

FLORESTA ALEATÓRIA APLICADA AO MAPEAMENTO DE SUSCETIBILIDADE A MOVIMENTO DE MASSA NA BACIA HIDROGRÁFICA DO RIO PALENA, CHILE

Francisco Helter Fernandes do Amaral¹, Renata Pacheco Quevedo², Edson Luís Piroli³

¹ Mestrando do Programa de Pós-graduação em Geografia da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, helter.fernandes@unesp.br; ² Programa de Pós-Graduação em Sensoriamento Remoto - PGSER, Divisão de Observação da Terra e Geoinformática - DIOTG, Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos 12227-010, Brasil, renata.quevedo@inpe.br ; ³ Professor adjunto do Programa de Pós-graduação em Geografia da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, edson.piroli@unesp.br

RESUMO

O mapeamento de suscetibilidade a deslizamentos de terra são uma importante ferramenta para a redução de risco de desastres. Apesar da disseminação de pesquisas utilizando métodos de aprendizagem de máquina para mapeamento de deslizamentos de terra ao redor do mundo, poucos foram desenvolvidos para a América do Sul. Assim, o presente estudo teve como objetivo a modelagem de suscetibilidade a deslizamentos de terra na bacia hidrográfica do rio *Palena*, localizado no sul do Chile, a partir do modelo Random Forest. Com base nos resultados, foi observado que o modelo alcançou um nível de excelência em seu desempenho (AUC = 0.9937). Além disso, o mapeamento permitiu verificar uma maior presença de áreas altamente suscetíveis junto às zonas de maior declividade, a montante da bacia. Por fim, pode-se inferir que o RF apresenta potencial para prever a suscetibilidade a deslizamentos de terra.

Palavras-chave — Aprendizagem de máquina, desastre natural, SIG, modelagem espacial, risco.

ABSTRACT

Landslide susceptibility mapping is an important tool for disaster risk reduction. Despite the dissemination of research using machine learning methods to map landslide susceptibility around the world, few of them were developed for South America. Thus, the present study aimed to model and map landslide susceptibility in the Palena River basin, located Southern Chile, using the Random Forest model. According to the results, it was observed that the generated model reached a level of excellence in its performance (AUC = 0.9937). Moreover, the mapping showed a greater presence of high and very high landslide susceptible areas along the steepest areas, mainly in the upstream portion of the basin. Finally, we could infer that the RF has the potential to predict landslide susceptibility.

Key words — Machine learning, natural disaster, GIS, spatial modeling, risk.

1. INTRODUÇÃO

Os deslizamentos de terra compreendem o deslocamento de rochas, lama, detritos ou solos situados ao longo das encostas^[1]. A velocidade e o volume de massa deslocada tornam os deslizamentos de terra um dos maiores perigos geológicos no mundo, sendo estes os principais desafios para a previsibilidade deste fenômeno^[1].

A precipitação é um dos principais fatores desencadeadores de deslizamento de terra em regiões montanhosas. Além desses, outros processos de matriz endógena como terremotos, maremotos e erupções vulcânicas, podem atuar como catalizadores.

Todos os anos são registrados diversos eventos de deslizamento de terra em várias partes do mundo, sendo responsáveis por gerar danos materiais significativos, impactando diretamente a economia local, além do registro de perdas da vida^[9]. Entre os anos 1995 e 2014, foram registradas mais de 3.876 ocorrências de deslizamento de terra em todo o mundo que acarretaram 163.658 mortes^[1]. Além disso, entre 1993 e 2002 os deslizamentos de terra foram responsáveis por 17% das mortes relacionadas à ocorrência de desastres causados por fenômenos naturais, provocando mais de 940 fatalidades por ano^[2].

Causados por desastres causados por fenômenos naturais No Chile, possui registros de deslizamento de terra ao longo das últimas seis décadas que provocaram mais de 251 mortes^[7], devido à sua localização na Cordilheira dos Andes. Como exemplo, em 16 de dezembro de 2017, a *villa Santa Lucía*, no sul do país, foi atingida por um deslizamento que provocou 22 mortes e destruiu mais de 50 residências^[10].

Nesse sentido, modelos capazes de representar espacialmente as feições de maior suscetibilidade a deslizamento de terra são fundamentais para auxiliar a formulação de políticas públicas de planejamento territorial e de redução de risco de desastres^[11].

Assim, o objetivo do presente estudo foi gerar um mapeamento de suscetibilidade a deslizamentos de terra para a bacia do rio *Palena*, utilizando o modelo de aprendizagem de máquina Floresta Aleatória (*Random Forest* – RF).

