https://rbgeomorfologia.org.br/
ISSN 2236-5664

Revista Brasileira de Geomorfologia

v. 25, nº 4 (2024)

http://dx.doi.org/10.20502/rbgeomorfologia.v25i4.2513

Research Article

Landslide Susceptibility Modeling Using Artificial Neural Networks in the Municipality of Joinville, southern Brazil

Modelagem com redes neurais artificiais da suscetibilidade a movimentos de massa do município de Joinville, SC, Brasil

Renato Ribeiro Mendonça [1], Guilherme Garcia de Oliveira [2] e Carlos Gustavo Tornquist [3]

[1] Federal University of Rio Grande do Sul - UFRGS, Graduate Program in Remote Sensing - PPGSR, Porto Alegre, Brazil. renatormendonca@gmail.com.
ORCID: https://orcid.org/0009-0008-1038-7698

[2] Federal University of Rio Grande do Sul - UFRGS, State Research Center for Remote Sensing and Meteorology - CEPSRM, Porto Alegre, Brazil. g.g.oliveira10@gmail.com.
ORCID: https://orcid.org/0000-0003-4197-5704

[3] Federal University of Rio Grande do Sul - UFRGS, Faculty of Agronomy, Department of Soil Science, Brazil. carlos.tornquist@ufrgs.br.
ORCID: https://orcid.org/0000-0002-5715-0654

Received: 29/11/2023; Accepted: 17/12/2024; Published: 29/12/2024

Abstract: Assessing landslide susceptibility in a municipality is crucial for disaster prevention, and Artificial Neural Networks (ANN´s) have proven effective in this analysis. This study aimed to model landslides susceptibility in the municipality of Joinville, Santa Catarina state, southern Brazil, using ANNs. The municipality has a significant history of such events, allowing for an inventory of occurrence areas (OC) through polygon mapping on satellite images. For non-occurrence areas (NO), a 1 km radius buffer was used, subtracting OC from it. Random points were generated at 10 m intervals, with a value of 1 for OC and 0 for NO. The explanatory variables were divided into three groups: (i) morphometric variables, (ii) horizontal distances to roads and structural lineaments, and (iii) geo-environmental cartographic databases. Five ANN´s configurations were tested. Validation employed metrics such as area under the ROC curve (AUC) and overall accuracy (ACC), with the best modeling yielding an AUC of 0.90 and ACC of 0.84. This result utilized all explanatory variables except land use and cover, which caused a slight bias in the ANN due to the predominance of landslides in forested areas in the inventory. Geology played a crucial role in determining susceptibility.

Keywords: Disaster Prevention, Machine Learning, Landslide Susceptibility.

Resumo: A avaliação da suscetibilidade a deslizamentos em um município é crucial na prevenção de desastres. Redes Neurais Artificiais (RNA) provaram ser eficazes nessa análise. Este estudo modelou a suscetibilidade a deslizamentos em Joinville, Santa Catarina, usando RNA. O município tem histórico significativo desses eventos, permitindo um levantamento de áreas de ocorrência (OC) através do mapeamento de cicatrizes em imagens de satélite. Para áreas de não ocorrência (NO), foi utilizado um buffer de 1 km, subtraindo as OC. Pontos aleatórios foram gerados a cada 10 m, com valor 1 para OC e 0 para NO. As variáveis explicativas foram divididas em três grupos: (i) morfométricas, (ii) distâncias horizontais para estradas e estruturas e (iii) dados cartográficos geoambientais. Cinco configurações de RNA foram testadas. Na validação, métricas como área sob a curva ROC (AUC) e acurácia global (ACC) foram usadas, com a melhor modelagem apresentando AUC de 0,90 e ACC de 0,84. Essa configuração usou todas as variáveis explicativas, exceto uso e cobertura da terra, causando um leve viés na RNA, devido ao predomínio de cicatrizes em áreas florestais no inventário. A geologia desempenhou um papel crucial na determinação da suscetibilidade.

Palavras-chave: Prevenção de Desastres, Aprendizado de Máquina, Suscetibilidade a Movimentos de Massa.

1. Introduction

A study of susceptibility to landslide events is an effective way to mitigate disasters caused by these phenomena and to technically support public decision-making for sustainable municipal land use and occupation. Susceptibility to landslides is understood as the propensity of slopes and terrains to develop landslides processes and related events (SOBREIRA; SOUZA, 2012), (BRESSANI; COSTA, 2013), (MINISTÉRIO DAS CIDADES, 2013). Several landslide models have been developed, using approaches based on physics (conceptual models), heuristic analysis, and statistical methods (LUO; LIU, 2017). Each approach offers a unique perspective to evaluate landslides susceptibility, considering specific and complex factors.

For example, the Geological Survey of Brazil (CPRM) and the Institute for Technological Research (IPT) use quantitative models combined with heuristic analysis. These models are derived from the relationship between the density of landslide areas in a given municipality, structural lineament density, and terrain slope (BITAR, 2014). This susceptibility model was designed to meet two key objectives: broad applicability across Brazil's diverse territory and rapid generation to support a continuous government program for mapping susceptibility charts, with the first efforts dating back to 2012.

One challenge of this method is the influence of the modeler's interpretation on the final model's quality. Input data, such as geostructural lineaments, rely on the modeler's interpretation of the digital elevation model (LAMBERTY; KEPEL FILHO; NORONHA, 2015).

The use of machine learning (ML) and artificial intelligence (AI) techniques is a robust way to assess the susceptibility of a region to Landslides. With the rapid advancements in remote sensing technologies in recent years, it is possible to collect a large volume of landslides related data more quickly and efficiently. Many researchers have utilized such data to model landslides susceptibility using AI methodologies (PASCALE et al., 2013; ZHU et al., 2018; EMANI, 2020; LUCCHESE; OLIVEIRA; PEDROLLO, 2021). These methods can analyze the complex relationships between landslides susceptibility and its influencing factors through large datasets (ZHU et al., 2018).

As highlighted by Lucchese, Oliveira, and Pedrollo (2021), artificial intelligence encompasses the theory and development of computer systems that simulate human intelligence. According to Zhu et al. (2018), several studies have applied AI to landslides susceptibility modeling (GÓMEZ; KAVZOGLU, 2004; PRADHAN; LEE, 2010; DOU et al., 2015; YAO et al., 2022), as these methods excel in assimilating extensive input datasets and landslides inventories, producing susceptibility maps with superior performance metrics (e.g., overall accuracy) compared to conventional modeling methods.

Among the various ML approaches, Artificial Neural Networks (ANNs) stand out for their flexibility regarding data scale and their ability to incorporate both qualitative and quantitative variables into their analyses (KAWABATA; BANDIBAS, 2009), regardless of data distribution (WANG et al., 1995). Studies achieving satisfactory results in adjusting ANNs for landslides susceptibility include Pradhan and Lee (2010), who conducted research in Cameron Highlands, Malaysia. Using an inventory of 324 landslides polygons over 293 km² and terrain attributes derived from a 10 m DEM, they employed a backpropagation ANN algorithm with 10 inputs, 22 hidden layer neurons, and 2 outputs. Results showed that slope was the most influential variable, with the model achieving an AUC (Area Under the ROC Curve) of 0.83.

Similarly, Quevedo et al. (2019a) modeled landslides susceptibility in the Rolante River Basin, northeastern Rio Grande do Sul, Brazil, using data from 308 landslide caused by an extreme rainfall event in 2017. Morphometric information extracted from a digital terrain model was used to train the ANN, yielding validation results with an AUC exceeding 0.9. Ullah (2022) proposed a multi-process geodynamic susceptibility mapping approach using three methods: ANN, logistic regression, and k-nearest neighbors (KNN). The ANN was trained on historical data of flash floods, debris flows, and shallow planar landslides from satellite images in Shangla District, Pakistan, using various geomorphometric parameters, geological maps, and land use data. The ANN outperformed the other methods, achieving AUC values of 0.98 for flash floods, 0.94 for landslides, and 0.98 for debris flows.

ANNs' extrapolation ability is particularly valuable in determining landslides susceptibility in areas with little or no landslide occurrences. According to Gameiro (2020), ANN models are empirical and may underperform when extrapolating beyond the training domain. However, the author noted that spatial extrapolation in geomorphologically homogeneous areas can produce satisfactory results, especially when terrain attributes are similar between training and testing areas. Sampling distances for non-occurrence data are a critical factor in obtaining better results, enabling the ANN to more effectively distinguish susceptible from non-susceptible areas. Gameiro's study sampled five different basins in the Serra Geral region to train ANNs for each dataset, emphasizing the influence of sampling on landslides susceptibility mapping. The author highlighted the importance of variables such as slope, LS factor, and elevation. Model accuracy improved with larger buffers for non-occurrence samples, and sample representativity impacted extrapolation capabilities.

This study aims to analyze landslides susceptibility using ANNs in the municipality of Joinville, located in Santa Catarina, southern Brazil. Joinville has a significant history of landslides events, the most notable occurring in November 2008 (ODEBRECHT et al., 2017). Due to this history, Joinville has a substantial landslides inventory that can be leveraged to construct an ANN model and generate an accurate landslides susceptibility map.

2. Materials and Methods

To develop an landslides susceptibility model using ANN, it is necessary to sample areas where landslides have occurred, as well as areas where these events have not taken place. According to Wang et al. (2019), identifying these areas makes it possible to relate them to terrain attributes associated with landslides and then extrapolate the results to other areas based on the assumption that future landslides will occur in environments similar to those of past events. Therefore, for the purpose of this study, landslides within the study area were mapped and identified. After collecting the landslide polygons, these areas were correlated with data on terrain morphology and its substrate, which directly affect stability. Finally, these data were used to train an efficient neural network to create a terrain susceptibility model for landslides. The flowchart illustrating the preparation of sampling data, processing, validation, and the presentation of the susceptibility map is shown in Figure 1.

