

# Suscetibilidade a Deslizamentos Utilizando Machine Learning: Um Estudo de Caso em Angra dos Reis - RJ

Amanda Alves da Silva<sup>1</sup>

Doutoranda em Engenharia Ambiental, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil | amandaalves@poli.ufrj.br

Marcos Barreto de Mendonça<sup>2</sup>

Professor Associado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil | mbm@poli.ufrj.br

André de Souza Avelar<sup>2</sup>

Professor Associado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil | andreavelar@acd.ufrj.br

**RESUMO:** A predição das probabilidades de ocorrência de deslizamentos subsidiam a elaboração de mapas de suscetibilidade, instrumento básico na gestão de riscos geológicos e no planejamento territorial sustentável. O presente estudo propõe a modelagem estatística da suscetibilidade a deslizamentos da região central de Angra dos Reis - RJ, por meio de algoritmos de machine learning, baseando-se em dados que caracterizam fatores condicionantes de deslizamentos e o inventário de ocorrências. Inicialmente, foi conduzida uma análise geoespacializada das cicatrizes e registros de deslizamentos para compilar um inventário, fundamentando-se na localização e características geométricas dos eventos. Doze fatores condicionantes foram selecionados, incluindo dados topográficos, de cobertura vegetal e geológicos, essenciais para entender os padrões de ocorrência. Através da análise de componentes principais (PCA), identificaram-se padrões complexos nas relações entre esses fatores, permitindo uma compreensão sobre quais deles mais influenciam na deflagração de ocorrências na área de estudo. Cinco diferentes algoritmos classificadores foram aplicados para prever a suscetibilidade, como Random Forest, SVM, Regressão Logística, LightGBM e XGBoost. O teste de validação dos resultados demonstrou um bom desempenho das modelagens utilizadas, destacando-se o modelo baseado em Random Forest, cuja avaliação da taxa de falsos negativos indicou uma melhor performance em relação aos demais algoritmos testados.

**PALAVRAS-CHAVE:** Deslizamentos, Suscetibilidade, Machine Learning.

**ABSTRACT:** Predicting the probability of landslides subsidizes the creation of susceptibility maps, a basic tool in geological risk management and sustainable territorial planning. This study proposes the statistical modeling of the susceptibility to landslides in the central region of Angra dos Reis - RJ, using machine learning algorithms, based on data that characterizes the conditioning factors of landslides and the inventory of occurrences. Initially, a geospatial analysis of landslide scars and records was conducted to compile an inventory, based on the location and geometric characteristics of the events. Twelve conditioning factors were selected, including topographic, vegetation cover and geological data, which are essential for understanding the patterns of occurrence. Through principal component analysis (PCA), complex patterns were identified in the relationships between these factors, allowing an understanding of which of them most influence the outbreak of occurrences in the study area. Five different classification algorithms were applied to predict susceptibility, including Random Forest, SVM, Logistic Regression, LightGBM and XGBoost. The test to validate the results showed that the models used performed well, with the Random Forest model standing out as having the best false negative rate compared to the other algorithms tested.

**KEYWORDS:** Landslides, Susceptibility, Machine Learning.

## 1. INTRODUÇÃO

Angra dos Reis, município situado no litoral sul do Rio de Janeiro (Figura 1), é palco de frequentes deslizamentos de terra, notadamente associados a eventos pluviométricos intensos. Grande parte de tais

eventos têm levado a situações de desastre, cada vez mais frequentes nas últimas décadas, devido à maior exposição de populações vulneráveis a tais ameaças. Em virtude disso, foi iniciada no ano de 2011 a elaboração de instrumentos técnico-científicos voltados para o estudo e análise de deslizamentos ocorridos na cidade (CPRM, 2011) e a previsão de futuros desastres associados a deslizamentos, incluindo cartas geotécnicas. As Cartas Geotécnicas produzidas até o momento incluem, as Cartas de Risco e de Suscetibilidade a Movimentos de Massa em escala de 1:5.000, desenvolvida pela UFRJ (UFRJ, 2014) e a Carta Geotécnica de Aptidão Urbana - CGU, elaborada pelo Serviço Geológico do Estado do Rio de Janeiro - DRM/RJ (NADE/DRM, 2015).