2. MATERIAL E MÉTODOS

2.1. Caracterização da área de estudo

A área de estudo encontra-se localizada entre as longitudes 70 865 900 E a 71 468 900 E e as latitudes 51 984 7800 S a 50 989 9400 S, compreendendo a bacia hidrográfica do rio *Palena*, região sul do Chile, a qual faz fronteira a leste com a Argentina. A bacia hidrográfica do rio *Palena* abarca as províncias de *Palena*, *Aysén* e *Coihaique*. A altitude varia do nível do mar a 2085 m (Figura 1), e declividade varia entre 0 a 87 (graus).

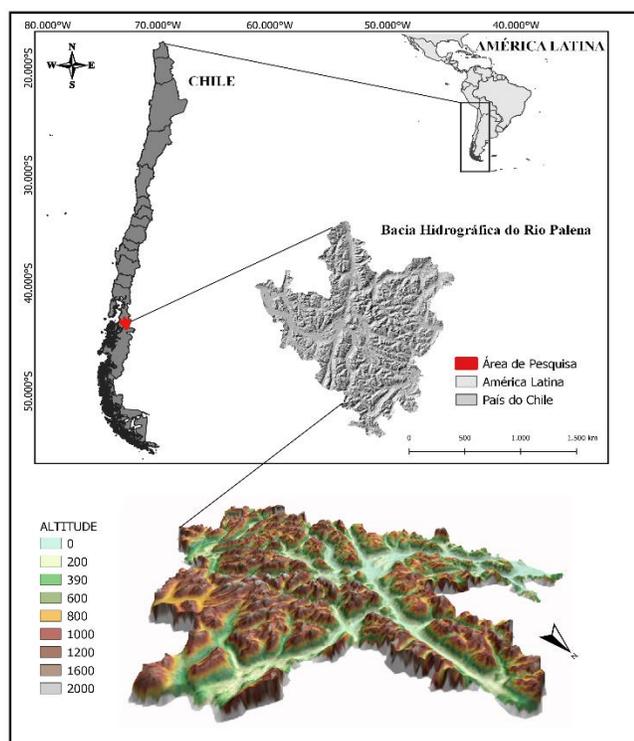


Figura 1. Mapa de localização da área de pesquisa (provisório)

2.2. Procedimentos metodológicos

A análise de suscetibilidade de deslizamento foi realizada em três etapas: 1) levantamento das variáveis e estruturação do banco de dados geoespacial; 2) seleção e implementação do modelo; e 3) avaliação do modelo e quantificação das áreas de suscetibilidade.

Primeiramente, gerou-se o inventário de deslizamento de terra na área de interesse de maneira manual a partir de interpretação visual de imagens de alta resolução [12]. Para isso, foram utilizadas imagens de satélite do programa *Google Earth Pro* referentes ao ano de 2021 e da coleção de imagens do Sentinel 1 e 2, entre janeiro de 2015 a dezembro de 2021. O inventário final resultou em 40 polígonos de deslizamento representados por 1.968 pontos e o mesmo quantitativo de pontos foram gerados aleatoriamente para descrever as áreas de não deslizamento. Para cada uma das

categorias de deslizamento e não deslizamento, se dividiu aleatoriamente as duas partes de 70% e 30%, para compor as amostras de treinamento e validação, respectivamente [3].

A seleção dos fatores condicionantes de deslizamentos foi realizada a partir de revisão bibliográfica para elencar os principais fatores utilizados na literatura internacional. Assim, foram considerados para este estudo os seguintes fatores: altitude, declividade, aspecto, curvatura do perfil, índice de umidade topográfica (TWI), índice de potência do fluxo (SPI), altura acima da drenagem mais próxima (HAND), litologia, Índice de Vegetação por Diferença Normalizada (MDVI), Índice de Água por Diferença Normalizada (NDWI) e Cobertura do Solo.

O algoritmo RF foi o modelo utilizado para gerar a suscetibilidade a deslizamento de terra. Este modelo é não paramétrico, sendo capaz de operar com dados contínuos e discretos e realizar operações de classificação e regressão. Por ser um modelo baseado em árvores de decisão, sua estruturação consiste em conjuntos de várias árvores de decisão individuais através, utilizando subconjuntos aleatórios de dados. O resultado final apresenta a média dos valores de previsão entre todas as árvores de decisão geradas. Nesta pesquisa, foi utilizado a regressão do RF para mensurar a suscetibilidade, variando de 0 (menor suscetibilidade) a 1 (maior suscetibilidade). A implementação do modelo foi feita através do pacote *RandomForest* na linguagem de programação R no software *RStudio* versão 4.0.2. O RF foi rodado com o *parameter ntree* definido como 500 (valor padrão do pacote) e o *mtry* como 3 (número aproximado da raiz quadrada do número de variáveis utilizadas).

