Machine learning-based method to map landslide susceptibility in urban areas

Francisco Monte Alverne de Sales Sampaio[1], Andrea Valli Nummer[2], Romário Trentin[2], João Paulo Delapasse Simioni[3], Luis Eduardo de Sousa Robaina[2]

[1] Santa Maria Federal University, Santa Maria RS, Brazil
[2] Santa Maria Federal University, Santa Maria RS, Brazil
[3] Rio Grande do Sul Federal University, Porto Alegre RS, Brazil
montealvernesampaio@gmail.com

Abstract
Landslides in urban areas can cause natural disasters with economic impacts and loss of life. The increase in these events is associated with population growth and disorderly urbanization. In Brazil, landslides resulted in 3,758 deaths between 1988 and 2022. Mapping susceptibility to planar landslides is vital for preventing and mitigating these disasters, and machine learning techniques, such as the Maximum Entropy Model (MAXENT), for example, have enabling the analysis and manipulation of large volumes of data, producing quick results with high levels of accuracy, becoming a valuable tool for minimizing damage in areas prone to landslides. This study sought to map the susceptibility to landslides in the urban area of Bento Gonçalves, RS, Brazil, using MAXENT. To this end, the work was divided into three stages: i) identification of scars for model training, ii) mapping of susceptibility to landslides and iii) validation of the model. To identify landslide scars, visual interpretation of satellite images made available by Planet Scope was used. To map susceptibility to landslides, the MAXENT machine learning model was used. The input data were points with landslide scars, visually identified through Planet Scope images. The results showed that MAXENT had an overall accuracy greater than 0.94 and the frequency ratio indicated a greater occurrence of scars in areas of high and very high susceptibility. Finally, the high potential of the MAXENT model for mapping susceptibility to landslides stands out.

Keywords
Landslide, Maxent, Urban area, Machine Learning.

1. Introduction
Landslides can cause natural disasters and are observed almost daily in various parts of the world, leading to significant economic losses and, in many cases, taking human lives (Huang; Zhao, 2018). In the last three decades, the number of natural disasters in various parts of the world has been increasing, mainly due to population growth, which has led to unplanned occupation, causing an intense process of urbanization and industrialization (Kobiyama et al., 2006; Robaina et al., 2010; Shah et al., 2022). In Brazil alone, a survey conducted by Macedo; Sandre (2022) found that 3,758 people lost their lives to landslides between 1988 and 2022. Furthermore, in May 2024, the State of Rio Grande do Sul experienced the largest climatic catastrophe in its history, with over 420 municipalities affected by floods or landslides. A survey conducted by the National Aeronautics and Space Act (NASA) identified more than 5,000 landslide scars in the state of Rio Grande do Sul in May 2024 alone. The municipality of Bento Gonçalves, located in the mountainous region of Rio Grande do Sul, recorded more than 140 landslides that resulted in 09 deaths and 05 missing persons.

Slides are landslides that occur on steeper slopes with higher speeds (m/h or even m/s) and are important in the evolution of slopes. They have a rupture surface that can be planar, circular (rotational), or wedge-shaped, following the structural weakness planes of earthen or rocky masses (Augusto Filho, 1993; IPT, 2014). The landslides discussed in this article are planar, typically occurring in shallow, younger soils or surface deposits on high-slope terrain. They have a relatively thin profile and a rectangular shape, with length much greater than width (Massad, 2010).

As a preventive method against landslides, various studies are conducted to map landslide susceptibility, such as Wicaksono et al. (2029); Kaminski (2020); Hana et al. (2022) Pathak & Devkota (2022) and in Brazil, Marcelino; Nunes & Kobiyama, 2006; Brasil, 2007 and Vanacôr & Rolim (2012); Dias et al. (2018), among others. These studies are used as the first preventive measure to reduce the impact of these phenomena in affected communities (Brasil, 2007).

Especially since the 2000s, machine learning techniques have enabled the analysis and manipulation of large volumes of data, producing rapid results with high accuracy levels. Numerous studies have applied machine learning for landslide susceptibility mapping. Quevedo et al. (2019) achieved 94% accuracy by applying machine learning to map landslide-susceptible areas in the Rolante River basin, RS, Brazil. Kornejady, Ownegh, and Bahremand (2017) and Ngo et al. (2021) applied machine learning-based techniques to map landslide susceptibility in India and Iran, respectively, achieving accuracies above 90%.