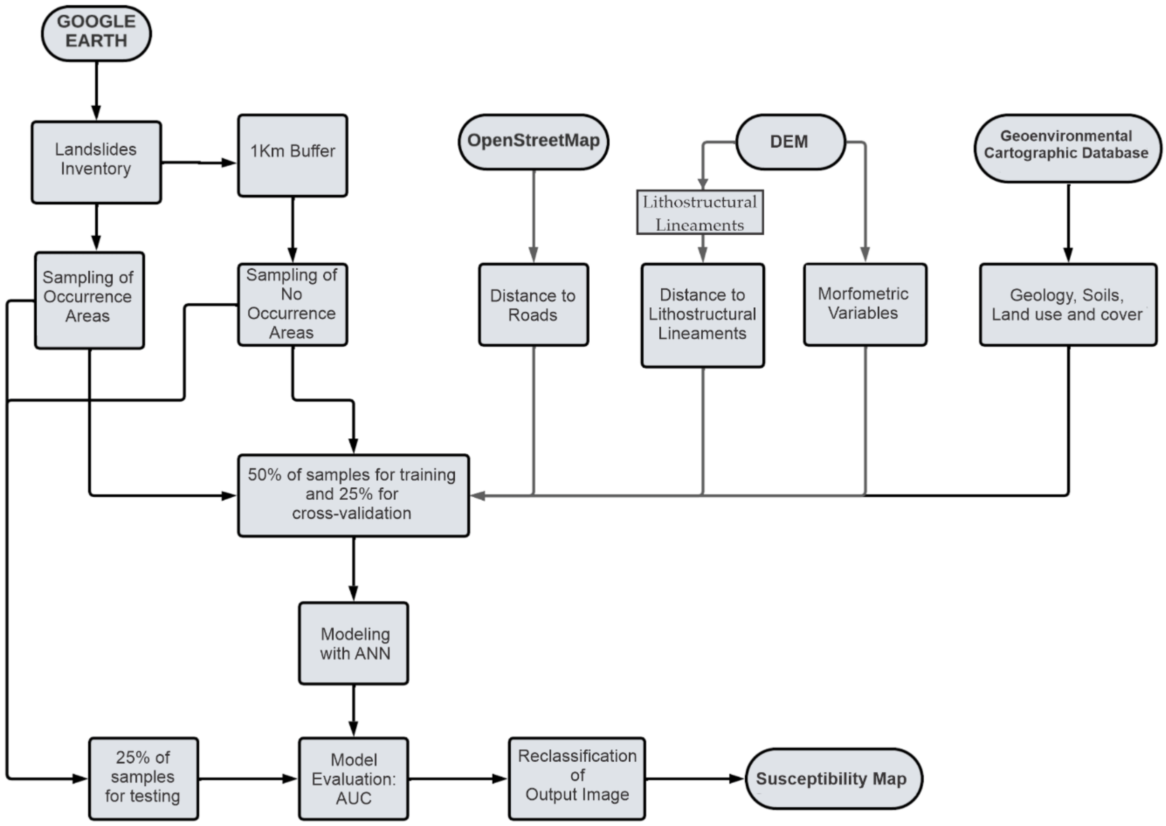

Figure 1. Methodological flowchart for obtaining the susceptibility model. Source: Organized by the author.

2.1. Study Area

The study of landslide susceptibility was conducted in the municipality of Joinville, Santa Catarina (Figure 1). Joinville is located in the northern part of the state and has the largest population. In addition to being the economic hub of the northern mesoregion, its urban area is expanding (IBGE 2022). According to Odebrecht et al. (2017), the coastline of Santa Catarina recorded extreme precipitation values between September and November 2008, accumulating more than 1000 mm of rainfall, with the peak on November 22 and 23. During this period, more than 800 landslide events occurred in the municipality; however, despite the high number, the incidence of landslide in the urbanized area was minimal, resulting only in material damage and no fatalities.

Figure 2 shows the landslides polygons identified in historical images from the Google Earth Pro software. To improve sampling and expand the landslides database, with the aim of diversifying the sample set and enhancing prior knowledge of the ANN model, as well as the possibility of applying the same model to adjacent municipalities, it was decided to extrapolate the municipal boundary. In addition to the Cubatão River basin, which covers the northern area of the municipality, the Itapocu River basin and its two tributaries, the Itapocuzinho and Jaraguá rivers, were included. The southern part of the Quiriri mountain range, to the north of the municipality, was also considered due to its significant collection of landslide polygons. Thus, the sum of these areas comprises the Sampling Area (SA), as shown in Figure 2.

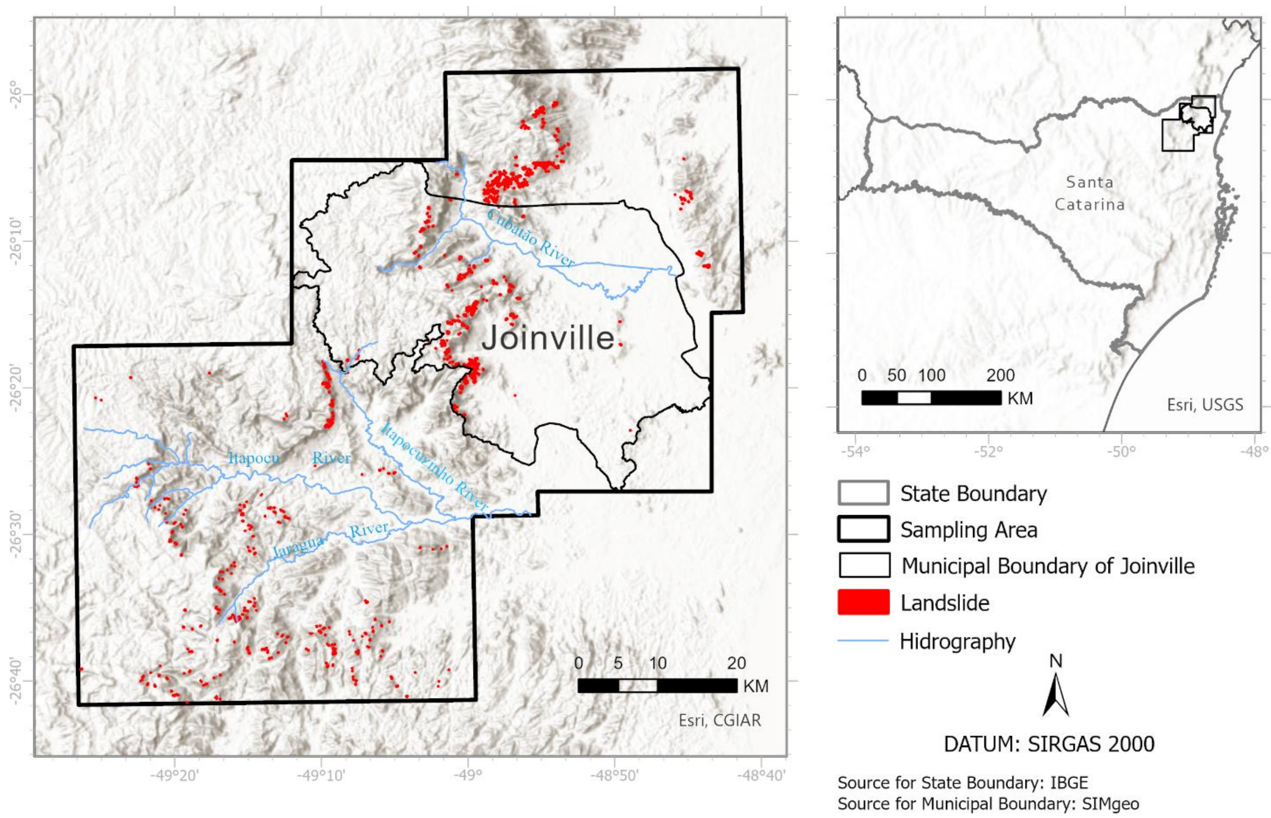

Figure 2. Sampling area of landslides for the study. Source: Organized by the author.

2.2. Sampling of occurrences and non-occurrences.

The mapping of the landslide polygons inventory was carried out through visual interpretation and vectorization, using remote sensing images from the Google Earth PRO software, as done in the works of Pham et al. (2017) and Wang et al. (2019). A total of 784 landslides were identified and mapped, including both the area of initiation and the area of impact of the landslides, from 2009 to 2018. The area of landslide impact within the perimeter of the Sampling Area (SA) was also mapped. After defining the occurrence areas, random points were generated within these areas using ArcGIS PRO 3.01 software (ESRI, 2023), with a minimum spacing of 10 meters between them, totaling 20,480 occurrence sampling points.

According to Lucchese, Oliveira, and Pedrollo (2021), one of the considerations related to sampling locations where landslides did not occur is the definition of a maximum distance from the landslide polygon. This determination is of great importance as it aims to prevent sample collection being restricted to areas close to landslides. A broader approach in sampling non-occurrence locations is effective for understanding the factors that influenced landslide in the polygon, in contrast to surrounding areas. Additionally, this approach considers the premise that rainfall, which can trigger landslide events, shows significant spatial variability. However, recommending the selection of areas too far from the landslides may be inappropriate, as there is no guarantee that both areas experienced the same amount of rainfall as the region of the landslides.

Therefore, for generating the non-occurrence area inventory, a buffer with a 1 km radius from the landslides was created, excluding the landslides areas, which prevents the collection of non-occurrence samples in these areas. The occurrence and non-occurrence areas are exemplified in Figure 3. Random points were then generated in the perimeter of this non-occurrence area, with the same spacing as the occurrence areas, limited to the same number of 20,480 sample points, maintaining a balanced (1:1) ratio between the two sample groups to avoid biasing the ANN model. Finally, the samples were classified with the value 1 for occurrence points and value 0 for non-occurrence points.

