90), there is a certain tendency toward overfitting, especially when no factor selection is addressed. LDA offers relatively lower adjustments (average around 0.82), while RF and ANN present very good fits in general (average around 0.88 and 0.84, respectively). In all these methods, starting from four common and non-collinear factors with a high correlation with all typologies (elevation, slope, TPI and lithology), an increase in AUC-ROC occurs when one or two additional factors specific for each typology are introduced. Then, they moderate their increase when non-collinear factors with lower correlation with the movements are included (introducing noise), and especially when collinear factors are also considered (producing overfitting). But whatever the method used, the LSM maps show a zoning as a consequence of geological control, which produces higher susceptibility to avalanches and debris flow in the lower part of the basin, with more resistant rocks, while a higher susceptibility to slides, earth flows and creep occurs in the upper part of the basin where less resistant rocks appear.

The importance of the choice of susceptibility modeling method and factor selection is also checked in order to avoid noise and overfitting. This ensures the development of robust and coherent models that can be in neighboring areas of this region. Thus, the use of RF or other machine learning methods is recommended due to their greater versatility and better behavior in nonlinear systems such as LSM. Likewise, a certain factor analysis is also recommended that allows a better understanding of the conditions in which landslides occur and also an improvement in the factor selection. In this case, models with six or eight factors are considered to provide the most reliable results, and thus, synthesis

maps have been prepared for shallower or deeper movements, with AUC-ROC values of 0.87 and 0.83, respectively.

The main limitations of this work are related to the quality and resolution of input data, especially the determinant factors. Thus, although the resolution of DEM can be considered as sufficient, its precision as a DEM derived from PalSAR could be limited. Nevertheless, the most limiting aspect is the resolution of thematic factors, especially the geological map as the main determinant factor; in addition, a better quality and temporal signification of the land cover, even with derived land use, and vegetation indices is also required. Meanwhile, another limiting aspect is the estimation of landslide activity that can allow not only a refinement of susceptibility maps but also the elaboration of hazard maps.

Thus, for future improvements and work, the introduction of new determinant factors, especially the improvement of those used in this work, is proposed, both derived from the DEM and thematic maps, which allow the models to be refined. This would also allow the analysis of the influence of spatial and thematic resolution on the models. Thus, robust models are developed, which can be applied to other neighboring areas. Automated methods for factor selection and complex models for susceptibility maps can be tested, but always subject to control by the analysts. Finally, the estimation of temporal probability based on the realization of multi-temporal inventories and the introduction of triggering factors (mainly rainfall) will allow the creation of hazard or threat maps in the study area and other areas of the mountain range.

Author Contributions: Conceptualization, M.C.H.-C. and T.F.; methodology, M.C.H.-C., C.C. and T.F.; software, P.E.B.-L.; validation, M.C.H.-C., L.P.C., I.L.H.-P. and P.E.B.-L.; formal analysis, M.C.H.-C.; investigation, M.C.H.-C., L.P.C. and I.L.H.-P.; resources, M.C.H.-C., I.L.H.-P., P.E.B.-L., M.S.-G. and T.F.; data curation, M.C.H.-C. and P.E.B.-L.; writing—original draft preparation, M.C.H.-C., L.P.C., I.L.H.-P., P.E.B.-L. and T.F.; writing—review and editing, C.C., J.D., M.S.-G. and T.F.; supervision, C.C., M.S.-G. and J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to Photogrammetric and Topometric Systems Research Group of the University of Jaen (TEP-213); Geologic Processes and Resources Group of the University of Jaen (RNM-325); Natural Hazards Lab of the Centre for Advanced Studies in Earth Sciences, Energy and Environment of the University of Jaen (CEACTEMA); Department of Earth and Marine Sciences (DiSTeM) of the University of Palermo (Italia); Department of Geographic and Environmental Engineering, UDCA and the Institute for Studies of Sectional Regime of Ecuador (IERSE), University of Azuay. We also acknowledge the help of Professor Sergio David Parra González, from the Faculty of Agricultural Sciences and Natural Resources of the Universidad de los Llanos (Colombia).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schuster, R.L. Socioeconomic significance of landslides. In *Landslides: Investigation and Mitigation*; Turner, A.K., Schuster, R.L., Eds.; Transportation Research Board Special Report 247; National Academy of Sciences: Washington, DC, USA, 1996; pp. 12–35.