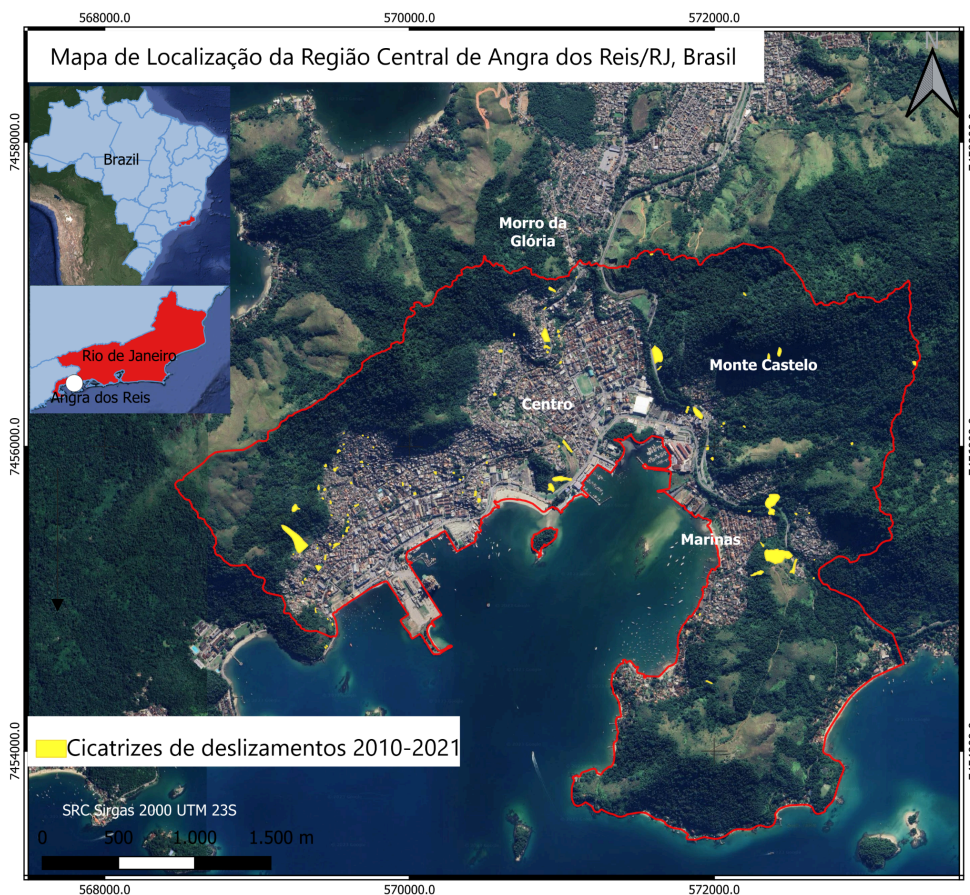


Figura 1: Área de estudo e localização das cicatrizes de deslizamentos de 12/ 2010 a 12/2021.

O mapeamento de suscetibilidade é um dos principais instrumentos cartográficos na gestão de riscos associados à movimentos de massa, uma vez que, a partir do seu conhecimento pode-se adotar medidas estruturais e não-estruturais para a redução dos riscos de forma racional. Atualmente, existem várias abordagens para a elaboração destes mapas, incluindo métodos heurísticos (baseados no julgamento subjetivo de profissionais especialistas da área considerando ponderando as características do terreno), determinísticos (modelagem matemática dos processos físico - geotécnico e hidrológico - envolvidos na deflagração de deslizamentos) e estatísticos (baseados em inventários e sua correlação com as características físicas do terreno) - Zêzere et al. (2017). Nos métodos estatísticos, os pesos dos condicionantes relacionados com os deslizamentos de terra são encontrados por meio de inventários de deslizamento, cuja abordagem é de que eventos passados de deslizamentos podem informar ocorrências futuras (Fell et al., 2008).

Nesse sentido, o presente trabalho propõe um modelo estatístico por meio de algoritmos de machine learning para o Mapeamento da Suscetibilidade a Deslizamentos (MSD), bem como a investigação dos Fatores Condicionantes a Deslizamentos (FCD) que mais influenciam a deflagração de deslizamentos para a região central de Angra dos Reis. O resultado final deste modelo dará suporte a um novo MSD para a área de estudo, ressaltando que a última elaboração de MSD ocorreu há aproximadamente dez anos. Mapas de

suscetibilidade e de risco necessitam de atualização com uma certa frequência, devido a possível melhoria de dados dos FCDs e/ou sua variação ao longo do tempo. Sabe-se, no entanto, que nem todas as municipalidades têm disponibilidade de recursos técnicos e financeiros para a execução desses estudos. Tem sido publicados alguns trabalhos sobre mapeamento de suscetibilidade a deslizamentos no Brasil por meio de modelagem probabilística baseada em machine learning (ex: Canavessi et. al, 2020; Hader et al., 2022), porém, ainda não contemplando a área do presente estudo.