Currently, there are various machine learning models used in diverse knowledge areas. The Maximum Entropy Model (MAXENT), for example, is a machine learning model that quantitatively calculates the probability of specific events occurring based on environmental indicators (Golkarian & Rahmati, 2018). One advantage of MAXENT is that the model allows for the addition of various variables as input layers, such as land use, precipitation, lithology, pedology, etc. MAXENT is widely employed in ecological and diversity models; however, recent studies show the model's efficacy for landslide susceptibility mapping (Shi, 2022). The general objective of this study is to map landslide susceptibility in the urban area of Bento Gonçalves, RS, using machine learning techniques.

1.1 Study area

This study was conducted in the urban area of the municipality of Bento Gonçalves, located in the mountainous region of the State of Rio Grande do Sul (Figure 1). Bento Gonçalves has a population of 123,151 inhabitants and ranks among the ten largest economies in Rio Grande do Sul, with notable sectors including furniture, wine, metallurgy, transportation, and fruit production (IBGE, 2022). The northern part of the municipality is bordered by the Taquari-Antas River and the eastern part by the Barracão Stream. Altitudes range from 73 to 759 meters, with an average altitude of 626 meters in the urban area. According to INMET (2024), average temperatures range between 27.3°C in the hottest month (January) and 8.5°C in the coldest month (July). Average precipitation ranges from 117.8 mm in the least rainy month (March) to 185.6 mm in the rainiest month (October).

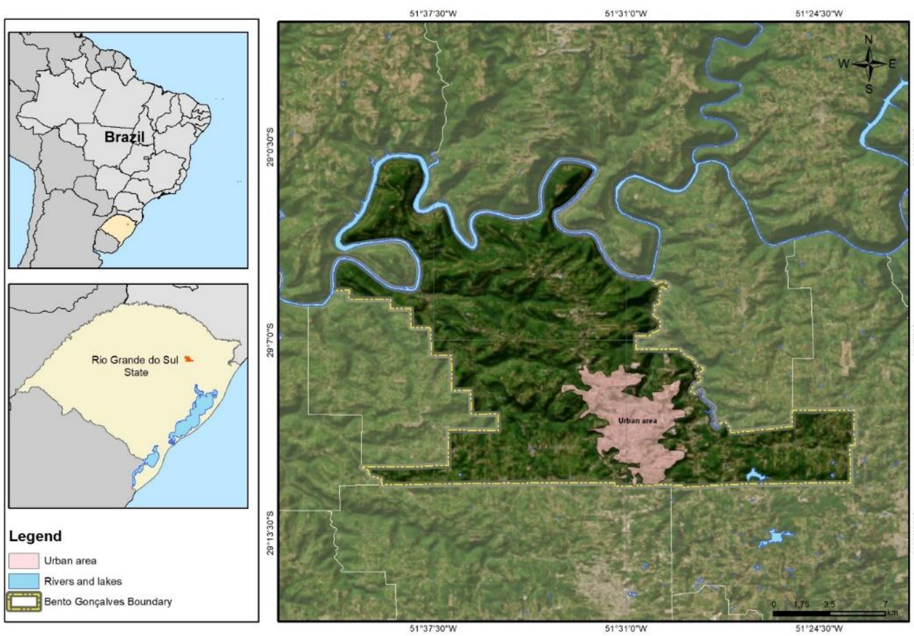

Fig. 1. Location of the municipality of Bento Gonçalves.

The southern region of Brazil does not have seasonality regarding rainfall behavior, and there is a probability of extreme rainfall occurring at any time of the year (Barbieri, 2005; Franchito et al., 2008). The work of Teixeira and Prieto (2020) evaluated extreme rainfall events in the state of Rio Grande do Sul, which were classified into two categories: daily and persistent. Daily extreme events are more frequent in spring and summer, while persistent extreme events are more frequent in winter. The spatial distribution of rainfall, in the case of persistent extreme events, showed that the eastern half of the state, where the municipality of Bento Gonçalves is located, has a higher frequency of persistent rainfall events with a large amount of rainfall.