A validação do modelo foi feita com base nos seguintes índices estatísticos: Taxa de Predição Positiva (TPP), Taxa de Predição Negativa (TPN), sensibilidade, especificidade e acurácia global. Além destes, a característica de operação do receptor ou Curva-ROC (*Receiver Operating Characteristic*) foi utilizada na validação do desempenho geral do modelo. Esta curva é amplamente utilizada como mecanismo de validação de modelos probabilísticos [4]. A curva ROC deriva da relação pareada entre diversos valores estadísticos, entre estes a sensibilidade e especificidade, que resultam no índice da Área Abaixo da Curva ROC ou AUC (*Area Under the Curve*), a qual permite validar quantitativamente os modelos. Quanto maior for o valor AUC, ou seja, quanto mais próximo de 1,0, melhor é o modelo. Os valores de AUC podem ser ranqueados em: ruim (0,5–0,6), médio (0,6–0,7), bom (0,7–0,8), muito bom (0,8–0,9) e excelente (0,9–1,0) [5].

3. RESULTADOS

3.1. Importância das variáveis

A importância de cada variável para a execução do mapa de suscetibilidade é apresentada na (Figura 2). Para o modelo gerado, o fator de maior representatividade foi a geologia, seguida por HAND, Altitude e os três fatores descritores das condições de cobertura da superfície, NDVI, NDWI, Uso e

Cobertura da Terra, respectivamente. As demais seis variáveis demonstraram contribuições abaixo de 50%, sendo a Curvatura de perfil a de menor importância.

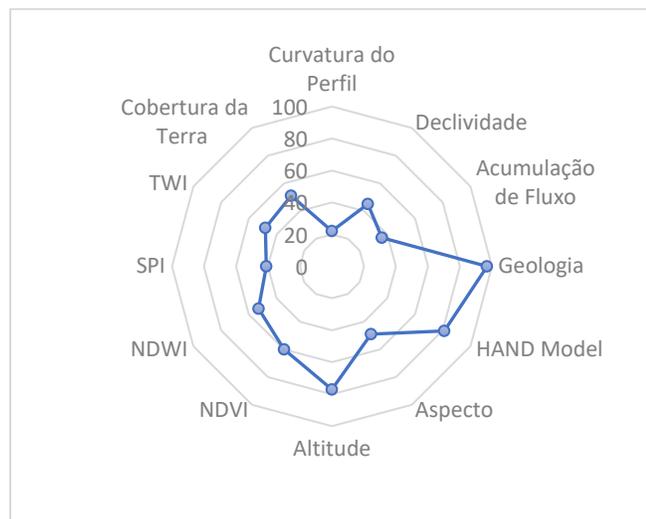


Figura 2. Classificação do nível de importância das variáveis explicativas do modelo

3.2. Mapeamento de suscetibilidade a deslizamento de terra utilizando o modelo RF

O mapeamento de suscetibilidade a movimentos de massa na bacia do Rio Palena, sul do Chile, é apresentado na (Figura 3), onde as classes de suscetibilidade foram escalonadas entre muito baixa e muito alta.

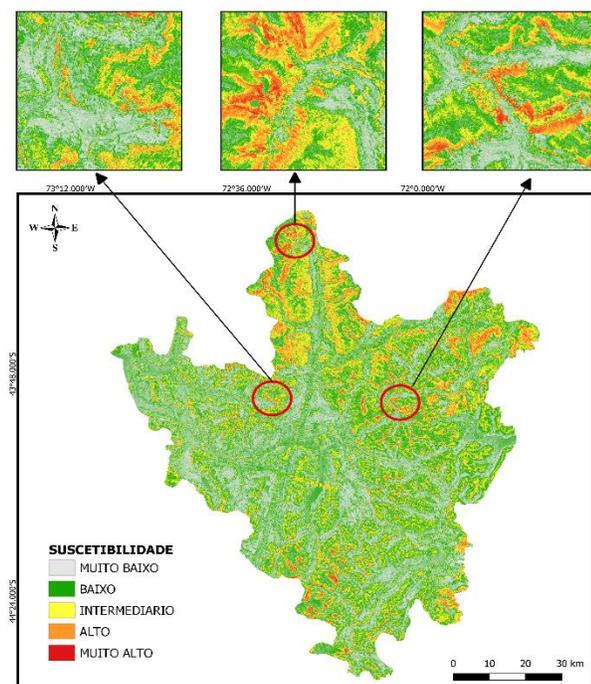


Figura 3. Mapa de suscetibilidade a deslizamentos de terra gerado através do modelo FR.