Nas últimas duas décadas, a disponibilidade de imagens de satélite gratuitas proporcionou uma melhoria substancial no mapeamento dos FCDs, permitindo ainda a integração destes dados com técnicas de Machine Learning (ML) e Deep Learning (DL) para o mapeamento de suscetibilidade, empregando algoritmos como as Random Forest (RF) e Support Vector Machines (SVM). Para avaliar a disseminação das técnicas ML e DL na comunidade científica para o mapeamento da suscetibilidade a deslizamentos, foi realizado um levantamento de artigos na base do *Scopus*. O mecanismo de busca foi baseado na estruturação de lógica booleana (and e or) entre as palavras-chaves “Landslides”, “Susceptibility” e “Machine Learning”, tendo sido encontrado um total de 1.605 referências após o refinamento de busca. Os estudos realizados por pesquisadores de diferentes partes do mundo indicam que os métodos baseados em ML e DL oferecem um grande potencial para o mapeamento de suscetibilidade a deslizamentos e que a escolha dos algoritmos que são implementados nos diversos modelos de suscetibilidade encontrados na literatura depende de diversos fatores, como a natureza dos dados disponíveis, a escala do estudo e os objetivos específicos da pesquisa.

## 2. METODOLOGIA

Neste capítulo serão abordados os procedimentos para tratamento de dados sobre o inventário de deslizamentos e FCDs, bem como os métodos estatísticos e suas respectivas definições empregados no MSD.

### 2.1. Dados e etapas do trabalho

O MSD da região central de Angra dos Reis iniciou-se com a seleção de registros de deslizamentos para a elaboração do Inventário de Deslizamentos (ID), baseado na localização, data e geometrias de deslizamentos anteriores em imagens do Google Earth (2010-2021). Ao mesmo tempo, foram mapeados 12 FCDs (Tabela 1), todos na resolução de 30x30 metros. A segunda etapa consistiu no empilhamento dos mapas supracitados e amostragem em grade sobre a área de estudo com resolução de 30x30 metros, correspondente com a escala de cada um dos FCDs mapeados. A terceira etapa foi a criação de um modelo de aprendizagem de máquina, selecionando do total de pontos amostrados em grade, aqueles que ficaram dentro os polígonos de cicatrizes de deslizamentos para representar as ocorrências; e apenas aqueles que atenderam ao critério de possuir declividade  $> 10^\circ$  e de se tratar de áreas planas onde não há potencialidade de deslizamentos para representar as não ocorrências.

Os dados obtidos foram, então, padronizados e encaminhados para implementação de algoritmos de estatística multivariada, como a Análise de Componentes Principais (PCA), com o intuito de identificar as relações entre FCDs segundo sua variabilidade. Em seguida, o cálculo probabilístico da suscetibilidade foi conduzido por cinco algoritmos classificadores, com o intuito de selecionar aquele que melhor se ajusta ao modelo. O MSD foi, por fim, validado através de métricas estatísticas como a matriz de confusão e a área sob a curva ROC (AUC), além da verificação da taxa de falsos negativos. Quanto mais próximo de 1 for o valor de AUC, melhor será a performance do modelo.

A elevação do terreno, aferida por meio de um modelo digital de elevação (DEM), fornece uma representação quantitativa da superfície terrestre e de seus produtos cartográficos derivados, como a declividade, Curvatura, Curvatura Plana, Curvatura Padrão e TWI. A ocorrência de deslizamentos é diretamente afetada pela declividade do terreno. As curvaturas representam a morfologia do terreno, com o sentido de identificar áreas côncavas, convexas ou planas. O aspecto indica a orientação azimutal dos taludes na área de estudo, além de ser relevante para a determinação do TWI. O TWI é responsável por descrever os atributos topográficos dos processos hidrológicos. Os terrenos sobre os quais recaem valores de TWI menores são reconhecidos como áreas propensas a deslizamentos (Achour e Pourghasemi, 2020).