2. Petley, D.N. Global Patterns of Loss of Life from Landslides. *Geology* **2012**, *40*, 927–930. [CrossRef]
3. UNDRR. *Global Annual Report 2019*; UNDRR: Geneva, Switzerland, 2019; Available online: https://gar.undrr.org/report-2019 (accessed on 30 May 2023).
4. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [CrossRef]
5. Muñoz, E.; Poveda, G.; Ochoa, A.; Caballero, H. Multifractal Analysis of Spatial and Temporal Distributions of Landslides in Colombia. In *Advancing Culture of Living with Landslides*; Springer International Publishing: Cham, Switzerland, 2017; pp. 1073–1079. [CrossRef]
6. Aristizábal, E.; Sánchez, O. Spatial and Temporal Patterns and the Socioeconomic Impacts of Landslides in the Tropical and Mountainous Colombian Andes. *Disasters* **2020**, *44*, 596–618. [CrossRef]
7. Kühnl, M.; Sapena, M.; Wurm, M.; Geiß, C.; Taubenböck, H. Multitemporal Landslide Exposure and Vulnerability Assessment in Medellín, Colombia. *Nat. Hazards* **2022**, 1–24. [CrossRef]

8. García-Delgado, H.; Petley, D.N.; Bermúdez, M.A.; Sepúlveda, S.A. Fatal Landslides in Colombia (from Historical Times to 2020) and Their Socio-Economic Impacts. *Landslides* **2022**, *19*, 1689–1716. [CrossRef]
9. Gómez, D.; García, E.F.; Aristizábal, E. Spatial and Temporal Landslide Distributions Using Global and Open Landslide Databases. *Nat. Hazards* **2023**, *117*, 25–55. [CrossRef]
10. CRED: EM-DAT, The International Disaster DataBase, Centre for Research on the Epidemiology of Disasters (CRED). Available online: https://www.emdat.be/ (accessed on 30 May 2023).
11. UNDRR: DesInventar Sendai. Available online: https://db.desinventar.org/ (accessed on 30 May 2023).
12. NASA: Open Global Landslide Catalog. Available online: https://gpm.nasa.gov/landslides/index.html (accessed on 30 May 2023).
13. UN-SPIDER: Global Fatal Landslide Database (GFLD—University of Sheffield). United Nations Office for Outer Space Affairs UN-SPIDER Knowledge Portal. Available online: https://un-spider.org/links-and-resources/data-sources/global-fatal-landslide-database-gfld-university-sheffield (accessed on 30 May 2023).
14. Ojeda Moncayo, J.; Donnelly, L. Landslides in Colombia and Their Impact on Towns and Cities. *IAEG Pap.* **2006**, *112*, 1–13.
15. Servicio Geológico Colombiano: Sistema de Información de Movimientos en Masa—SIMMA. Available online: http://simma.sgc.gov.co/# (accessed on 30 May 2023).
16. CORPES. *Mapa de Amenazas Geológicas Por Remoción en Masa y Erosión del Departamento de Cundinamarca*; Ingeominas: Bogotá, Colombia, 1998.
17. Varnes, D.J. *Landslide Hazard Zonation: A Rewiew of Principles and Practise*; UNDRR: Geneva, Switzerland, 1984.
18. Isaza-Restrepo, P.A.; Martínez Carvajal, H.E.; Hidalgo Montoya, C.A. Methodology for Quantitative Landslide Risk Analysis in Residential Projects. *Habitat Int.* **2016**, *53*, 403–412. [CrossRef]
19. Smith, H.; Coupé, F.; Garcia-Ferrari, S.; Rivera, H.; Castro Mera, W.E. Toward Negotiated Mitigation of Landslide Risks in Informal Settlements: Reflections from a Pilot Experience in Medellín, Colombia. *Ecol. Soc.* **2020**, *25*, art19. [CrossRef]
20. Ayala-García, J.; Dall'Erba, S. The Impact of Preemptive Investment on Natural Disasters. *Pap. Reg. Sci.* **2022**, *101*, 1087–1103. [CrossRef]
21. García-Delgado, H.; Machuca, S.; Medina, E. Dynamic and Geomorphic Characterizations of the Mocoa Debris Flow (March 31, 2017, Putumayo Department, Southern Colombia). *Landslides* **2019**, *16*, 597–609. [CrossRef]
22. Pertuz-Paz, A.; Monsalve, G.; Loaiza-Úsuga, J.C.; Caballero-Acosta, J.H.; Agudelo-Vélez, L.I.; Sidle, R.C. Linking Soil Hydrology and Creep: A Northern Andes Case. *Geosciences* **2020**, *10*, 472. [CrossRef]
23. Soeters, R.; Van Westen, C.J. Slope instability recognition, analysis and zonation. In *Landslides, Investigation and Mitigation*; Turner, A.K., Schuster, R.L., Eds.; Transportation Research Board, National Research Council, Special Report 247; National Academy Press: Washington, DC, USA, 1996; pp. 129–177.