Observando o mapa de suscetibilidade (Figura 3), pode-se averiguar um padrão na concentração de áreas com maior suscetibilidade em regiões de alta declividade que acompanham o perfil de elevação das cadeias montanhosas, características da área de pesquisa. Por outro lado, as estruturas de planícies presentes na base e no topo das montanhas são apresentadas com suscetibilidade baixa e muito baixa, sendo estas últimas, as classes predominantes na bacia. Também se nota uma concentração das áreas de alta e muito alta suscetibilidade acompanhando os perfis de montanhas da parte norte, nordeste e em menor densidade na porção sul da bacia.

Quanto à porcentagem de área para cada categoria de suscetibilidade, a categoria de suscetibilidade muito baixa foi predominante, abrangendo 47% (360.917 km²) do total de área da bacia (772.337 km²), seguida pela categoria de baixa suscetibilidade que representa 33% (254.730 km²), a categoria intermediária aparece com 11% da bacia (87.747 km²), e, por fim, as feições de alta e muito alta suscetibilidade aparecem com as menores representatividade espacial, com 6% (45.490 km²) e 3% (14.453 km²), respectivamente.

3.3. Avaliação do desempenho e precisão do modelo

O resultado da AUC para o modelo RF adotado demonstrou robustez do modelo, alcançando um valor de 0.9937 (Figura 4), podendo ser considerado como excelente de acordo com o ranqueamento das classificações AUC^[5].

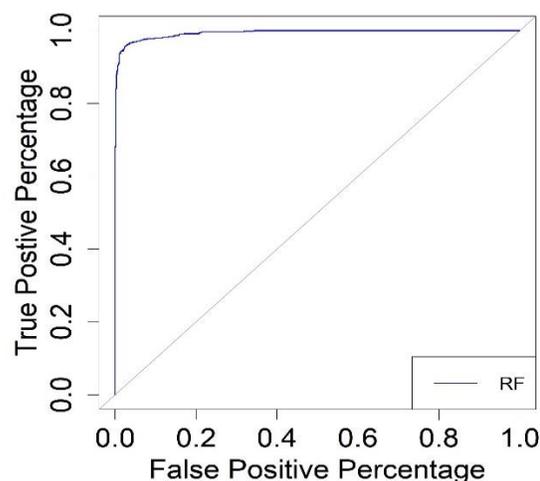


Figura 4. Análise da curva ROC do modelo de deslizamento usando conjunto de dados de teste

A análise de precisão da previsão do modelo baseou-se nas métricas estatísticas calculadas através da matriz de confusão. As métricas referentes à taxa de predição positiva e à taxa de predição negativa, responsáveis por mostrar se o modelo foi capaz de prever corretamente as áreas de ocorrência e não ocorrência de deslizamento de terra. Nesta pesquisa, a TPP obteve valor maior (0.9694) do que a TPN (0.9502)

Por sua vez, ao verificar o comportamento da estatística de sensibilidade (0.9500) e especificidade (0.9695), podemos notar que ambas obtiveram valores elevados e substancialmente aproximados. Contudo, a sensibilidade permaneceu com pontuação inferior à especificidade, indicando que o modelo tende a classificar de forma mais precisa as zonas de não deslizamento, apesar de o volume de amostra ter sido similar para ambas as classes de deslizamentos e não deslizamentos. Por fim, a acurácia global foi de 0.9597, um valor relativamente elevado, que reforça a boa estimativa gerada pelo modelo.

4. DISCUSSÃO

O deslizamento de terra são fenômenos potencialmente geradores de grandes desastres naturais em todo o mundo, provocando anualmente prejuízos financeiros substanciais. Decorrente disso, o mapeamento da suscetibilidade se torna uma importante ferramenta para a estruturação de medidas de redução de risco de desastres. Nesse contexto, várias pesquisas têm utilizado algoritmos de aprendizagem de máquina e inteligência artificial para o desenvolvimento e implementação de modelos de suscetibilidade a deslizamento para diversas localidades do mundo. Contudo, para a América Latina existem poucas pesquisas com esse perfil. Assim a presente pesquisa aplicou o algoritmo RF para gerar uma modelagem de suscetibilidade a movimento de massa em uma área de montanha abarcada pela bacia hidrográfica do rio *Palena*, situado na porção sul do Chile. Ao compara os resultados estatísticos de validação da acurácia do modelo e a espacialização das zonas de maior e menor suscetibilidade com outras pesquisas ^{[10]; [6]; [11]}. Foram observadas aproximações entre os resultados, reforçando a eficácia do modelo gerado.