A fim de identificar os níveis de cobertura vegetal e de ocupação urbana na área de estudo, optou-se por utilizar índices espectrais obtidos em imagens de satélite LANDSAT 8, como o índice da diferença

normalizada da vegetação (NDVI) e o Índice da Diferença Normalizada para Construções (NDBI), respectivamente. O NDVI além de ser um bom mecanismo para identificar áreas com cobertura vegetal incipiente ou ausente, é um bom parâmetro para observar potenciais cicatrizes de deslizamentos pretéritos. Já o NDBI, reflete as influências das atividades de ocupação humana na evolução regional de deslizamentos de terra (Huang et al., 2022).

Os dados de solos foram adquiridos através da Carta Geotécnica de Aptidão Urbana, que se divide em duas cartas distintas: a Carta Geológico-Geotécnica Específica sobre Escorregamentos de Angra dos Reis (CGUi) e a Carta Geotécnica de Aptidão Urbana Específica quanto ao Potencial de Ocorrência de Escorregamentos (CGUf) (NADE/DRM, 2015). Estas cartas fornecem um detalhamento do meio físico da de Angra dos Reis, com ênfase na ocorrência de escorregamentos nas encostas e nas diferentes unidades geológico-geotécnicas essenciais para entender os riscos de movimentos de massa e ordenamento territorial, classificadas como: Afloramentos Rochosos (AF), Solos Rasos Sobre Rocha (S/R), Solos Residuais Espessos (SR), Zonas de Concentração de Blocos in situ (BR), Depósitos de Corrida de Massa (CMD) e Depósitos de Tálus (TA). Essa classificação detalhada é fundamental para uma avaliação precisa do potencial de ocorrência de escorregamentos em diferentes áreas da região de Angra dos Reis.

As informações acerca dos aspectos geológicos e geomorfológicos foram obtidas em UFRJ (2014), que descreveu a predominância de predominância dos Granitos Mombaça e ortognaisses associados ao Complexo Rio Negro e compartimentou geomorfológicamente a área de estudo em cristas isoladas e serras baixas, depósitos tecnogênicos, escarpas de borda de planaltos, ilhas costeiras, morros altos, planícies fluvio-marinhas, planícies marinhas e rampas de colúvio/depósito de tálus.

Tabela 1: FCDs considerados no trabalho e suas respectivas fontes e referências de uso

Fator Condicionante	Fonte	Referência Acadêmica
Elevação	ASTER GDEM	Canavessi et. al., 2020; Hader, 2022; Merghadi, 2020; Achour e Pourghasemi, 2020; Chen e Chen, 2021; Gupta e Shukla, 2022; Huang et al., 2022.
Declividade	ASTER GDEM	
Curvatura, Curvatura Plana e Curvatura Padrão	ASTER GDEM	
Índice de Molhamento Topográfico (TWI)	ASTER GDEM	
Aspecto	ASTER GDEM	
Cobertura Vegetal	LandSat 08 (NDVI)	
Ocupação Urbana	LandSat 08 (NDBI)	
Litologia	Mapa Geológico	UFRJ, 2014
Geomorfologia	Mapa Geológico	
Solos	Carta Geotécnica de Aptidão Urbana	NADE/DRM, 2015

## 2.2. Análise de Influências

O PCA é uma técnica estatística multivariada usada para identificar e compreender as influências dos fatores condicionantes na deflagração de deslizamentos (Solanki e Gupta, 2022). Ao examinar os dados de entrada, este algoritmo é capaz de identificar padrões e relações entre esses fatores, destacando quais exercem maior influência na ocorrência de deslizamentos. Além disso, o PCA desempenha um papel importante para lidar com a complexidade dos dados de entrada, permitindo não apenas uma análise mais simplificada, como também a compreensão das interações complexas entre os FCDs e suas contribuições para explicar a variabilidade observada.