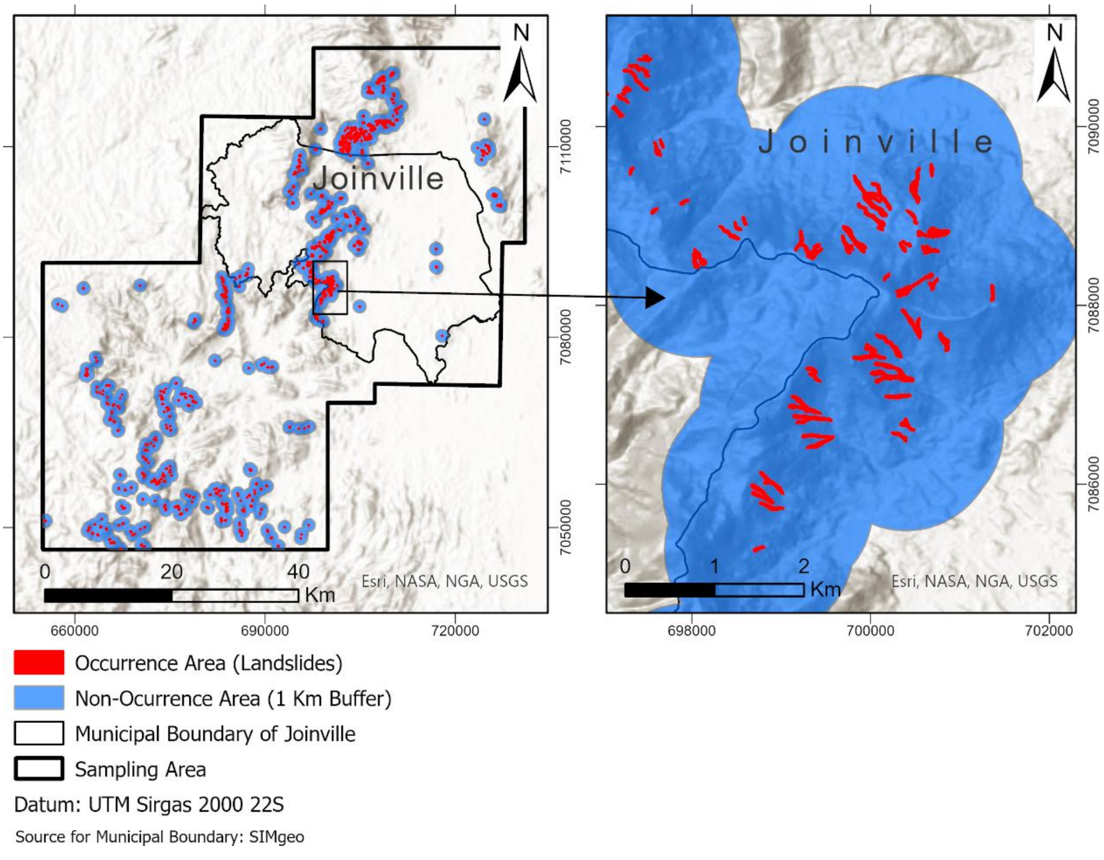

Figure 3. Areas of landslide occurrence and non-occurrence (1 km radius buffer). Source: Organized by the author.

2.3. Dataset

The research by Reichenbach et al. (2018) highlights several variables that help explain the occurrence of landslides, making the careful selection of variables crucial for creating a reliable and robust susceptibility model. In this context, input data were selected to train the network that, in addition to influencing landslides in general, were in tune with the geomorphology and nature of the substrate of the Sampling Area (SA). The selected parameters were categorized into three sets: those related to terrain morphometry, those related to distances between structural lineaments and roads, and those derived from cartographic sources with qualitative geo-environmental characteristics. This comprehensive approach aims to ensure a more accurate and robust model of landslides susceptibility for the specific region under investigation. The selected variables are presented in Table 1.

Table 1. Explanatory variables used for modeling landslides susceptibility in Joinville, Santa Catarina State.

Variable	Source
Altimetry (ALT)	SDS DEM 1m.
Aspect (ASP)	SDS DEM 1m.
Slope (SLP)	SDS DEM 1m.
Horizontal Distance to Roads (HDR)	Open Street Map (2023)
Horizontal Distance to Lithostructural Lineaments (HDLL)	SDS DEM 1m intepetration.
Topographic Wetness Index (TWI)	SDS DEM 1m.
Geology	Geological Map of the State of Santa Catarina - CPRM (2014)
Land Use and Cover	MAPBIOMAS (2021)
Pedology	IBGE (2021)

The processing to obtain the morphometric data for the model was based on the Digital Elevation Model (DEM) provided by the Secretariat of State for Sustainable Economic Development (SDS) of the Government of the State of Santa Catarina. This DEM has spatial and altimetric resolution of 1 meter (SIGSC, 2017) and was obtained through aerial survey. The data is available for free on the SIGSC portal. For better computational performance, this DEM was resampled to a pixel size of 5 meters. The processing to obtain this dataset was carried out in ArcGIS PRO 3.01 software (ESRI, 2023). To ensure that the morphometric data is on the same scale, a network of equidistant points every 5 meters, both vertically and horizontally, was created. Then, the values of each morphometric variable were extracted for each point. With these values, a raster for each variable was generated using the Topo to Raster tool in ArcGIS PRO 3.01 software

2.3.1. Morfometric variables

The morphometric variables chosen for the study are: Altimetry (ALT), Slope (SLP), Aspect (ASP), and Topographic Wetness Index (TWI). These variables are shown in Figure 4. The ALT data are obtained from the DEM, where each pixel represents the altitude of a point. The study area has an altimetric range from 0 to 1540 meters. The SLP of the terrain is expressed in degrees and is directly related to the stability of the terrain. The ASP, or Aspect, has values in degrees ranging from 0 to 360. Some studies indicate that ASP has an indirect influence on slope stability due to the effects of precipitation, wind, and solar radiation impacting a specific slope face, resulting in distinct conditions of moisture, weathering, and prevailing vegetation on the terrain (BRAGAGNOLO et al., 2020 citing CHEN et al., 2017; DING et al., 2017; TIEN BUI et al., 2017).

According to Beven & Kirkby (1979), the Topographic Wetness Index (TWI) assesses the surface runoff and moisture accumulation in the terrain. It is defined as a logarithmic function of slope and contributing area, as shown in Equation 1.

$$TWI = ln\,(\frac{A_s}{tanB}) \tag{1}$$

where: A is the contributing area and B is the slope of the terrain. The TWI aims to represent the topographic control of soil moisture. According to Tien Bui et al. (2017), the susceptibility to landslides can be estimated as a function of the relationship between the topographic effects on the hydrological response of an area, directly influencing the pore pressure in the soil.

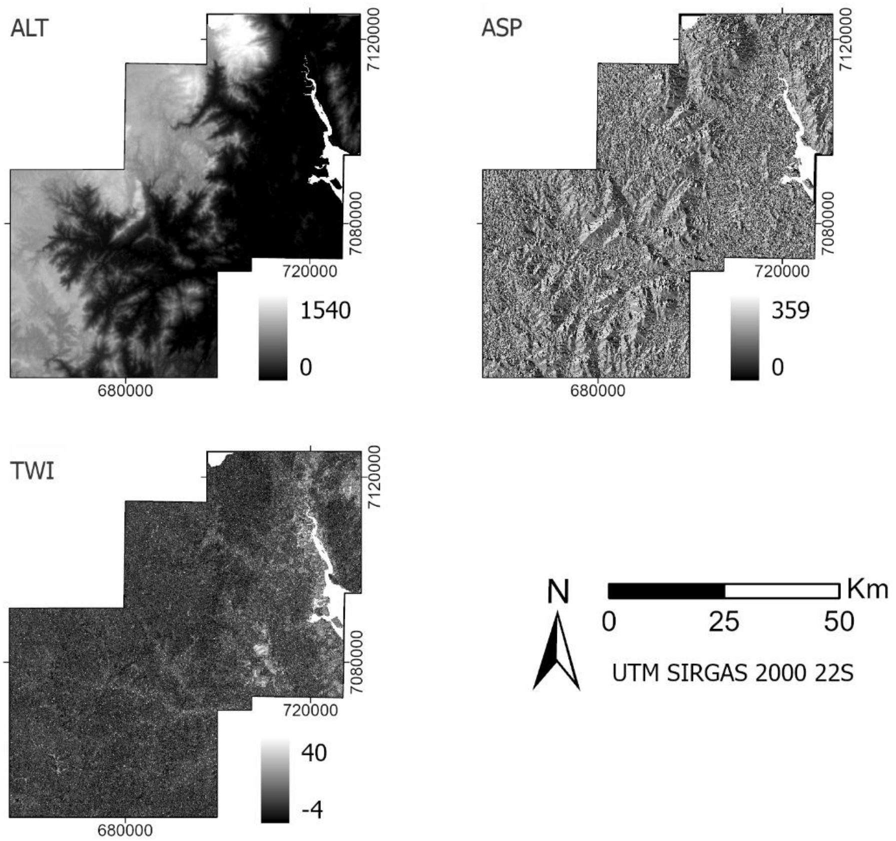

Figure 4. Selected morphometric variables for the study: Altimetry (ALT); Aspect (ASP), Slope (SLP), and Topographic Wetness Index (TWI). Source: Organized by the author.

2.3.2. Horizontal Distance to Roads.

In the study, two distance variables were used: Horizontal Distance to Roads (HDR) and Horizontal Distance to Lithostructural Lineaments (HDLL) (Figure 5). HDR is related to the degree of human intervention in the terrain, which can induce landslides on slopes cut by roads. To obtain the HDR, OpenStreetMap was used through the OSMDowloader plugin in QGIS 3.28, followed by the calculation of the Euclidean distance. On the other hand, HDLL is associated with the degree of fracturing in the rock substrate, resulting from geological structural deformations of a ruptile type. The HDLL was obtained by delimiting valleys defined by linear structures in the Digital Elevation Model (DEM), followed by vectorization and conversion into an image, also using the Euclidean distance calculation.