5. CONCLUSÕES

A referida pesquisa mostrou a utilidade do algoritmo de aprendizagem de máquina RF na determinação de áreas suscetíveis a deslizamentos de terra utilizando fatores explicativos com resolução espacial ajustados a escala média. O mapeamento de suscetibilidade a deslizamento de terra gerado para a área da bacia hidrográfica do rio *Palena* permitiu identificar as áreas de maior suscetibilidade a deslizamento de terra. Assim, verificou-se que este mapeamento é um importante instrumento técnico-científico, que pode ser utilizado para orientar a gestão local e regional na elaboração de medidas de planejamento e gestão territoriais que visem minimizar a ocorrência de desastres causados por deslizamentos de terra. Por fim, ressalta-se que com a passagem temporal e a ocorrência de novos episódios de deslizamentos na área, deve-se atualizar o modelo, a fim de garantir a eficiência do mesmo.

6. REFERÊNCIAS

- [1] T. Davies. Landslide Hazards, Risks, and Disasters: Introduction. *Landslide Hazards, Risks, and Disasters*, Elsevier Inc., pp. 16, 2015.
- [2] Lacasse, S., Nadim, F., Kalsnes, B., Living with risk of landslide. *Geotechnical Engineering Journal of SEAGS & AGSSEA* 41 (4 December 2010 ISSN 0046-5828), 2010.
- [3] Rahmati O, Kornejady A, Deo RC. Spatial prediction of landslide susceptibility using random forest algorithm. Singapore: Springer, pp 281-292, 2021.
- [4] Pham BT, Tien Bui D, Prakash I, Dholakia MB. Classifier set based on fuzzy rotation forest rules for spatial prediction of landslides using GIS. *Nat. v.83: pp.97-127*, 2016.
- [5] Rasyid AR, Bhandary NP, Yatabe R. Frequency ratio performance and logistic regression model in creating the GIS-based landslide susceptibility map on Lompobattang Mountain, Indonesia. *Geoenvironmental Disasters. v.3: pp.19*. 2016.
- [6] Merghadi A, Yunus AP, Dou J, Whiteley J, Pham BT, Bui DT, Avtar R, Abderrahmane B. Machine learning methods for landslide susceptibility studies: a comparative view of the performance of algorithms. *Earth Sci Rev. 207* (September 2022): 103225, 2020.
- [7] EM-DAT, CRED / UCLouvain, Brussels, Belgium. <www.emdat.be>.
- [8] Ubydul Haque, Paula F. da Silva, Graziella Devoli, Jürgen Pilz, Bingxin Zhao, Asmaa Khaloua, Wahyu Wilopo, Peter Andersen, Ping Lu, Juneseok Lee, Taro Yamamoto, David Keellings, Jian-Hong Wu, Gregory E. Glass. The human cost of global warming: Deadly landslides and their triggers (1995–2014). *Science of The Total Environment, v. 682: pp. 673-684*, 2019.
- [9] Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, *Nat. Hazards Earth Syst. Sci., v.18: pp.2161–2181*, 2018.
- [10] M. A. Somos-Valenzuela, J. E. Oyarzún-Ulloa, I. J. Fustos-Toribio, N. Garrido-Urzuá, and N. Chen, “The mudflow disaster at Villa Santa Lucía in Chilean Patagonia: understandings and insights derived from numerical simulation and postevent field surveys,” *Nat. Hazards Earth Syst. Sci.*, v. 20: pp. 2319–2333, 2020.
- [11] Carrión-Mero, Paúl and Briones-Bitar, Josué and Morante-Carballo, Fernando and Stay-Coello, David and Blanco-Torrens, Roberto and Berrezueta, Edgar. Evaluation of Slope Stability in an Urban Area as a Basis for Territorial Planning: A Case Study. *Applied Sciences, v.11*, 2021.
- [12] Rabby, Y.W., Li, Y. An integrated approach to map landslides in Chittagong Hilly Areas, Bangladesh, using Google Earth and field mapping. *Landslides, v. 16: pp. 633–645*, 2019.