2. Materials and Methods

This research was divided into three stages: i) acquisition of landslide scar samples; ii) definition of the layers used in the study; and iii) landslide susceptibility mapping. The acquisition of landslide scar samples was based on the visual interpretation of Planet Scope satellite images dated June 10, 2024. For this, a meticulous work of identifying landslide

2.

scars that occurred in Bento Gonçalves in May 2024 was carried out. For each landslide scar, a point was created at the rupture surface, and the geographical coordinates of the location were obtained (Figure 2). A total of 36 landslide scars were identified in Bento Gonçalves. These points were used as training samples and subsequent validation in the MAXENT machine learning model. Thus, the model was divided into 70% samples for training and 30% for validation. The layers used in the study were slope (SRTM 30 meters), distance to highways, distance to first-order rivers, distance to lineaments, and slope shape.

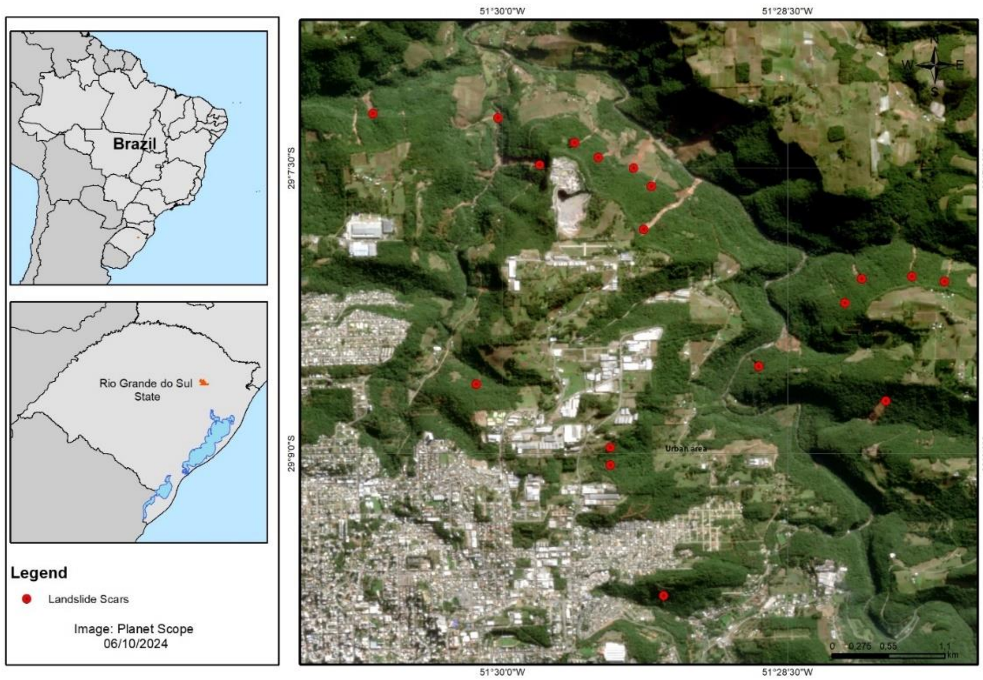

Fig. 2. Example of the identification of landslide scars used for training and validation.

The choice of these layers was based on their widespread use in the literature. Slope, for example, was described by Brito et al. (2016) as the most relevant factor for terrain predisposition to landslides in a study conducted in the municipality of Porto Alegre. According to the authors, slope is directly proportional to the speed of movement and, therefore, to the soil's transport capacity.

First-order rivers were evaluated by Vanacôr and Rolim (2012) in a study on landslide susceptibility in the northeastern region of the state of Rio Grande do Sul. The authors explain that these areas, where surface runoff begins to develop, are high-moisture areas and, under conditions of heavy rainfall, can lead to soil saturation and movement.

In addition to first-order rivers, Vanacôr and Rolim (2012) identified that the distance to highways is also an important determinant of landslides, as the road system modifies the natural water flow patterns and the slope shape through cuts and embankments. Regarding lineaments, Ramli et al. (2010) mention that this factor constitutes preferential paths for water percolation and plays an important role in rock weathering by forming planes of weakness, contributing to the reduction of resistance parameters. Pimentel and Bricalli (2023) evaluated the relationship between the density of lineaments and the occurrence of landslides in the municipality of Vitória-ES and found a strong association between the density of geological structures and the occurrence of landslides.

Finally, slope shape plays a crucial role in predisposing to landslides, influencing stress distribution, drainage, and slope stability. Ohlmacher (2007) identified a strong correlation between slope shape and landslide occurrence. According to the author, slopes with planar curvature are the most susceptible to landslides, with concave slopes being slightly more susceptible than convex slopes.