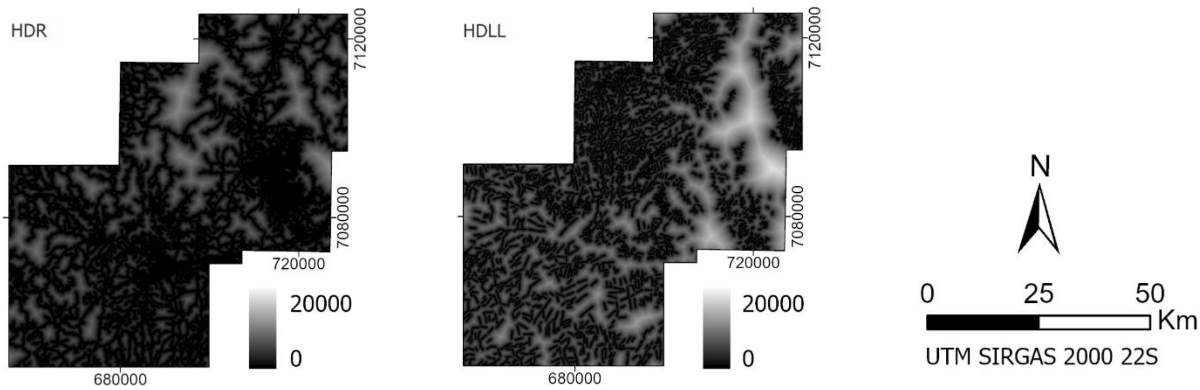

Figure 5. Horizontal Distance to Roads (HDR) and Horizontal Distance to Lithostructural Lineaments (HDLL) variables for the Sampling Area (SA). Source: Organized by the author.

2.3.3. Geoenvironmental Cartographic Database

Geology is widely used in susceptibility models (WANG et al., 2019). Different lithologies influence various levels of susceptibility (CHEN et al., 2017), as this variable is responsible for defining the shear strength of the materials that make up the slopes, playing a crucial role in determining their stability condition (ZÊZERE et al., 2017). For this study, lithology was obtained from the Geological Map of the State of Santa Catarina, provided by the Geological Service of Brazil - CPRM (2014), available at a scale of 1:500,000. Although it has a smaller scale than the one used for morphometric variables, this data was chosen due to its availability throughout the study area and its qualitative nature. The geological units were reinterpreted as lithological units to simplify the input into the neural network model (ANN).

According to Pham et al. (2016), the study of Land Use and Cover is relevant for identifying areas susceptible to landslides, especially considering the presence and removal of vegetation. For this purpose, the dataset from the Annual Mapping Project of Land Cover and Use of Brazil (MAPBIOMAS), collection 7.1, corresponding to the year 2021, at a scale of 1:100,000, was used. The land use and cover classes were simplified to Level 1, i.e., grouped according to the type of formation. The pedological data used in the study was provided by the Brazilian Institute of Geography and Statistics - IBGE (2021), at a scale of 1:250,000, and grouped according to the texture of each pedotype. Figure 6 shows the qualitative data used as input for training the neural network model.

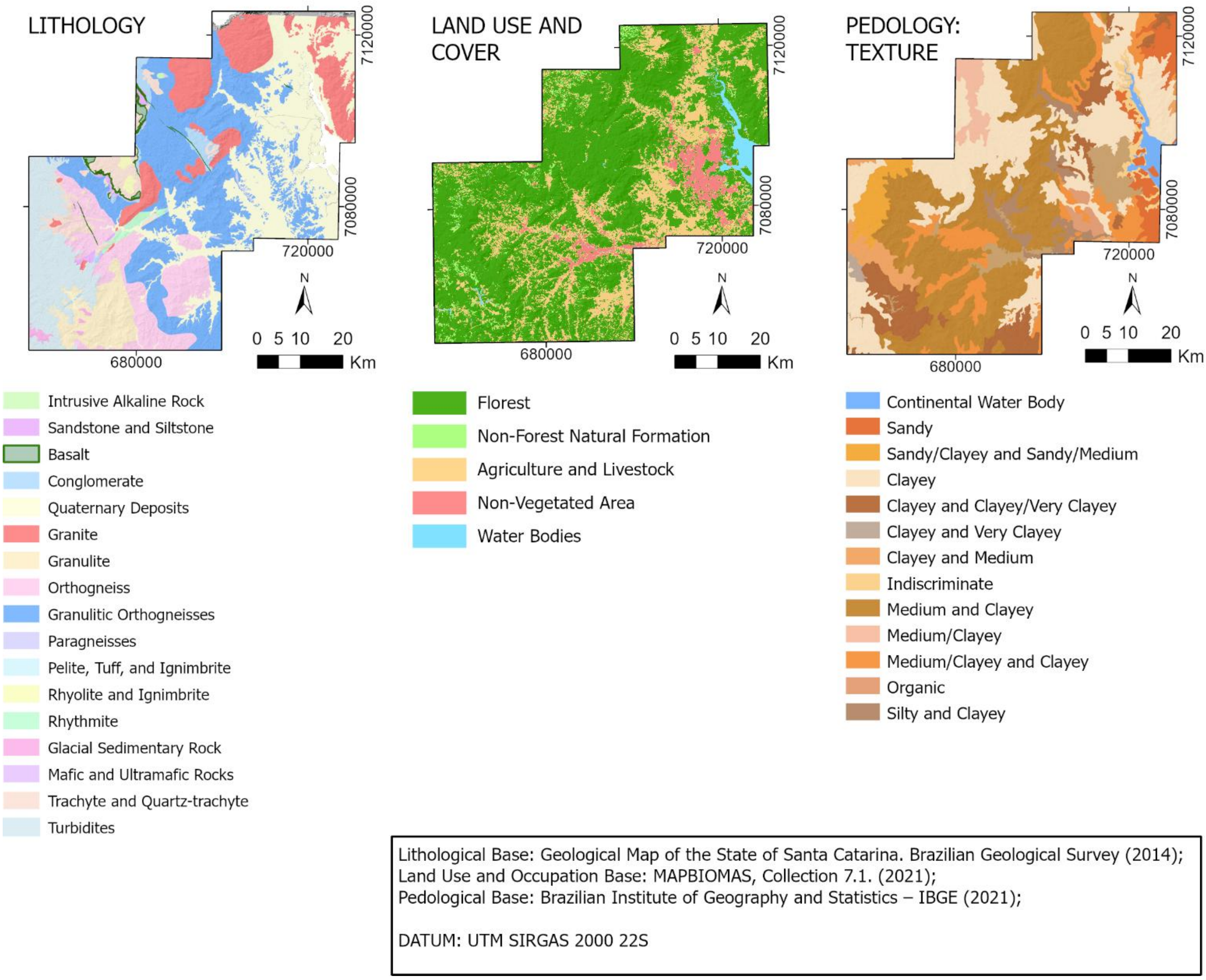

Figure 6. Geoenvironmental and qualitative database for the study area. Source: Organized by the author.

Since the classes of the qualitative maps of lithology, pedology, and land use and cover do not have an ordinal relationship, a one-hot encoding approach was adopted, where each class was represented as a combination of images with 0 or 1 codes. For the 17 lithology classes, 5 binary images were defined, which when combined allow the neural network (ANN) to differentiate all the classes. For the 4 land use and cover classes, 2 binary images were defined, while for the 14 soil type classes, 4 binary images were defined.

Finally, once the input dataset for the ANN was defined, the values of the variables were extracted for each sampling point, both for occurrence and non-occurrence cases. This extraction was performed using ArcGIS PRO 3.01 (ESRI, 2023). Additionally, an exploratory data analysis was conducted using boxplot-type graphs to investigate the relationship between the explanatory variables and the dependent variable.

2.4. Artificial Neural Networks

For this study, a Multilayer Perceptron (MLP) model of Artificial Neural Networks (ANNs) was employed, following the backpropagation method with multiple layers proposed by Rumelhart et al. (1986). The MATLAB 2021 software was used for network training, applying the Delta Rule (WIDROW; HOFF, 1960) for updating synaptic weights, through a script developed by the authors. Additionally, a parallel series cross-validation technique was adopted to avoid overfitting the network. The values of the variables were normalized through linear transformations, and the activation function used in the neurons was the sigmoid function.

Tests were conducted with five ANN configurations. The first configuration included all the input data from the proposed base (ANN 1). The other configurations excluded the qualitative data individually: one ANN without the geological data (ANN 2), another without the land use and cover data (ANN 3), and a third without the pedological data (ANN 4). Finally, an ANN with only morphometric data and distances (ANN 5) was tested to assess the impact of these data on the outcome. The maximum number of iterations or initializations for each ANN was set to 5, and the number of learning cycles was set to 50,000 for all configurations. The number of neurons for training the ANN followed the equation 2x+1, where x is the number of input variables.

The samples were divided into three sets: 50% for training, 25% for testing, and 25% for cross-validation. For model validation, global accuracy (ACC) and Area Under the Curve (AUC) metrics were used, according to Delong et al. (1988). The validations followed the same procedure as adopted by Lucchese et al. (2021) and Oliveira et al. (2019). Global accuracy is calculated by the following equation:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

In which: *TP* is the True Positives index, *TN* is the True Negatives index, *FP* is the False Positives index, and *FN* is the False Negatives index.

AUC is the area under the ROC curve, which expresses the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR). The samples were divided into several groups of two classes using different thresholds. For each division, we calculated the TPR and FPR, thus constructing the ROC curve (FAWCETT, 2006). The relationships are given by the following equations:

$$TPR = \frac{TP}{(TP + FN)} \tag{3}$$

and,

$$FPR = \frac{FP}{(FP + TN)} \tag{4}$$

In which: *TP* is the True Positives index, *TN* is the True Negatives index, *FP* is the False Positives index, and *FN* is the False Negatives index.

The model result is obtained in the form of an image containing values in the range of 0 to 1. These values were reclassified into 5 classes using natural breaks statistics, and each class was related to levels of susceptibility to landslides.

3. Results and discussion

For an initial statistical analysis, occurrence and non-occurrence samples for each morphometric measure were plotted in boxplot graphs, as shown in Figure 7. The topographic analysis revealed that most of the landslides occur at altitudes between 400 and 750 meters, with a median of 539.86 meters and a standard deviation of 226.01 meters. However, unlike the studies by Kawata and Bandibas (2009) and Chen et al. (2017), which identified this as a predominant attribute, in Joinville, the occurrence and non-occurrence data overlap, making it difficult to determine a preferred altitude for landslide occurrence. This discrepancy may be related to the topography of Joinville, where, despite a large range of altitudes, the terrain morphology consists of flattened forms separated by a large plateau escarpment, with the highest number of landslide occurrences concentrated in this escarpment.