### 2.3. Análise Probabilística

Para realizar a predição probabilística de deslizamentos, foi concluído que os pesquisadores empregaram uma série de métodos estatísticos e algoritmos de machine learning para quantificar a suscetibilidade a deslizamentos, como a Regressão Logística (LR), Support Vector Machine (SVM), e outros algoritmos baseados em árvores de decisão, como a Random Forest (RF), o LightGBM (LGB) e o XGBoost (XGB). A LR é uma ferramenta estatística capaz de modelar a probabilidade de deslizamentos, podendo ainda ser empregada de maneira complementar na etapa de análise de influências, com o intuito de observar o comportamento global das interações entre FCDs, como é o caso deste estudo. O RF, é um método que utiliza múltiplas árvores de decisão para criar um modelo robusto capaz de capturar padrões complexos nos dados, sendo particularmente eficaz para lidar com conjuntos de dados grandes e heterogêneos, com alta dimensionalidade e interações não lineares entre variáveis (Breiman, 2001). Da mesma maneira, o LightGBM (LGB) e o XGBoost (XGB) são algoritmos baseados em *gradient boosting* que se destacam pela sua capacidade de lidar com dados desbalanceados e de gerar modelos com alta precisão e capacidade preditiva (Ke, 2017; Chen, 2016). Outra abordagem amplamente utilizada é a Support Vector Machine (SVM), que mapeia os dados em um espaço multidimensional e distingue classes através de um hiperplano de separação (Vapnik, 1995). O SVM é eficaz na identificação de padrões complexos e na modelagem de relações não lineares entre variáveis, sendo uma ferramenta para a previsão de suscetibilidade a deslizamentos em diferentes contextos geográficos e ambientais.

### 3. RESULTADOS

Após a aplicação do PCA combinado com a LR, foi verificado que, com a redução da dimensionalidade dos dados de entrada com 12 FCDs (que são também as variáveis preditoras do modelo), apenas oito componentes principais são capazes de explicar 95% da variabilidade do modelo, além de revelar padrões distintos e inter-relacionados entre os FCDs (Figura 2), tanto para as ocorrências de deslizamento (representadas pelos pesos positivos) quanto para a não-ocorrência (representadas pelos pesos negativos).