To map landslide susceptibility in the urban area of Bento Gonçalves, the MAXENT machine learning model was used (Phillips et al., 2006). MAXENT is a machine learning model that quantitatively calculates the probability of occurrence of a given event in a study area using Bayesian rule based on environmental indicators, which in this case are the variables: slope, first-order rivers, roads and highways, lineaments, and slope shape (Rahmati et al., 2016). Kerekes, Poszet, and Gál (2018) emphasize that MAXENT works with three main inputs:

2.

i) The location of known presence points.
ii) A study area.
iii) Explanatory variables, or covariates, that describe the environmental factors that may be related to presence in the study area (Kerekes, Poszet, and Gál, 2018).

The MAXENT model was applied using ArcGIS Pro 3.2 software. For model validation, cross-validation was used with the k-fold method (Burman, 1989). In k-fold cross-validation, the scar points are divided into k data subsets (in this study, 3 groups were used). The model then trains on two groups, minus one (k-1) of the data sets, and subsequently evaluates the model on the data set not used for training. This process is repeated 3 times, with a different subset reserved for evaluation (and excluded from training) each time. The model also presents the regression coefficient of all explanatory variables used.

3. Results

The results of the MAXENT machine learning model showed that the omission rate, i.e., the proportion of cases where a system fails to detect or report events, was 0.2773. The omission rate refers to the proportion of locations where landslide scars were verified but were not correctly predicted by the model as areas of very high susceptibility. It is worth noting that a low omission rate (less than 0.30) indicates that the model has a good ability to predict a given phenomenon, while a high omission rate (> 0.70) suggests that the model may be underestimating the occurrence areas or not adequately capturing the environmental factors that influence landslides.
The overall accuracy was 0.9403. This measure evaluates the model's overall precision in correctly predicting areas with higher landslide susceptibility. This measure is calculated by comparing the model's predictions with the observed occurrence data.
The susceptibility map (Figure 3) of the urban area of Bento Gonçalves was divided into five different classes according to the probability of landslide occurrence. The classification intervals of the landslide susceptibility map are shown in Table 1.

Table 1. Landslide Susceptibility Classes

	Very Low	**Low**	**Moderate**	**High**	**Very High**
Classes	0 – 0.2	0.2 – 0.4	0.4 – 0.6	0.6 – 0.8	> 0.8

It was found that 85.4% of the urban area of Bento Gonçalves is located in areas with low or very low landslide susceptibility. These areas have gently undulating terrain, with few slopes and intense urban occupation, and are distant from first-order rivers, highways, and lineaments. According to IBGE (2024), this area contains 48,654 buildings, which corresponds to 89% of all urban occupations in the municipality.
The moderate class corresponds to 9.1% of the urban area and is located in areas with undulating terrain near slopes, where first-order rivers are present. This class covers 3,862 buildings, which corresponds to 7% of the buildings in the urban area of the municipality. Finally, 5% of the urban area is located in high or very high landslide susceptibility zones, encompassing a total of 2,087 buildings. These areas are predominantly located on the edges of the urban area, where the terrain ranges from strongly undulating to steep.

2.