The analysis of natural terrain slope (SLP) showed that angles above 30° are susceptible to landslides; the median slope is calculated at 35.74°, with a standard deviation of 10.46°. The SLP variable is recognized in several studies as one of the most significantly associated factors with landslide occurrence (PRADHAN; LEE, 2010; OLIVEIRA et al., 2018; QUEVEDO et al., 2019b). Similarly, the SLP variable stands out as the one that most effectively differentiates the occurrence and non-occurrence sample sets in Joinville, compared to other variables. Therefore, it is believed that this is the explanatory variable for the landslide model studied here. It is important to note that most areas with a slope above 30° in Joinville are associated with the plateau escarpment in the NE-SW direction, meaning this structural trend appears to influence susceptibility. In agreement with the previous information, the analysis of Aspect (ASP) indicated that most landslides occur in areas belonging to the second quadrant, meaning the direction of movement is towards the SE.

HDR data indicate that landslides are more frequent as they move away from the main roadways. In contrast, it is observed that landslides have a preferential occurrence near structural lineaments, suggesting that the primary nature of landslides in this region is influenced by local geological and geomorphological factors. Furthermore, the presence of human activities seems to have a reduced contribution to inducing these landslides within the studied sample area. The concentration of low TWI values indicates reduced topographic moisture retention in the locations of occurrence. However, the medians of the TWI indices for occurrence and non-occurrence are very close, making it difficult to establish a direct correlation.

Figure 8 presents the relationship between landslides and geoenvironmental data. It is noteworthy that a significant majority of landslides occur in lithotypes related to granites, orthogneisses, granulites, and orthogneiss granulites. Regarding land use and cover, landslides are predominantly associated with the Forest class. Additionally, the soil texture predominant in landslide areas is primarily related to soils of medium and clayey composition, presumably derived from high-grade igneous and metamorphic rocks. Overall, landslide occurrences in Joinville are associated with vegetated areas, with a shallow weathering profile, derived from granite and metamorphic rocks, with average slopes of 35.74°.

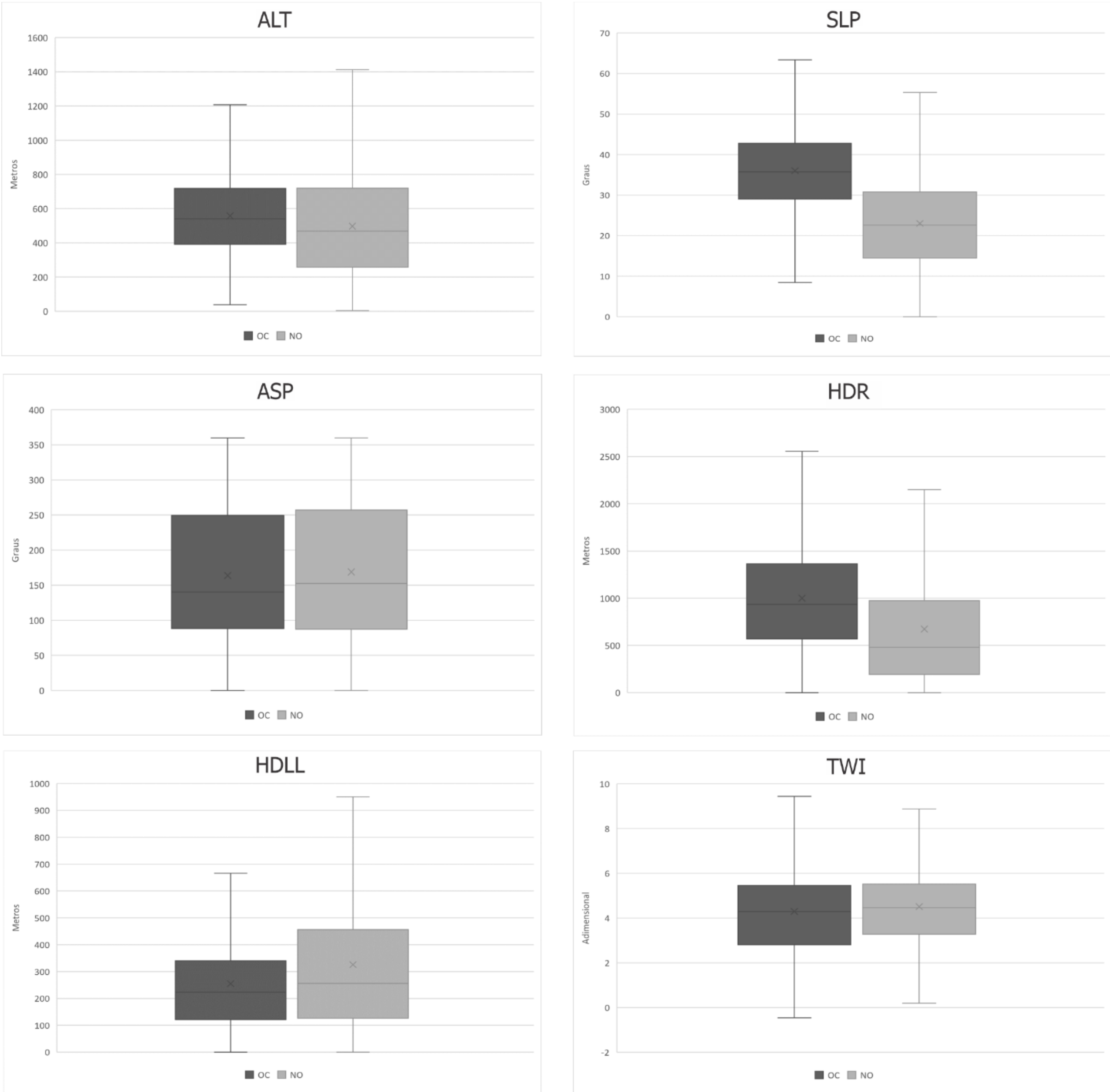

Figure 7. Boxplot graphs related to each morphometric data, divided into occurrence (OC) and non-occurrence (NO) areas. Legend: ALT is the altitude in meters; SLP is the slope in degrees; ASP is the Aspect in degrees; HDR is the horizontal distance from roads in meters; HDLL is the horizontal distance from lithostructural lineaments in meters; and TWI is the topographic moisture index (dimensionless). Source: Organized by the author.

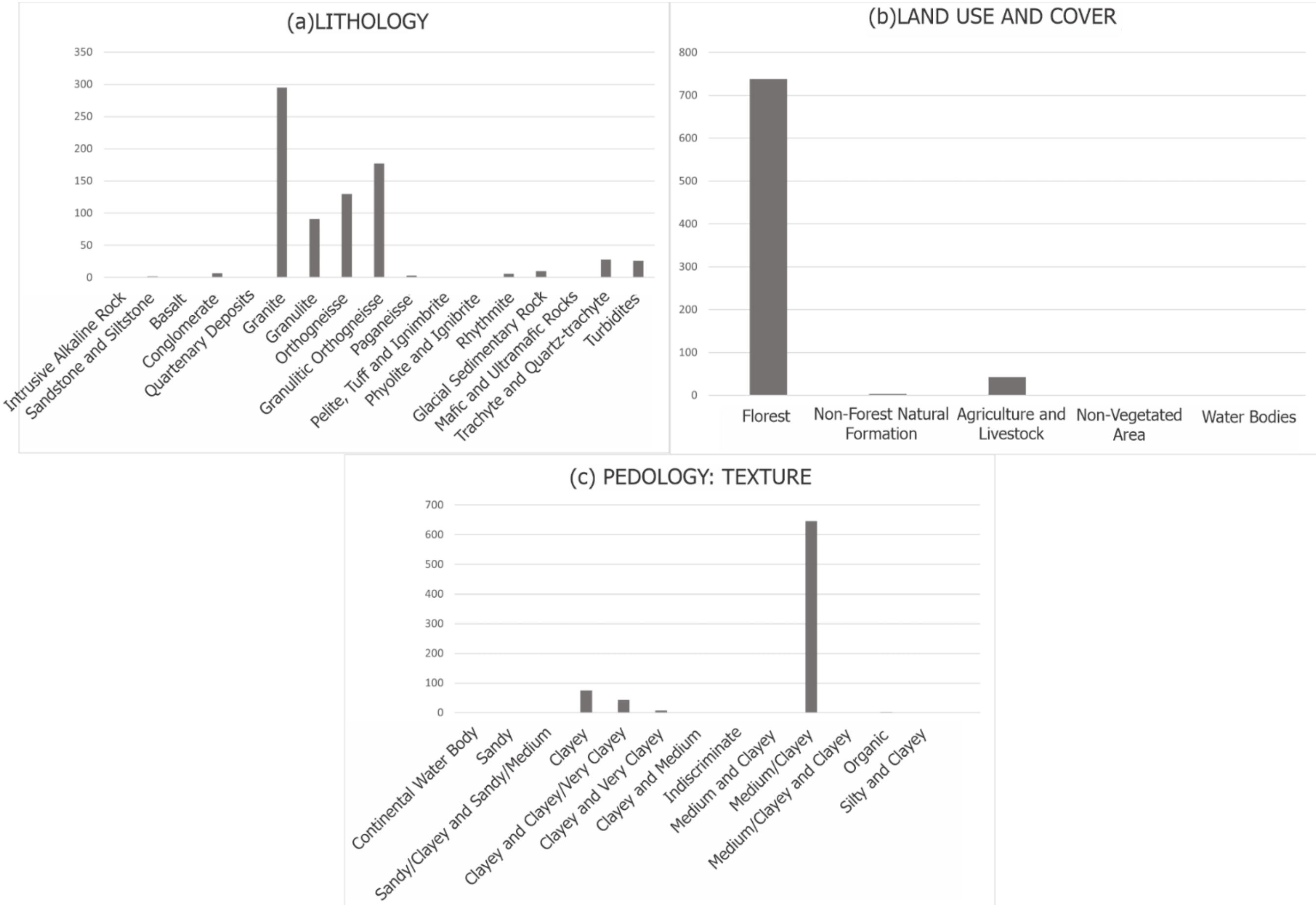

Figure 8 – Number of Landslide Occurrences versus Geoenvironmental Cartographic Data: (a) Number of occurrences x Lithology; (b) Number of occurrences x Land Use and Cover; and (c) Pedology**.**

The AUC and ACC values obtained in the five ANN simulations are presented in Table 2. The objective was to select models with an accuracy greater than 80% (PRADHAN; LEE, 2010; DOU et al., 2015), a criterion that all models exceeded. Although the simulations in this study showed approximately similar accuracy results, simulation ANN 3 achieved the best performance, with AUC values of 0.9 and ACC close to 84% for the municipality of Joinville. These results are similar to those found by Quevedo et al. (2019) in their model for the Rio Rolante Watershed (AUC > 0.9) and by Gameiro (2020) in models produced for the southern, central, and northeastern regions of the state of Santa Catarina (AUC between 0.87 and 0.93). This parity is significant, as these regions share very similar geomorphological conditions. The output products calculated by the ANN are shown in Figure 9, highlighting the main differences in the models calculated by the ANN. Figure 10 shows the total area of each susceptibility class for the five ANN simulations.