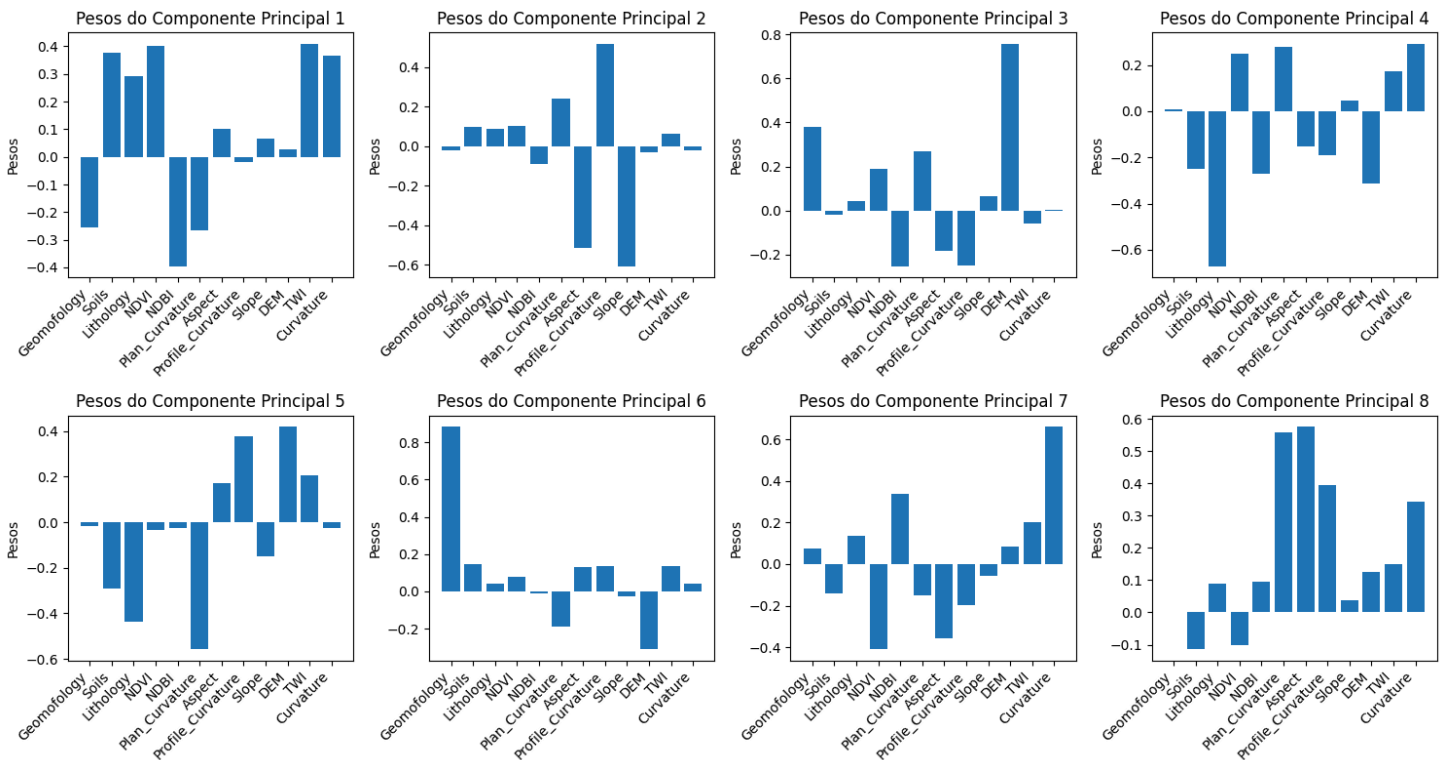


Figura 2: Quantificação dos pesos de influência para cada um dos FCDs, em cada um dos 8 PCAs

O primeiro componente principal revelou uma forte correlação entre os FCDs do solo, litologia, NDVI, TWI e curvatura, sugerindo interações entre a composição do solo, morfologia do terreno e o nível de cobertura da vegetação. No segundo componente, foi observado uma relação entre características inerentes à curvatura do terreno e à orientação das encostas, apontando para padrões geomorfológicos específicos. O terceiro componente indicou uma associação entre características da superfície da terra, vegetação e elevação do terreno. O quarto componente principal destacou a relação entre solos e litologia com o NDBI, sugerindo padrões de urbanização relacionados à composição do solo. No quinto componente, a associação entre solos, litologia com a curvatura do terreno, enfatizando a associação entre os parâmetros geológicos e fisiográficos da área de estudo. O sexto componente indicou uma possível relação entre a geomorfologia, curvatura do terreno e NDBI, apontando reflexões sobre o ordenamento da ocupação urbana ao longo das encostas. O sétimo componente ressaltou a associação entre solos, litologia e orientação das encostas. Por fim, o oitavo componente principal evidenciou a relação entre características do solo e da litologia com a curvatura do perfil do terreno, contribuindo para uma compreensão mais abrangente das variações na morfologia do terreno.

Na etapa de análise probabilística (Figura 3a), os resultados indicam que os modelos baseados em RF, LightGBM e XGB tiveram desempenhos excepcionais e com alta precisão na identificação de áreas suscetíveis. O SVM também teve um bom desempenho, enquanto a RL teve a menor acurácia, sugerindo menor capacidade de discriminação entre as classes. Todos os modelos apresentaram AUCs elevadas, com XGBoost alcançando a maior pontuação, seguido por Random Forest e LightGBM, todos próximos a 1. O SVM e a Regressão Logística tiveram AUCs um pouco mais baixas, mas ainda significativas, acima de 0,85.

No sentido de escolher o melhor modelo preditivo entre os cinco algoritmos utilizados, foi realizada uma análise da taxa de falsos negativos para cada um dos modelos, com o intuito de verificar a capacidade de identificação de áreas de não ocorrência, quando na verdade seriam ocorrências positivas. A Figura 3b mostra que o modelo baseado em RF apresenta a menor taxa de falsos negativos, enquanto a regressão logística apresenta a maior taxa. Considerando-se os diferentes algoritmos de machine learning empregados na pesquisa, observa-se que o modelo para elaboração do MSD baseado em RF foi o que proporcionou os melhores resultados, podendo ser considerado de alta acurácia para o caso estudado. A Figura 4 apresenta o MSD baseado em RF.