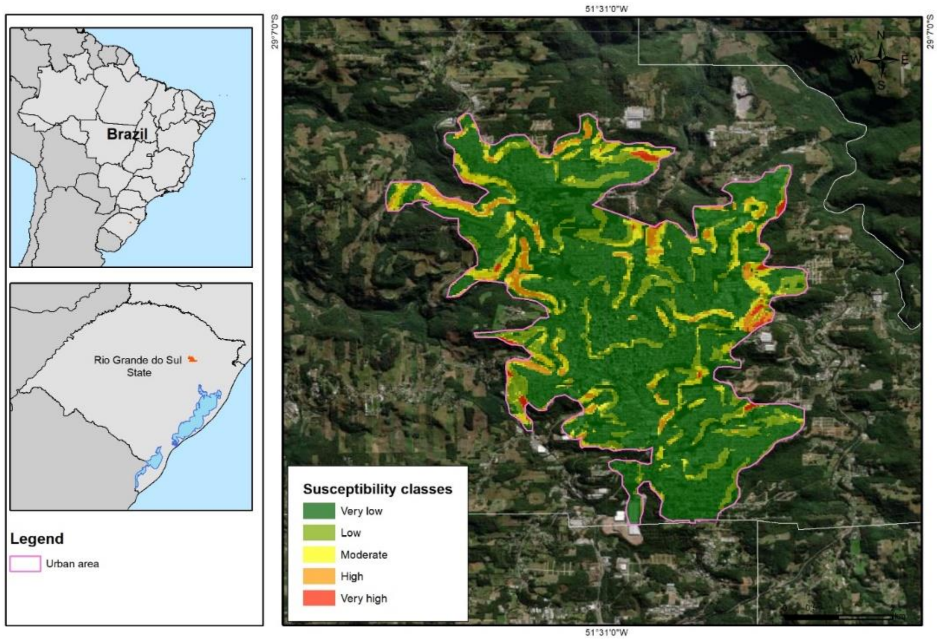

Fig. 3. Landslide Susceptibility in the Urban Area of Bento Gonçalves, RS, Using Machine Learning.

4. Discussion

First, regarding landslide susceptibility in the urban area of Bento Gonçalves, it was found that 5.1% of the urban area of the municipality is located in areas with high or very high landslide susceptibility. By cross-referencing data from these areas with the number of buildings obtained through the IBGE Demographic Census (2024), it is evident that 2,087 buildings are located in areas with high or very high landslide susceptibility.

A study by the Geological Survey of Brazil (Bellettini; Silva, 2015) identified six areas with high and very high landslide susceptibility in the urban area of Bento Gonçalves. The total area of these six zones, as delineated by the Geological Survey of Brazil (Bellettini; Silva, 2015), is 8.98 hectares. When comparing these areas with the landslide-susceptible zones identified using MAXENT, a 77.6% overlap was observed, thus demonstrating the high potential of the MAXENT machine learning algorithm in delineating areas susceptible to landslides.

The application of MAXENT for landslide susceptibility mapping is still a novelty in the methodological field (Boussouf; Fernández; Hart, 2023). A literature search on the topic reveals that even a review of the literature on machine learning and landslides by Ado et al. (2022) does not include the MAXENT approach in its discussions. This is evident in Tehrani et al. (2022), who conducted a study demonstrating recent advances in the application of machine learning for landslide mapping, excluding the MAXENT model.

A search in major scientific repositories (Science Direct, Scopus, and Scielo) found only six articles involving the MAXENT model for landslide susceptibility mapping. In one of the first studies using MAXENT for landslide mapping, Yuan et al. (2017) achieved an overall accuracy of 84.5% when applying the model in Zhaoqing (China). The difference in accuracy compared to our study may be attributed to the peculiarities of the study areas or the different layers used in the model. In addition to slope, slope shape, distance to lineaments, distance to highways, and distance to first-order rivers, Yuan et al. (2017) also used the Normalized Difference Vegetation Index (NDVI), precipitation, and lithology.

Our accuracy results are similar to those found by Boussouf; Fernández and Hart (2023), who applied MAXENT to map landslide susceptibility in Almería, Spain, achieving an overall accuracy above 90%. The study by Boussouf; Fernández and Hart (2023) demonstrated high precision, highlighting the importance of factors such as slope, elevation, and proximity to geological faults as major contributors to landslide susceptibility.

5. Conclusions

This study aimed to evaluate the application of the MAXENT machine learning algorithm for mapping landslide susceptibility in the urban area of the municipality of Bento Gonçalves. Our results highlighted the high potential of

2.

MAXENT, achieving accuracies above 0.94. Regarding the areas of the municipality, it is noteworthy that most of the urban area of Bento Gonçalves (85.4%) is located in areas of low or very low landslide susceptibility. These areas also have the highest urban occupation. Additionally, 9.1% of the municipality's urban area is located in areas classified as having moderate landslide susceptibility. Finally, 5% of the urban area of Bento Gonçalves is located in areas with high or very high landslide susceptibility. These areas contain 2,087 buildings, which therefore require special attention from public authorities.

The findings of this study underscore the importance of considering landslide susceptibility in decision-making related to urban planning and environmental risk management. It is essential to adopt appropriate prevention and mitigation measures in the areas identified as high risk, especially those with urban occupations. Furthermore, the results of this study can provide support for the development of public policies and strategies for environmental protection, ensuring the safety of communities and the preservation of natural resources in the urban area of Bento Gonçalves.