Table 2. Performance metric of the models for the different sample sets.

		ANN 1	ANN 2	ANN 3	ANN 4	ANN 5
AUC	**Training**	0.901	0.873	0.909	0.903	0.849
	Test	0.882	0.854	0.887	0.875	0.836
	Cross-validation	0.887	0.866	0.893	0.885	0.846
ACC	**Training**	0.848	0.807	0.856	0.843	0.782
	Test	0.823	0.789	0.829	0.814	0.768
	Cross-validation	0.828	0.803	0.838	0.822	0.776

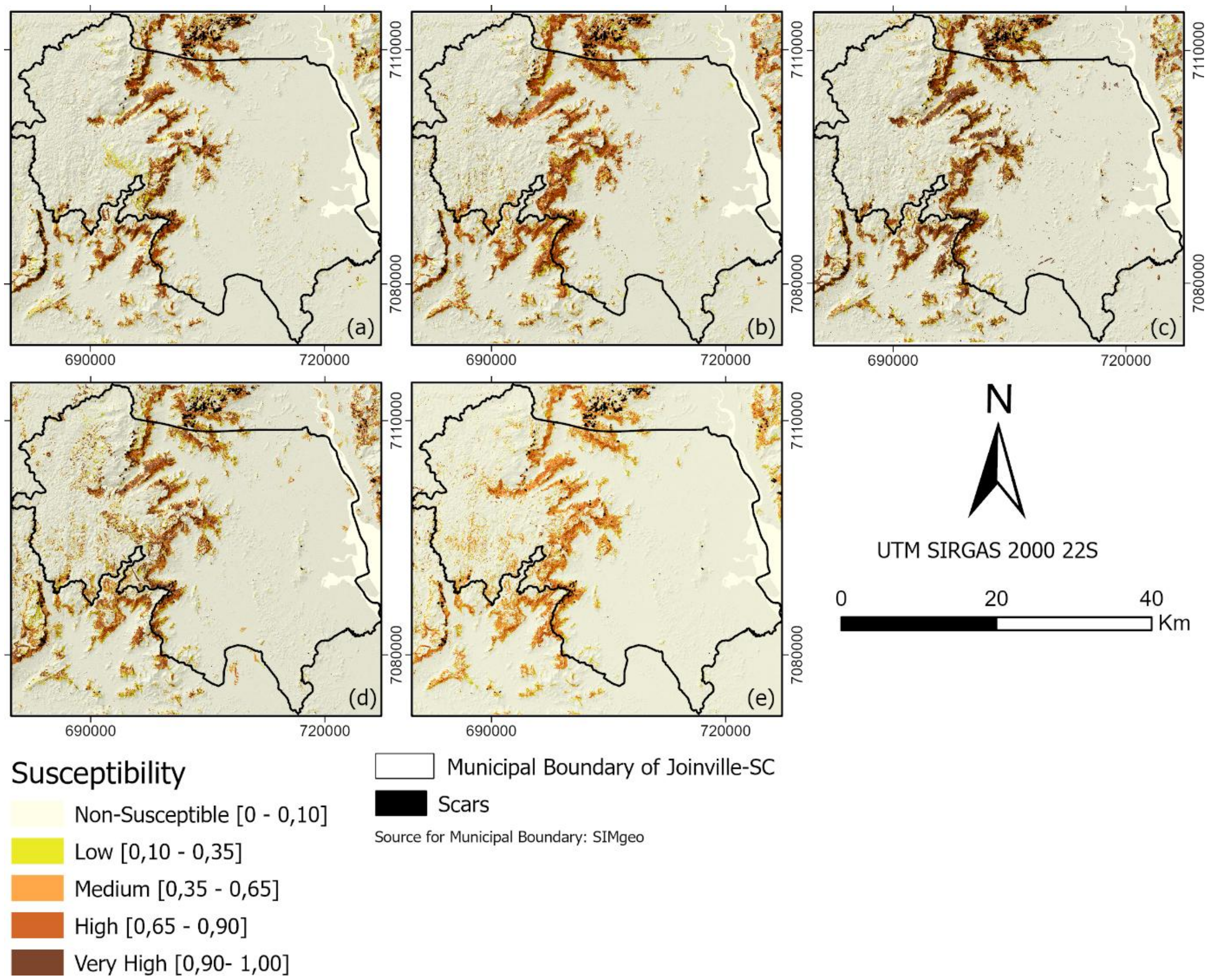

Figure 9. Output maps calculated by the ANN for each sample set – (a) ANN1: All input data for the model; (b) ANN2: Without lithology data; (c) ANN3: Without land use and cover data, showing the best accuracy results; (d) ANN4: Without soil data; and (e) ANN5: Without all qualitative geoenvironmental cartographic data. Source: Organized by the author.

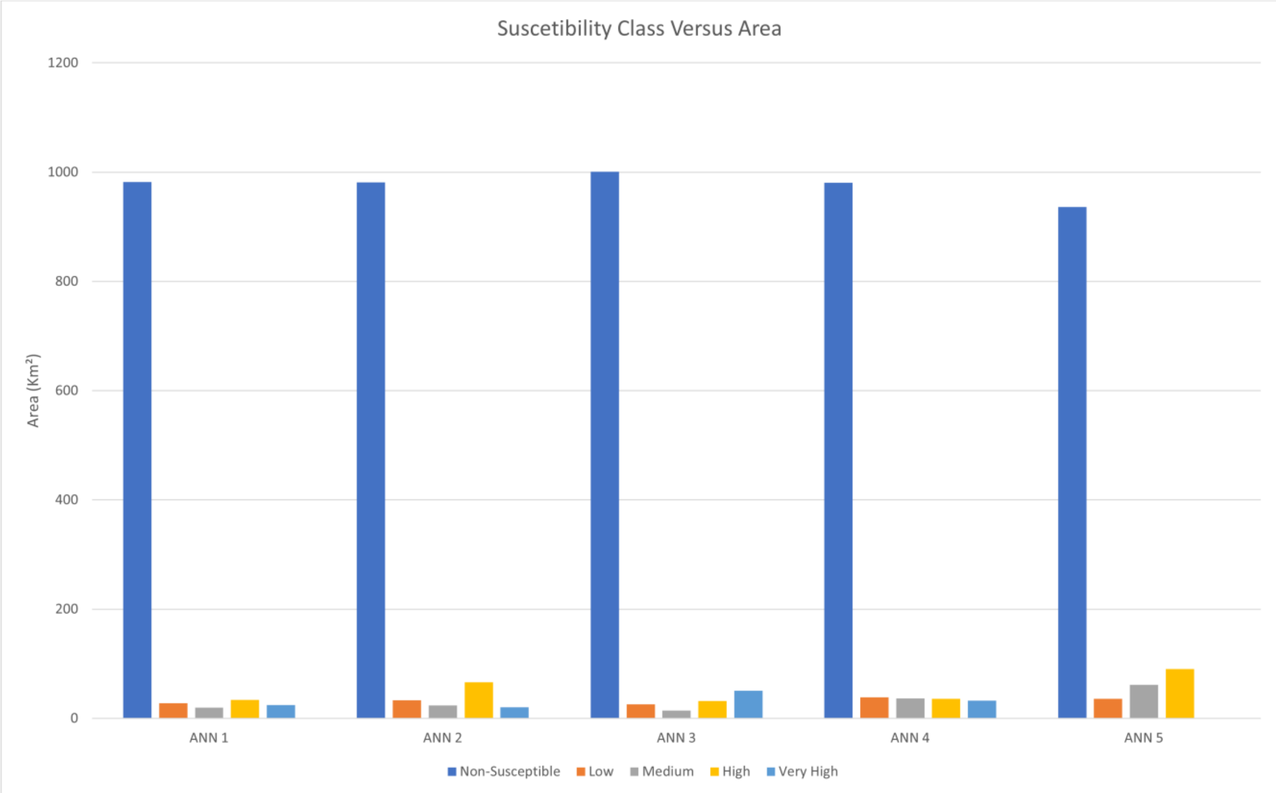

Figure 10. Total area of each susceptibility class for the five ANN simulations. In an initial analysis, it is observed that areas classified as not susceptible maintain similar values across all simulations, which can be attributed to the extensive flat regions of the municipality. The variations in the results are more pronounced in the classes susceptible to landslides.

The ANN 3 simulation, with the best accuracy results, was made with all variables except the land use and cover data. It is believed that this result is due to the higher concentration of landslides samples in the Forest class of the land use and cover attribute, as shown in Figure 8. In other words, this attribute is introducing a bias in the network, making it assume that where there is forest, there are occurrences, thereby reducing the weight of other factors, such as slope. For this reason, when this attribute is included in the network, the network's ability to generalize to other areas is diminished, such as the municipal urban center, which has areas of landslide occurrences and is included in the non-vegetated class of land use and cover. For the variables in the ANN 3 simulation, the relative importance percentages or Relative Contribution Index (RCI) were calculated according to Oliveira, Pedrollo, and Castro (2015) and are shown in Table 3.