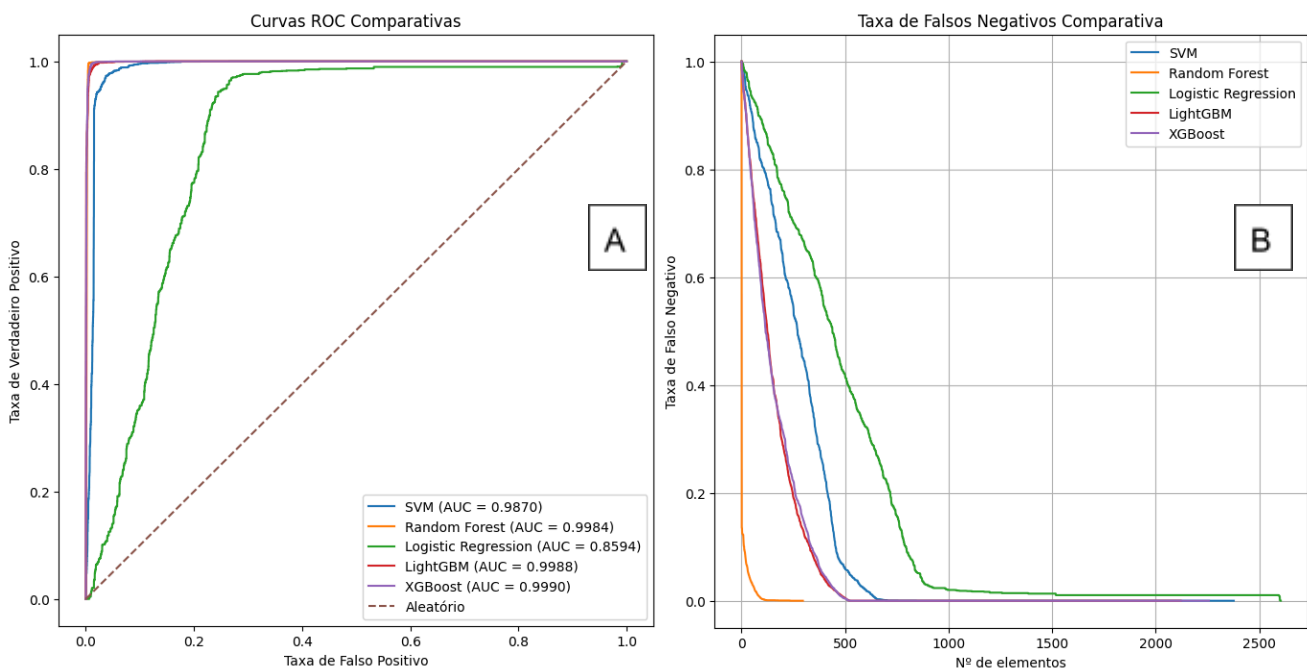


Figura 3: Comparativo de curvas ROC-AUC e de taxas de falsos negativos

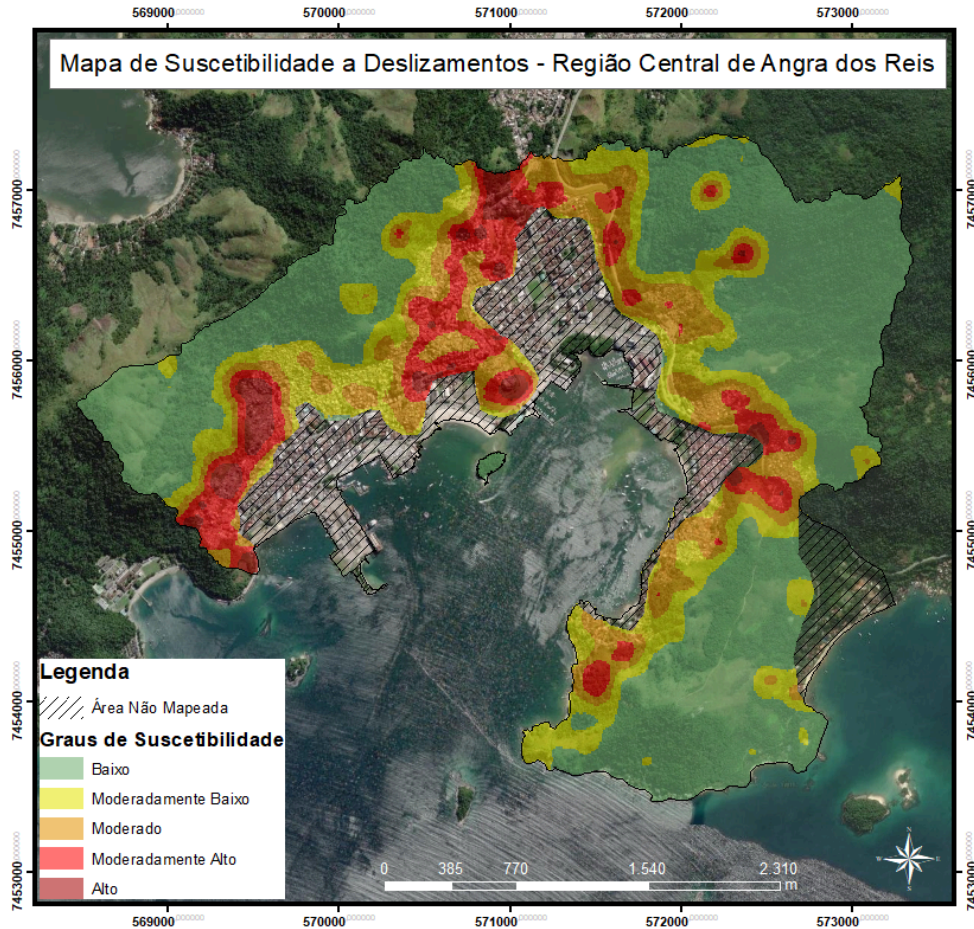


Figura 4: Mapa de Suscetibilidade a Deslizamentos da Região Central de Angra dos Reis baseado em Random Forest