24. Palacio Cordoba, J.; Mergili, M.; Aristizábal, E. Probabilistic Landslide Susceptibility Analysis in Tropical Mountainous Terrain Using the Physically Based r.Slope.Stability Model. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 815–829. [CrossRef]
25. Hidalgo, C.A.; Vega, J.A. Probabilistic Landslide Risk Assessment in Water Supply Basins: La Liboriana River Basin (Salgar-Colombia). *Nat. Hazards* **2021**, *109*, 273–301. [CrossRef]
26. Marín, R.J.; Velásquez, M.F.; Sánchez, O. Applicability and Performance of Deterministic and Probabilistic Physically Based Landslide Modeling in a Data-Scarce Environment of the Colombian Andes. *J. S. Am. Earth Sci.* **2021**, *108*, 103175. [CrossRef]
27. Marín, R.J.; Mattos, Á.J.; Fernández-Escobar, C.J. Understanding the Sensitivity to the Soil Properties and Rainfall Conditions of Two Physically-Based Slope Stability Models. *Boletín Geol.* **2022**, *44*, 93–109. [CrossRef]
28. García-Delgado, H. The San Eduardo Landslide (Eastern Cordillera of Colombia): Reactivation of a Deep-Seated Gravitational Slope Deformation. *Landslides* **2020**, *17*, 1951–1964. [CrossRef]
29. Prada-Sarmiento, L.F.; Cabrera, M.A.; Camacho, R.; Estrada, N.; Ramos-Cañón, A.M. The Mocoa Event on March 31: Analysis of a Series of Mass Movements in a Tropical Environment of the Andean-Amazonian Piedmont. *Landslides* **2019**, *16*, 2459–2468. [CrossRef]
30. Cheng, D.; Cui, Y.; Su, F.; Jia, Y.; Choi, C.E. The Characteristics of the Mocoa Compound Disaster Event, Colombia. *Landslides* **2018**, *15*, 1223–1232. [CrossRef]
31. Vargas-Cuervo, G.; Rotigliano, E.; Conoscenti, C. Prediction of Debris-Avalanches and -Flows Triggered by a Tropical Storm by Using a Stochastic Approach: An Application to the Events Occurred in Mocoa (Colombia) on 1 April 2017. *Geomorphology* **2019**, *339*, 31–43. [CrossRef]
32. Braab, E.E. Innovative Approaches to Landslide Hazard and Risk Mapping. In Proceedings of the 4th International Symposium on Landslides, Toronto, ON, Canada, 16–21 September 1984; pp. 307–323.
33. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A Review of Statistically-Based Landslide Susceptibility Models. *Earth Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
34. Irigaray, C.; Fernández, T.; Chacón, J. Comparative Analysis of Methods for Landslide Susceptibility Mapping. In *Landslides, Proceedings of the 8th International Conference and Field Trip on Landslides, Granada, Spain, 27–28 September 1996*; CRC Press: Boca Raton, FL, USA; pp. 373–384.
35. Irigaray, C.; Fernández, T.; El Hamdouni, R.; Chacón, J. Verification of Landslide Susceptibility Mapping: A Case Study. *Earth Surf. Process. Landf.* **1999**, *24*, 537–544.
36. Fernández, T.; Irigaray, C.; El Hamdouni, R.; Chacón, J. Methodology for Landslide Susceptibility Mapping by Means of a GIS. Application to the Contraviesa Area (Granada, Spain). *Nat. Hazards* **2003**, *30*, 297–308. [CrossRef]

37. Chung, C.-J.F.; Fabbri, A.G. Probabilistic Prediction Models for Landslide Hazard Mapping. *Photogramm. Eng. Remote Sens.* **1999**, *65*, 1389–1399.
38. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic Landslide Hazard Assessment at the Basin Scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]
39. Baeza, C.; Corominas, J. Assessment of Shallow Landslide Susceptibility by Means of Multivariate Statistical Techniques. *Earth Surf. Process. Landf.* **2001**, *26*, 1251–1263. [CrossRef]
40. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine Learning Methods for Landslide Susceptibility Studies: A Comparative Overview of Algorithm Performance. *Earth Sci. Rev.* **2020**, *207*, 103225. [CrossRef]
41. Pradhan, B.; Lee, S. Regional Landslide Susceptibility Analysis Using Back-Propagation Neural Network Model at Cameron Highland, Malaysia. *Landslides* **2010**, *7*, 13–30. [CrossRef]
42. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. *Landslides* **2016**, *13*, 361–378. [CrossRef]
43. Pourghasemi, H.R.; Rahmati, O. Prediction of the Landslide Susceptibility: Which Algorithm, Which Precision? *Catena* **2018**, *162*, 177–192. [CrossRef]
44. Salazar Gutiérrez, L.F.; Menjivar Flores, J.C.; Martínez Carvajal, H.E. Susceptibility Factors of Drainage Basins to Shallow Landslides in Coffee-Growing Areas in the Department of Caldas, Colombia. *Environ. Earth Sci.* **2021**, *80*, 145. [CrossRef]
45. Goyes-Peñafiel, P.; Hernandez-Rojas, A. Landslide Susceptibility Index Based on the Integration of Logistic Regression and Weights of Evidence: A Case Study in Popayan, Colombia. *Eng. Geol.* **2021**, *280*, 105958. [CrossRef]
46. Calderón-Guevara, W.; Sánchez-Silva, M.; Nitescu, B.; Villarraga, D.F. Comparative Review of Data-Driven Landslide Susceptibility Models: Case Study in the Eastern Andes Mountain Range of Colombia. *Nat. Hazards* **2022**, *113*, 1105–1132. [CrossRef]
47. Aristizábal, E.; López, S.; Sánchez, O.; Vásquez, M.; Rincón, F.; Ruiz-Vásquez, D.; Restrepo, S.; Valencia, J.S. Evaluación de La Amenaza Por Movimientos En Masa Detonados Por Lluvias Para Una Región de Los Andes Colombianos Estimando La Probabilidad Espacial, Temporal, y Magnitud. *Rev. Boletín Geol.* **2019**, *41*, 85–105. [CrossRef]
48. Aristizábal, E.; Morales-García, P.; Vásquez-Guarín, M.; Ruíz-Vásquez, D.; Palacio-Córdoba, J.; Ángel-Cárdenas, F.P.; Caballero-Acosta, H.; Ordóñez-Carmona, O. 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 Geol.* **2022**, *43*, 199–217. [CrossRef]
49. Ruiz-Vásquez, D.; Aristizábal, E. Landslide Susceptibility Assessment in Mountainous and Tropical Scarce-Data Regions Using Remote Sensing Data: A Case Study in the Colombian Andes. In *EGU General Assembly*; EGU: Vienna, Austria; 8–13 April 2018, p. 3408.
50. Correa Muñoz, N.A.; Higidio Castro, J.F. Determination of Landslide Susceptibility in Linear Infrastructure. Case: Aqueduct Network in Palacé, Popayan (Colombia). *Ing. Investig.* **2017**, *37*, 17–24. [CrossRef]
51. Pradhan, A.M.S.; Lee, J.-M.; Kim, Y.-T. Semi-Quantitative Method to Identify the Vulnerable Areas in Terms of Building Aggregation for Probable Landslide Runout at the Regional Scale: A Case Study from Soacha Province, Colombia. *Bull. Eng. Geol. Environ.* **2019**, *78*, 5745–5762. [CrossRef]
52. Valencia Ortiz, J.A.; Martínez-Graña, A.M. A Neural Network Model Applied to Landslide Susceptibility Analysis (Capitanejo, Colombia). *Geomat. Nat. Hazards Risk* **2018**, *9*, 1106–1128. [CrossRef]
53. Renza, D.; Cárdenas, E.A.; Martinez, E.; Weber, S.S. CNN-Based Model for Landslide Susceptibility Assessment from Multispectral Data. *Appl. Sci.* **2022**, *12*, 8483. [CrossRef]
54. Cullen, C.A.; Al Suhili, R.; Aristizabal, E. A Landslide Numerical Factor Derived from CHIRPS for Shallow Rainfall Triggered Landslides in Colombia. *Remote Sens.* **2022**, *14*, 2239. [CrossRef]