Table 3. Relative importance of the variables used in the ANN 3 simulation**.**

Class	RCI Importance
LITOLOGY	33%
PEDOLOGY:TEXTURE	24%
ALTIMETRY	11%
SLOPE	11%
HORIZONTAL DISTANCE TO ROADS	10%
HORIZONTAL DISTANCE TO LITHOSTRUCTURAL LINEAMENTS	5%

When comparing the result maps produced by ANN 3 and ANN 1, which represent simulations with the highest and second-highest performance, respectively, and despite having similar accuracy results, it is visually apparent that ANN 1 tends to classify susceptible areas on the slopes of the urban center more conservatively. This difference in spatial pattern, even in cases of similar accuracy, has been evidenced in previous studies, such as those by Oliveira et al. (2019) and Gameiro (2020). Therefore, as emphasized by Gameiro (2020), the visual analysis of maps derived from empirical mathematical models, which are based on the relationships between input variables and expected output, plays a fundamental role.

The model resulting from ANN 4 simulation shows the third-best performance in terms of accuracy, displaying a more cautious tendency in categorizing the slopes in the eastern part of the municipality, when compared to the models generated by ANN s 1 and 3. Additionally, this model presents a higher amount of areas classified as susceptible at medium levels, compared to the two previously mentioned models. This variation in classification is likely due to the exclusion, in the model, of the separation between the classes of argisols and neosols. The latter class is characterized by shallow soils found in steeper terrain, which are directly associated with the predominant occurrences in the northern portion of the municipality.

Simulations ANN 2 and ANN 5 show lower accuracies than the others, especially ANN 5. Despite having AUC values above 0.8, ANN 5 shows an overall accuracy below this threshold. Both models were simulated without the geological framework, giving more weight to SLP and HDLL. A visual analysis of these results shows a reduced ability to generalize in areas with fewer occurrence samples, with a more conservative classification compared to the other models, emphasizing the medium susceptibility class. This highlights the importance of the geological substrate for a more precise weighting of the susceptibility model for Joinville.

To conclude the process of creating the susceptibility map for the municipality of Joinville, the image resulting from ANN 3 simulation, previously reclassified into 5 susceptibility classes, was converted to vector format using ArcGIS PRO 3.01 software (ESRI, 2023). To adjust the data to a 1:10,000 scale, polygons smaller than 10 km^2, or 1 ha, were removed, and a smoothing function was applied. The result of this procedure can be seen in Figure 11.

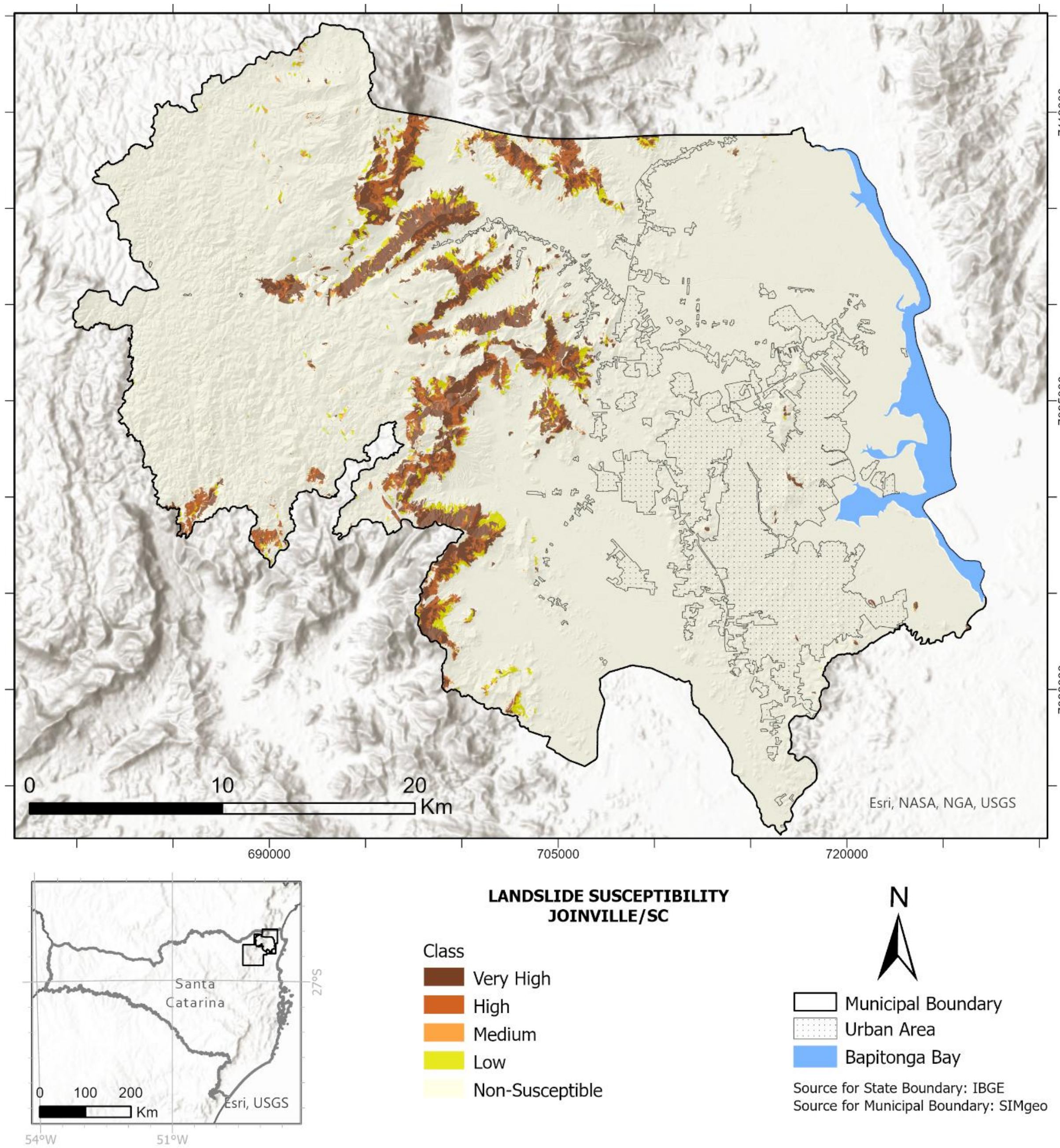

Figure 11. Landslide Susceptibility Map of the Municipality of Joinville-SC. Source: Organized by the author.

4. Conclusions

The results show that it was possible to obtain a landslide susceptibility model for the municipality of Joinville-SC, with satisfactory global accuracy and AUC indices, through the use of ANNs. Of all the models tested, simulation ANN 3 showed the best performance, with AUC values of 0.9 and ACC of 84%. Eighteen variables and 30 neurons in the hidden layer were used for the ANN application.

The analysis of morphometric variables shows a significant importance of the SLP and ASP data, indicating an important structural geological conditioning factor for the region. It is emphasized here that the SLP variable is likely the explanatory variable for the model proposed here. The HDR and HDLL data corroborate this analysis and demonstrate that landslides in the study area have low anthropogenic induction.

Qualitative data, when used, should be selected according to the context of landslides in the study area. In this work, it was found that the geological nature of the municipality has an important influence on the determination of land susceptibility. On the other hand, the use and occupation data caused a slight network bias, which slightly reduced accuracy and diminished the generalization ability for areas without landslide occurrence sampling. Furthermore, higher resolution data would lead to better results.

This study demonstrates that ANNs are a robust tool for landslide susceptibility assessment, helping in disaster prevention and sustainable urban planning. The model developed for the municipality of Joinville proved effective in identifying areas prone to landslides, providing information for territorial management and risk reduction. However, it is essential to validate the data published here in the field to ensure its effectiveness.

Author Contributions: Writing, sampling, data preparation, geoprocessing, maps, simulations, and analysis: R.R. Mendonça; ANN script for Matlab: G.G. Oliveira; Revisions: G.G. Oliveira and C.G. Tornquist.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the National Council for Scientific and Technological Development (CNPq) for financial support through Call CNPq No. 09/2022 - Research Productivity Scholarships - PQ, process 311009/2022-0, and Call CNPq/MCTI/FNDCT No. 18/2021 – Track A – Emerging Groups, process 408489/2021-9.

Conflict of Interest: The authors declare that there is no conflict of interest**.**