References

Ado, M., Amitab, K., Maji, A.K., Jasi ´Nska, E., Gono, R., Leonowicz, Z., Jasi ´Nski, M. (2022). Landslide Susceptibility Mapping Using Machine Learning: A Literature Survey. *Remote Sens*., 14. https://doi.org/10.3390/rs14133029.

Augusto Filho, O. 1993. *Escorregamentos em Encostas Naturais e Ocupadas: análise e controle*. In: Curso de Geologia Aplicada ao Meio Ambiente. São Paulo: IPT - DIGEO.

Barbieri, P.R.B (2005). *Caracterização da Estação Chuvosa nas Regiões Sul e Sudeste do Brasil Associado com a Circulação Atmosférica*. Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

Bellettini, A. da. S. & Silva, D. R. A. (2015). *Ação emergencial para delimitação de áreas em alto e muito alto risco a enchentes, inundações e movimentos de massa: Bento Gonçalves, RS.* CPRM. https://rigeo.cprm.gov.br/jspui/handle/doc/20089.

Boussouf, S., Fernández, T. & Hart, A. B. (2023). Landslide susceptibility mapping using maximum entropy (MaxEnt) and geographically weighted logistic regression (GWLR) models in the Río Aguas catchment (Almería, SE Spain). *Natural Hazards*, 117, 207–235. https://doi.org/10.1007/s11069-023-05857-7

Brasil. (2007). *Mapeamento de riscos em encostas e margens de rios*. Ministério das Cidades, Instituto de Pesquisas Tecnológicas – IPT.

Brito, M. M., Weber, E. J., Krigge, V. S., Leitzke, F. P. (2016). Analysis of Landslide Conditioning Factors in Porto Alegre Municipality based on Historical Data. *Revista Brasileira de Cartografia*, 68(1), 1853–1872.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross validation and the repeated learning-testing methods. *Biometrika*. 76, 503–514.

Dias, T.N. (2018). Avaliação da Suscetibilidade a Escorregamentos Rasos com Base na Aplicação de Estatística Bivariada: Resultados Preliminares. *Revista do Departamento de Geografia*, 35, 34-42.

Franchito, S.H., Rao, V.B., Barbieri, P.R.B, Santo, C.M. (2008). Rainy-season duration estimated from OLR versus rain gauge data and the 2001 drought in Southeast Brazil. *Journal of Applied Meteorology and Climatology*, 47(5), 1493-1499.

Golkarian, A. & Rahmati, O. (2018). Use of a maximum entropy model to identify the key factors that influence groundwater availability on the Gonabad Plain, Iran. *Environ Earth Sci*, 77(369). https://doi.org/10.1007/s12665-018-7551-y

Huang, Y. & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena*, 165, 520–529.

2.

IBGE - Instituto Brasileiro de Geografia e Estatística. (2024). *IBGE Cidades: Bento Gonçalves*. Government of Brazil. https://cidades.ibge.gov.br/.

IBGE - Instituto Brasileiro de Geografia e Estatística. (2024). *Coordenadas dos Endereços do Censo Demográfico de 2022*. Government of Brazil. https://censo2022.ibge.gov.br/apps/pgi/#/home/.

INMET - Instituto Nacional de Meteorologia. (2024). *Banco de Dados Meteorológicos*. Government of Brazil. https://bdmep.inmet.gov.br/

IPT - Instituto de Pesquisas Tecnológicas do Estado de São Paulo. (2014). *Cartas de suscetibilidade a movimentos gravitacionais de massa e inundações*. São Paulo: IPT.

Kaminky, M. (2020). The Impact of Quality of Digital Elevation Models on the Result of Landslide Susceptibility Modeling Using the Method ofWeights of Evidence**.** *Geosciences*, 10(488), 1-21. doi:10.3390/geosciences10120488

Kerekes, A., Poszet, S. L., & Gál, A. (2018). Landslide susceptibility assessment using the maximum entropy model in a sector of the Cluj–Napoca Municipality, Romania. Revista De Geomorfologie, 20, 130-146. https://doi.org/10.21094/rg.2018.039

Kobiyama, M. et al. (2010). *Prevenção de desastres naturais: conceitos básicos*. LABCIMA/UFPR. http://www.labclima.ufsc.br/files/2010/04/Livro_Prevencao_de_Desastres_Naturais.pdf

Kornejady, A., Ownegh, M. & Bahremand, A. (2017). Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena*, 152, 144-162. https://doi.org/10.1016/j.catena.2017.01.010

Macedo, E. S. de & Sandre, L. H. (2022). Mortes por deslizamentos no Brasil: 1988 a 2022. *Revista Brasileira de Geologia de Engenharia e Ambiental*, 12(1), 110-117.
Marcelino, E. V., Nunes, L. H. & Kobiyama, M. (2006). Mapeamento De Risco De Desastres Naturais Do Estado De Santa Catarina. *Caminhos de Geografia*, 7(17), 72–84.