55. Aristizábal Giraldo, E.V.; García Aristizábal, E.; Marín Sánchez, R.; Gómez Cardona, F.; Guzmán Martínez, J.C. Rainfall-Intensity Effect on Landslide Hazard Assessment Due to Climate Change in North-Western Colombian Andes. *Rev. Fac. Ing. Univ. Antioquia* **2022**, *103*, 51–66. [CrossRef]
56. Marín, R.J.; García, E.F.; Aristizábal, E. Effect of Basin Morphometric Parameters on Physically-Based Rainfall Thresholds for Shallow Landslides. *Eng. Geol.* **2020**, *278*, 105855. [CrossRef]
57. Marín, R.J.; Velásquez, M.F.; García, E.F.; Alvioli, M.; Aristizábal, E. Assessing Two Methods of Defining Rainfall Intensity and Duration Thresholds for Shallow Landslides in Data-Scarce Catchments of the Colombian Andean Mountains. *Catena* **2021**, *206*, 105563. [CrossRef]
58. Gutiérrez Alvis, D.E.; Bornachera Zarate, L.S.; Mosquera Palacios, D.J. Sistema de Alerta Temprana Por Movimiento En Masa Inducido Por Lluvia Para Ciudad Bolívar (Colombia). *Ing. Solidar.* **2018**, *14*, 26. [CrossRef]
59. Ramos-Cañón, A.M.; Prada-Sarmiento, L.F.; Trujillo-Vela, M.G.; Macías, J.P.; Santos-R, A.C. Linear Discriminant Analysis to Describe the Relationship between Rainfall and Landslides in Bogotá, Colombia. *Landslides* **2016**, *13*, 671–681. [CrossRef]
60. Velásquez, N. Assessment of Deep Convective Systems in the Colombian Andean Region. *Hydrology* **2022**, *9*, 119. [CrossRef]
61. Ortega, L.C.; Cañón, J.E. Correlative Analysis of Climate Impacts in an Andean Municipality of Colombia. *Rev. Cienc. Agrícolas* **2022**, *39*, 143–159. [CrossRef]

62. Parravano, V.; Teixell, A.; Mora, A. Influence of Salt in the Tectonic Development of the Frontal Thrust Belt of the Eastern Cordillera (Guatiquía Area, Colombian Andes). *Interpretation* **2015**, *3*, SAA17–SAA27. [CrossRef]
63. Chicangana, G.; Kammer, A. Evolución Tectónica de La Cordillera Oriental de Colombia. Desde La Apertura Del Océano Iapeto Hasta la Conformación de la Pangea: Una Visión Preliminar. Primera Parte: Aspectos Geológicos. *Geol. Colomb.* **2013**, *38*, 64–74.
64. Servicio Geológico Colombiano: Mapa geológico de Colombia. 2019. Available online: https://www2.sgc.gov.co/MGC/Paginas/mgc2M2019.aspx (accessed on 30 May 2023).
65. DANE: Censo Nacional de Población y Vivienda. 2018. Available online: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018 (accessed on 30 May 2023).
66. Pulido, O.; Gómez, L.S. Geología Plancha 266 Villavicencio. In *Memoria Explicativa*; Ingeominas: Bogotá, Colombia, 2001.
67. JAXA-METI. ALOS Systematic Observation Strategy—PALSAR. Available online: https://www.eorc.jaxa.jp/ALOS/en/obs/palsar_strat.htm (accessed on 30 May 2023).
68. Alaska Satellite Facility. Available online: https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/ (accessed on 30 May 2023).
69. Servicio Geológico Colombiano. Atlas Geológico de Colombia. 2015. Available online: https://www2.sgc.gov.co/MGC/Paginas/agc_500K2015.aspx (accessed on 30 May 2023).
70. Copernicus Open Access Hub. Available online: https://scihub.copernicus.eu/dhus/ (accessed on 30 May 2023).
71. Ideam. Atlas Climatológico de Colombia. Available online: http://atlas.ideam.gov.co/visorAtlasClimatologico.html (accessed on 30 May 2023).
72. Google Earth. Available online: https://www.google.es/intl/es/earth/index.html (accessed on 30 May 2023).
73. QGIS 3. A Free and Open Source Geographic Information System. 2023. Available online: https://www.qgis.org/en/site/ (accessed on 30 May 2023).
74. SAGA. System for Automated Geoscientific Analyses. Available online: https://saga-gis.sourceforge.io/en/index.html (accessed on 30 May 2023).