#### 4. CONCLUSÃO

Por meio da análise de inventário e do mapeamento de fatores condicionantes, foi possível identificar padrões nas interações ambientais e geomorfométricas relacionadas aos eventos de deslizamento. A aplicação de técnicas de estatística multivariada, como PCA, permitiu uma compreensão mais profunda dos principais fatores que influenciam de maneira complexa a deflagração na ocorrência desses eventos, destacando-se de maneira global os FCDs Solos, Litologia, NDVI, NDBI, Curvatura Plana, Aspecto e TWI.

Dentre os resultados obtidos, a performance do modelo baseado em RF na previsão da suscetibilidade a deslizamentos se destacou em relação aos demais algoritmos classificadores não apenas em virtude de sua acurácia e valor de AUC, mas sobretudo devido ao fato de apresentar menor taxa de falsos negativos preditos. O trabalho realizado comprova o grande potencial de uso do método estatístico baseado em machine learning para a obtenção de um dos principais instrumentos para a gestão de riscos geológicos, o mapa de suscetibilidade a deslizamentos. O embasamento do mapeamento da suscetibilidade através de modelagem estatística reduz a subjetividade no processamento de dados, além de motivar novos estudos para o restante do município e outras regiões..

#### AGRADECIMENTOS

Agradecemos à CAPES/Brasil - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Brasil (código de financiamento 001) e CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio à pesquisa.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Achour, Y., & Pourghasemi, H. R. (2020). How do machine learning techniques help in increasing accuracy of landslide susceptibility maps?. *Geoscience Frontiers*, 11(3), 871-883.
- Breimann, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.
- Canavesi, V., Segoni, S., Rosi, A., Ting, X., Nery, T., Catani, F., & Casagli, N. (2020). Different approaches to use morphometric attributes in landslide susceptibility mapping based on meso-scale spatial units: A case study in Rio de Janeiro (Brazil). *Remote Sensing*, 12(11), 1826.
- Chen, T., & Chen, X. (2021). GIS-based landslide susceptibility assessment using an optimized ensemble learning approach of bagging trees, random subspace, and artificial neural network models. *Catena*, 205, 105460.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering geology*, 102(3-4), 85-98.
- Gupta, A., & Shukla, P. (2022). Multi-criteria approach for landslide susceptibility assessment in the eastern Himalayas: a case study of Tista River Basin, Sikkim, India. *Arabian Journal of Geosciences*, 15(1), 30.
- Hader, P. R. P., Kaiser, I. M., Manzato, G. G., & Peixoto, A. S. P. (2019). Hazard Assessment of Landslides Disasters in the City of Cubatão, State of São Paulo, Brazil. In *International Congress on Engineering and Sustainability in the XXI Century* (pp. 1087-1101). Cham: Springer International Publishing.
- Huang, F., Chen, J., Liu, W., Huang, J., Hong, H., & Chen, W. (2022). Regional rainfall-induced landslide hazard warning based on landslide susceptibility mapping and a critical rainfall threshold. *Geomorphology*, 408, 108236.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., ... & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225.
- NADE/DRM (2015). Carta Geotécnica de Aptidão Urbana - CGU, elaborada pelo Serviço Geológico do Estado do Rio de Janeiro - DRM/RJ.
- Solanki, A., Gupta, V., & Joshi, M. (2022). Application of machine learning algorithms in landslide susceptibility mapping, Kali valley, Kumaun Himalaya, India. *Geocarto International*, 37(27), 16846-16871.
- UFRJ (2014). Carta de Risco e de Suscetibilidade a Movimentos de Massa 1:5.000, desenvolvidas pelas equipes da COPPE UFRJ e Geoheco.
- Vapnik, V. N. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- Zêzere, J. L., Pereira, S., Melo, R., Oliveira, S. C., & Garcia, R. A. (2017). Mapping landslide susceptibility using data-driven methods. *Science of the total environment*, 589, 250-267.