References

1. ALVALÁ, R.C.S.; BARBIERI, A. Desastres Naturais. In: NOBRE C.A.; MARENGO,J.A. Mudanças climáticas em rede: um olhar interdisciplinar. São José dos Campos, SP: INCT,2017. 608 p. ISBN 978-85-7917-463-6.
2. BITAR, O. Y. Cartas de suscetibilidade a movimentos gravitacionais de massa e inundações - 1:25.000: nota técnica explicativa. São Paulo: IPT; Brasília: CPRM, 2014. 42 p.
3. BRESSANI, L. A.; COSTA, E. A da. Mapeamento geotécnico: suscetibilidade, perigo, vulnerabilidade técnica, risco e risco instalado. In: CONGRESSO BRASILEIRO DE GEOLOGIA DE ENGENHARIA E AMBIENTAL, 14., 2013, Rio de Janeiro. Anais. Rio de Janeiro: ABGE, 2013.
4. BRAGAGNOLO, L.; DA SILVA, R. V.; GRZYBOWSKI, J. M. V. Artificial neural network ensembles applied to the mapping of landslide susceptibility. Catena, v. 184, p. 104240, 2020. DOI: https://doi.org/10.1016/j.catena.2019.104240.
5. CEMADEN - Centro Nacional de Monitoramento e Alerta de Desastres Naturais. Deslizamentos. Disponível em: http://www2.cemaden.gov.br/deslizamentos/. Acesso em: 15 de mar. de 2023.
6. CHEN, W., POURGHASEMI, H.R., PANAHI, M., KORNEJADY, A., WANG, J., XIE, X., CAO, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology 297, 69–85. 2017. https://doi.org/10.1016/j.geomorph.2017.09.007.
7. CPRM - SERVIÇO GEOLÓGICO DO BRASIL. Mapa geológico do estado de Santa Catarina. Porto Alegre: CPRM, 2014. Escala 1:500.000.
8. DELONG, E.R., DELONG, D.M., CLARKE-PEARSON, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845. 1988. DOI: https://doi.org/10.2307/2531595
9. DOU, J.; YAMAGASHI, H.; POURGHASEMI, H.R.; YUNUS, A.P.; SONG, X.; XU, Y. & ZHU, Z. An integrated artificial neural network model for landslide susceptibility assessment of Osado Island, Japan. Natural Hazards, 78: 1749-1776. 2015.DOI: https://doi.org/10.1007/s11069-015-1799-2
10. EMAMI, S. N., YOUSEFI, S., POURGHASEMI, H. R., TAVANGAR, S., & SANTOSH, M. A comparative study on machine learning modeling for mass movement susceptibility mapping (a case study of Iran). Bulletin of Engineering Geology and the Environment, 79, 5291-5308. 2020. DOI: https://doi.org/10.1007/s10064-020-01915-7
11. Esri Inc. Arcgis PRO (versão 3.0.1). Redlands, Estados Unidos, 2023.
12. FAWCETT, T. An introduction to ROC analysis. Pattern recognition letters, v. 27, n. 8, p. 861-874, 2006. https://doi.org/10.1016/j.patrec.2005.10.010
13. GAMEIRO, S. (2020). Capacidade de Generalização e extrapolação espacial de redes neurais artificiais no mapeamento a suscetibilidade de deslizamento. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul. Centro Estadual de Pesquisas em Sensoriamento Remoto e Metereologia. Porto Alegre. 85p.
14. GOMEZ, H.; KAVZOGLU, T.. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engineering Geology, v. 78, n. 1-2, p. 11-27, 2005. DOI: https://doi.org/10.1016/j.enggeo.2004.10.004
15. Google Inc. Google Earth Pro. Mountain View, California, Estados Unidos, 2023.

16. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Disponível em: https://cidades.ibge.gov.br/brasil/sc/joinville/panorama. Consultado em 02 de abril de 2022.
17. Instituto Brasileiro de Geografia e Estatística. Base de Solos. 2021. Disponível em: https://www.ibge.gov.br/geociencias/informacoes-ambientais/pedologia/10871-pedologia.html. Acesso em: maio de 2023.
18. KAWABATA, D., BANDIBAS, J. Geomorphology Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology 113 (1-2), 97–109. 2009. https://doi.org/10.1016/j.geomorph.2009.06.
19. LAMBERTY, D.; KEPEL FILHO, J. L.; NORONHA, F. L,. Carta de risco a movimentos de massa e inundação do município de Porto Alegre, RS: uma abordagem preliminar a partir dos dados de suscetibilidade e vulnerabilidade. In: CONGRESSO BRASILEIRO DE GEOLOGIA DE ENGENHARIA E AMBIENTAL, 15., Gramado, 2015. Anais... Gramado: ABGE, 2015. 1 CD ROM.
20. LUCCHESE, L. V.; DE OLIVEIRA, G.G.; PEDROLLO, O. C. Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using Artificial Neural Networks. Catena, v. 198, p. 105067, 2021. DOI: https://doi.org/10.1016/j.catena.2020.10506
21. LUCCHESE, L. V.; DE OLIVEIRA, G. G.; PEDRON, F. A.; SAMUEL-ROSA, A.; DALMOLIN, R. S. D. Variação das características pedológicas e classificação taxonômica de Argissolos derivados de rochas sedimentares. Revista Brasileira de Ciência do Solo, v. 36, p. 1-9, 2012. https://doi.org/10.1590/S0100-06832012000100001
22. LUO, W., LIU, C.-C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. Landslides 15 (3), 465–474. 2017.DOI: https://doi.org/10.1007/s10346-017-0893-9.
23. MAPBIOMAS. Disponível em: http://mapbiomas.org.Acesso em: 10 de março. 2023.
24. MINISTÉRIO DAS CIDADES. Capacitação em mapeamento e gerenciamento de risco. [S.l.,]. Disponível em: http://www.defesacivil.mg.gov.br/conteudo/arquivos/manuais/Mapeamento/mapeamento-grafica.pdf. Acesso em: 9 set. 2013.
25. OpenStreetMap contributors. OpenStreetMap. Disponível em: https://www.openstreetmap.org/. Acesso em: março de 2023.
26. ODEBRECHT, Edgar et al. Acontecimentos e consequências. Atuação da ABMS e exemplo de um Laudo de Diagnóstico e sua aplicação em Joinville/SC. ANAIS COBRAE, v. 1, p. 1–8, 2017. Disponível em: http://www.adfiducia.com.br/artigos/20170301-134757_modelo-artigo-cobrae-2013----artigo---v3.pdf. Acesso em: 05 setembro de 2022.
27. OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M.R. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. Engineering Applications of Artificial Intelligence, v. 40, p. 47-61, 2015.
28. OLIVEIRA, G. G.; RUIZ, L. F. C.; GUASSELLI, L. A.; HAETINGER, C. Random Forest and artificial neural networks in landslide susceptibility modeling: a case study of the Fão River Basin, Southern Brazil. Natural Hazards, 99, 1049, 2019. DOI: 10.1007/s11069-010-03795-x
29. PASCALE, Stefania et al. Landslide susceptibility mapping using artificial neural network in the urban area of Senise and San Costantino Albanese (Basilicata, Southern Italy). In: Computational Science and Its Applications–ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24-27, Proceedings, Part IV 13. Springer Berlin Heidelberg, 473-488. 2013.
30. PHAM, B.T., BUI, D.T., PRAKASH, I., DHOLAKIA, M. Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. Nat. Hazards 83, 97–127. 2016. https://doi.org/10.1007/s11069-016-2304-2
31. PHAM, B.T., BUI, D.T., PRAKASH, I., NGUYEN, L.H., DHOLAKIA, M. A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. Environmental Earth Sciences 76, 371. 2017. https://doi.org/10.1007/s12665-017-6689-3
32. PRADHAN, B.; LEE, S. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides, v. 7, n. 1, p. 13–30, 2010. DOI: https://doi.org/10.1007/s10346-009-0183-2
33. QGIS Development Team. QGIS Geographic Information System (versão 3.30). 2023. Disponível em: <http://qgis.osgeo.org>.
34. QUEVEDO, R. P., OLIVEIRA, G. G., GAMEIRO, S., RUIZ, L. F. C., & GUASSELLI, L. A. Modelagem de áreas suscetíveis a movimentos de massa com redes neurais artificiais. XIX SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2910-2913. 2019a.
35. QUEVEDO, R. P., GUASSELLI, L. A., DE OLIVEIRA, G. G., & RUIZ, L. F. C. Modelagem de Áreas Suscetíveis a Movimentos de Massa: Avaliação Comparativa De Técnicas De Amostragem, Aprendizado de Máquina e Modelos Digitais de Elevação. Revista Geociências, 38(3), 781-795. 2019b. https://doi.org/10.5016/geociencias.v38i3.14019

36. REICHENBACH, P., Rossi, M., MALAMUD, B., MIHIR, M., GUZZETTI, F. A review of statistically-based landslide susceptibility models. Earth Sci. Rev. 180, 60–91. 2018. DOI: https://doi.org/10.1016/j.earscirev.2018.03.001
37. RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. Nature, 323, 533-536, 1986. DOI: https://doi.org/10.1038/323533a0
38. SIG SC. Metadados Matriciais v1 [Recurso eletrônico]. Santa Catarina, Brasil: Secretaria de Estado da Administração, 2017. Disponível em: http://sigsc.sc.gov.br/1_Metadados_Matriciais_v1.pdf. Acesso em: 31 de jul. de 2023.
39. SOBREIRA, F. G.; SOUZA, L. A. de. Cartografia geotécnica aplicada ao planejamento urbano. Revista Brasileira de Geologia de Engenharia e Ambiental, v. 2, n. 1, p. 79-97, 2012.
40. TIEN BUI, D., TUAN, T.A., HOANG, N.-D., THANH, N.Q., NGUYEN, D.B., VAN LIEM, N., PRADHAN, B., Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14 (2), 447–458. 2017. DOI: https://doi.org/10.1007/s10346-016-0711-9
41. ULLAH, K., WANG. Y., ZHICE FANG, Z., WANG L., RAHMAN, M. Multi-hazard susceptibility mapping based on Convolutional Neural Networks, Geoscience Frontiers, Volume 13, Issue 5, 2022,101425, ISSN 1674-9871. DOI: https://doi.org/10.1016/j.gsf.2022.101425.
42. WANG, D., PU, R., GONG, P., YANG, R. Predicting forest yield with an artificial neural network and multiple regression. Chinese University of Hong Kong, Hong Kong. 771–780. 1995.
43. WANG, Y.; FANG, Z.; HONG, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the total environment, v. 666, p. 975-993, 2019. DOI: https://doi.org/10.1016/j.scitotenv.2019.02.263
44. WIDROW, B.; HOFF, M. E. Adaptive Switching Circuits. In: IRE WESCON CONVENTION RECORD, New York: IRE Part, 1960. 96– 104.
45. YAO, J., QIN, S., QIAO, S., LIU, X., ZHANG, L., & Chen, J. Application of a two-step sampling strategy based on deep neural network for landslide susceptibility mapping. Bulletin of Engineering Geology and the Environment, 81(4), 148. 2022. DOI: https://doi.org/10.1007/s10064-022-02615-0
46. ZÊZERE, J. L., PEREIRA, S., MELO, R., OLIVEIRA, S. C., & GARCIA, R. A. Mapping landslide susceptibility using data-driven methods. Science of the total environment, 589, 250-267. 2017. DOI: https://doi.org/10.1016/j.scitotenv.2017.02.188.
47. ZHU, A.X., MIAO, Y., WANG, R., ZHU, T., DENG, Y., LIU, J., HONG, H. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. CATENA 166, 317–327. 2018. DOI: https://doi.org/10.1016/j. catena.2018.04.003.