75. Rstudio. Available online: https://www.rstudio.com/categories/rstudio-ide/ (accessed on 30 May 2023).
76. Chacón, J.; Irigaray, C.; Fernández, T.; El Hamdouni, R. Engineering Geology Maps: Landslides and Geographical Information Systems. *Bull. Eng. Geol. Environ.* **2006**, *65*, 341–411. [CrossRef]
77. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K. Landslide Inventory Maps: New Tools for an Old Problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [CrossRef]
78. Varnes, D.J. Slope Movements Types and Processes. In *Landslides Analysis and Control*; Schuster, R.L., Krizek, R.J., Eds.; National Academy of Sciences: Washington, DC, USA, 1978; pp. 11–33.
79. WP/WLI; Cruden, D.M. A Suggested Method for Describing the Activity of a Landslide. *Bull. Int. Assoc. Eng. Geol.* **1993**, *47*, 53–57.
80. Bravo-López, E.; Fernández Del Castillo, T.; Sellers, C.; Delgado-García, J. Analysis of Conditioning Factors in Cuenca, Ecuador, for Landslide Susceptibility Maps Generation Employing Machine Learning Methods. *Land* **2023**, *12*, 1135. [CrossRef]
81. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Third Earth Resources Technology Satellite. In *Monitoring Vegetation Systems in the Great Plains with ERTS*; Technical Presentations; NASA: Washington, DC, USA, 1974; Volume I, pp. 309–317.
82. Pacheco-Quevedo, R.; Velastegui-Montoya, A.; Montalván-Burbano, N.; Morante-Carballo, F.; Korup, O.; Rennó, C.D. Land use and land cover as a conditioning factor in landslide susceptibility: A literature review. *Landslides* **2023**, *20*, 967–982. [CrossRef]
83. DeGraff, J.V.; Romesburg, H.C. Regional Landslide Susceptibility Assessment for Wildland Management: A Matrix Approach. In *Thresholds in Geomorphology*; Coates, D.R., Vitek, J.D., Eds.; Routledge: Boston, MA, USA, 1980; pp. 401–414.
84. Irigaray Fernández, C.; Fernández del Castillo, T.; El Hamdouni, R.; Chacón Montero, J. Evaluation and Validation of Landslide-Susceptibility Maps Obtained by a GIS Matrix Method: Examples from the Betic Cordillera (Southern Spain). *Nat. Hazards* **2007**, *41*, 61–79. [CrossRef]
85. Li, L.; Lan, H.; Wu, Y. How Sample Size Can Effect Landslide Size Distribution. *Geoenviron. Disasters* **2016**, *3*, 18. [CrossRef]
86. Shao, X.; Ma, S.; Xu, C.; Zhou, Q. Effects of Sampling Intensity and Non-Slide/Slide Sample Ratio on the Occurrence Probability of Coseismic Landslides. *Geomorphology* **2020**, *363*, 107222. [CrossRef]
87. Huang, F.; Cao, Z.; Jiang, S.-H.; Zhou, C.; Huang, J.; Guo, Z. Landslide Susceptibility Prediction Based on a Semi-Supervised Multiple-Layer Perceptron Model. *Landslides* **2020**, *17*, 2919–2930. [CrossRef]
88. Sameen, M.I.; Pradhan, B.; Bui, D.T.; Alamri, A.M. Systematic Sample Subdividing Strategy for Training Landslide Susceptibility Models. *Catena* **2020**, *187*, 104358. [CrossRef]
89. Dornik, A.; Drăguţ, L.; Oguchi, T.; Hayakawa, Y.; Micu, M. Influence of Sampling Design on Landslide Susceptibility Modeling in Lithologically Heterogeneous Areas. *Sci. Rep.* **2022**, *12*, 2106. [CrossRef]
90. Carrara, A. Multivariate Models for Landslide Hazard Evaluation. *J. Int. Assoc. Math. Geol.* **1983**, *15*, 403–426. [CrossRef]
91. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
92. Dou, J.; Yunus, A.P.; Tien Bui, D.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of Advanced Random Forest and Decision Tree Algorithms for Modeling Rainfall-Induced Landslide Susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [CrossRef] [PubMed]
93. Nefeslioglu, H.A.; Sezer, E.; Gokceoglu, C.; Bozkir, A.S.; Duman, T.Y. Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey. *Math. Probl. Eng.* **2010**, *2010*, 901095. [CrossRef]

94. Miner, A.; Vamplew, P.; Windle, D.J.; Flentje, P.; Warner, P. A Comparative Study of Various Data Mining Techniques as Applied to the Modeling of Landslide Susceptibility on the Bellarine Peninsula, Victoria, Australia. In *Geologically Active, Proceedings of the 11th IAEG Congress of the International Association of Engineering Geology and the Environment, Auckland, New Zealand, 5–10 September 2010*; CRC Press: Boca Raton, FL, USA, 2010.