Massad, F. (2010). *Obras de Terra: curso Básico de Geotecnia*. 2. ed. São Paulo: Oficina de Textos.

Ngo, T. Q. et al. (2021). Landslide Susceptibility Mapping Using Single Machine Learning Models: A Case Study from Pithoragarh District, India. *Advances in Civil Engineering*, 1, 1-19. https://doi.org/10.1155/2021/9934732

Ohlmacher, C. (2000). The Relationship Between Geology and Landslide Hazards of Atchison, Kansas, and Vicinity. [Report]. Kansas Geological Survey.
Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(4), 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pathak, L. & Devkota, K. C. (2022). Landslide susceptibility assessment in the Rangun Khola watershed of far western Nepal. *Journal of Nepal Geological Society, 63, 33–44.* https://doi.org/10.3126/jngs.v63i01.50827

Pimentel, T. B. & Bricalli, L. L. (2023). The Relationship between Lineament Patterns and Mass Movements in the Municipality of Vitória (Espírito Santo, Southeast Brazil). *Sociedade & Natureza*, 35(1).

Rahmati, O., Reza, H., Assefa, P. & Melesse, M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *CATENA*, 137, 360-372.

Ramli, M. F. et al. (2010). Lineament mapping and its application in landslide hazard assessment: a review. *Bull Eng Geol Environ*, 69, 215–233. https://doi.org/10.1007/s10064-009-0255-5

Rana, K., Malik, N. & Ozturk, U. (2022). Landsifier v1.0: a Python library to estimate likely triggers of mapped landslides. *Nat. Hazards Earth Syst. Sci*., 22, 3751–3764.

2.

Robaina, L. E. D. S. et al. (2010). Análise espaço-temporal das ocorrências de inundações e movimentos de massa no município de Caxias do Sul, RS. *Ciência e Natura*, 32(1), 159–172.

Shah, B. et al. (2022). Developing landslide hazard scenario using the historical events for the Kashmir Himalaya. *Nat Hazards*, 114, 3763–3785. https://doi.org/10.1007/s11069-022-05542-1
Shi, X. D., Yin, Q., Sang, Z. Y., Zhu, Z. L., Jia, Z. K., Ma, L. Y. (2022). Prediction of potentially suitable areas for the introduction of Magnolia wufengensis under climate change. *Ecol. Indic*., 1(127).

Tehrani, F. S., Calvello, C., Liu, Z., Zhang, L. & Lacasse, S. (2022). Machine learning and landslide studies: recent advances and applications. *Natural Hazards*, 114, 1197–1245. https://doi.org/10.1007/s11069-022-05423-7.

Teixeira, M. S. & Prieto, R. B. (2020). Eventos Extremos de Chuva no Estado do Rio Grande do Sul, Brasil, entre 2004 e 2013. *Revista Brasileira de Meteorologia*, 35, 45-52. http://dx.doi.org/10.1590/0102-7786351027

Vanacôr, R. R. N. & Rolim, S. S. B. A. (2012). Mapeamento da suscetibilidade a deslizamentos usando técnicas de estatística bivariada e sistema de informações geográficas na Região Nordeste do Rio Grande do Sul. *Revista Brasileira de Geomorfologia*, 13, 15–28.

Wicaksono, Y. S., Sihombing, F.M.H. & Indra, T.L. (2020). Landslide susceptibility map of Bogor Area using analytical hierarchy process. *IOP Conf. Series: Earth and Environmental Science*, 538**.** 10.1088/1755-1315/538/1/012050

Yuan, S., Huang, G., Xiong, H., Gong, Q., Chen, J. & Wang, J.(2021). Maximum Entropy-Based Model of High-Threat Landslide Disaster Distribution in Zhaoqing, China. *Journal of Risk Analysis and Crisis Response*, 7(3).