95. Bravo-López, E.; Fernández Del Castillo, T.; Sellers, C.; Delgado-García, J. Landslide Susceptibility Mapping of Landslides with Artificial Neural Networks: Multi-Approach Analysis of Backpropagation Algorithm Applying the Neuralnet Package in Cuenca, Ecuador. *Remote Sens.* **2022**, *14*, 3495. [CrossRef]
96. Ciaburro, G.; Venkateswaran, B. *Neural Network with R: Smart Models Using CNN, RNN, Deep Learning, and Artificial Intelligence Principles*; Packt Publishing Ltd.: Birmingham, UK, 2017; Volume 91.
97. Günther, F.; Fritsch, S. Neuralnet: Training of Neural Networks. *R J.* **2010**, *2*, 30–38. [CrossRef]
98. Zare, M.; Pourghasemi, H.R.; Vafakhah, M.; Pradhan, B. Landslide Susceptibility Mapping at Vaz Watershed (Iran) Using an Artificial Neural Network Model: A Comparison between Multilayer Perceptron (MLP) and Radial Basic Function (RBF) Algorithms. *Arab. J. Geosci.* **2013**, *6*, 2873–2888. [CrossRef]
99. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. Hybrid Integration of Multilayer Perceptron Neural Networks and Machine Learning Ensembles for Landslide Susceptibility Assessment at Himalayan Area (India) Using GIS. *Catena* **2017**, *149*, 52–63. [CrossRef]
100. Gholamy, A.; Kreinovich, V.; Kosheleva, O. Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation. *Dep. Tech. Rep.* **2018**, *1209*, 1–6.
101. Vu, H.L.; Ng, K.T.W.; Richter, A.; An, C. Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation. *J. Environ. Manag.* **2022**, *311*, 114869. [CrossRef]
102. Zou, K.; O'Malley, A.; Mauri, L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation* **2007**, *115*, 654–657. [CrossRef]
103. Yilmaz, I. A Case Study from Koyulhisar (Sivas-Turkey) for Landslide Susceptibility Mapping by Artificial Neural Networks. *Bull. Eng. Geol. Environ.* **2009**, *68*, 297–306. [CrossRef]
104. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide Susceptibility Mapping Using Frequency Ratio, Analytic Hierarchy Process, Logistic Regression, and Artificial Neural Network Methods at the Inje Area, Korea. *Environ. Earth Sci.* **2013**, *68*, 1443–1464. [CrossRef]
105. Flórez-García, A.C.; Pérez Castillo, J.N. Técnicas Para La Predicción Espacial de Zonas Susceptibles a Deslizamientos. *Av. Investig. Ing.* **2019**, *16*, 20–48. [CrossRef]
106. Yi, Y.; Zhang, W.; Xu, X.; Zhang, Z.; Wu, X. Evaluation of Neural Network Models for Landslide Susceptibility Assessment. *Int. J. Digit. Earth* **2022**, *15*, 934–953. [CrossRef]
107. Aslam, B.; Zafar, A.; Khalil, U. Comparative Analysis of Multiple Conventional Neural Networks for Landslide Susceptibility Mapping. *Nat. Hazards* **2023**, *115*, 673–707. [CrossRef]
108. Chung, C.J.F.; Fabbri, A.G. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat. Hazards* **2003**, *30*, 451–472. [CrossRef]
109. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of Prediction Capability of the Artificial Neural Networks for Mapping Landslide Susceptibility in the Turbolo River Catchment (Northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [CrossRef]
110. Hungr, O.; Leroueil, S.; Picarelli, L. The Varnes Classification of Landslide Types, an Update. *Landslides* **2014**, *11*, 167–194. [CrossRef]
111. Fernández, T.; Pérez García, J.L.; Gómez López, J.M.; Cardenal, F.J.; Moya-Giménez, F.; Delgado, J. Multitemporal Landslide Inventory and Activity Analysis by Means of Aerial Photogrammetry and LiDAR Techniques in an Area of Southern Spain. *Remote Sens.* **2021**, *13*, 2110. [CrossRef]
112. Gariano, S.L.; Guzzetti, F. Landslides in a Changing Climate. *Earth Sci. Rev.* **2016**, *162*, 227–252. [CrossRef]
113. Correa, O.; García, F.; Bernal, G.; Cardona, O.D.; Rodriguez, C. Early Warning System for Rainfall-Triggered Landslides Based on Real-Time Probabilistic Hazard Assessment. *Nat. Hazards* **2020**, *100*, 345–361. [CrossRef]
114. García-Delgado, H.; Contreras, N.M. Historical Distribution for Landslides Triggered by Earthquakes in the Colombian Region. In Proceedings of the XIII International Symposium on Landslides, Cartagena, Colombia, 15–19 June 2020.
115. Bermúdez, M.A.; Velandia, F.; García-Delgado, H.; Jiménez, D.; Bernet, M. Exhumation of the Southern Transpressive Bucaramanga Fault, Eastern Cordillera of Colombia: Insights from Detrital, Quantitative Thermochronology and Geomorphology. *J. S. Am. Earth Sci.* **2021**, *106*, 103057. [CrossRef]
116. Grima, N.; Edwards, D.; Edwards, F.; Petley, D.; Fisher, B. Landslides in the Andes: Forests Can Provide Cost-Effective Landslide Regulation Services. *Sci. Total Environ.* **2020**, *745*, 141128. [CrossRef] [PubMed]
117. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How Can Statistical Models Help to Determine Driving Factors of Landslides? *Ecol. Modell.* **2012**, *239*, 27–39. [CrossRef]
118. Zhu, A.-X.; Miao, Y.; Yang, L.; Bai, S.; Liu, J.; Hong, H. Comparison of the Presence-Only Method and Presence-Absence Method in Landslide Susceptibility Mapping. *Catena* **2018**, *171*, 222–233. [CrossRef]
119. Costanzo, D.; Rotigliano, E.; Irigaray, C.; Jiménez-Perálvarez, J.D.; Chacón, J. Factors Selection in Landslide Susceptibility Modelling on Large Scale Following the Gis Matrix Method: Application to the River Beiro Basin (Spain). *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 327–340. [CrossRef]

120. Meena, S.R.; Puliero, S.; Bhuyan, K.; Floris, M.; Catani, F. Assessing the Importance of Conditioning Factor Selection in Landslide Susceptibility for the Province of Belluno (Region of Veneto, Northeastern Italy). *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 1395–1417. [CrossRef]
121. Liu, L.L.; Yang, C.; Wang, X.M. Landslide Susceptibility Assessment Using Feature Selection-Based Machine Learning Models. *Geomech. Eng.* **2021**, *25*, 1–16.
122. Chacón Montero, J.; Irigaray Fernández, C.; Fernández del Castillo, T. Large to Middle Scale Landslides Inventory, Analysis and Mapping with Modelling and Assessment of Derived Susceptibility, Hazards and Risks in a GIS. In *International Congress International Association of Engineering Geology*; A.A. Balkema: Rotterdam, The Netherlands, 1994; pp. 4669–4678.
123. van Westen, C.; van Asch, T.; Soeters, R. Landslide hazard and risk zonation—Why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [CrossRef]
124. Huabin, W.; Gangjun, L.; Weiya, X.; Gonghui, W. GIS-based landslide hazard assessment: An overview. *Prog. Phys. Geogr. Earth Environ.* **2005**, *29*, 548–567. [CrossRef]
125. Korup, O.; Stolle, A. Landslide Prediction from Machine Learning. *Geol. Today* **2014**, *30*, 26–33. [CrossRef]
126. Sahin, E.K. Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. *SN Appl. Sci.* **2020**, *2*, 1308. [CrossRef]
127. Deng, H.; Wu, X.; Zhang, W.; Liu, Y.; Li, W.; Li, X.; Zhou, P.; Zhuo, W. Slope-Unit Scale Landslide Susceptibility Mapping Based on the Random Forest Model in Deep Valley Areas. *Remote Sens.* **2022**, *14*, 4245. [CrossRef]
128. Wei, A.; Yu, K.; Dai, F.; Gu, F.; Zhang, W.; Liu, Y. Application of Tree-Based Ensemble Models to Landslide Susceptibility Mapping: A Comparative Study. *Sustainability* **2022**, *14*, 6330. [CrossRef]
129. Bilbao, I.; Bilbao, J. Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. In Proceedings of the Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 5–7 December 2017; pp. 173–177. [CrossRef]
130. Lv, L.; Chen, T.; Dou, J.; Plaza, A. A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102713. [CrossRef]