

GEOTECNOLOGIAS E ANÁLISES ESTATÍSTICAS MULTIVARIADAS APLICADAS EM OBRAS LINEARES

Mauro Candido Filho

Mestrando (PPGEM-DEMIN), Univ. Federal de Ouro Preto, Brasil, mauro.candido@aluno.ufop.edu.br

Tatiana Barreto dos Santos

Professora, Universidade Federal de Ouro Preto, Ouro Preto, Brasil, tatiana.santos@ufop.edu.br

Allan Erlikhman Medeiros Santos

Professor, Universidade Federal de Ouro Preto, Ouro Preto, Brasil, allan.santos@ufop.edu.br

RESUMO: Este estudo se propõe a analisar se um conjunto de dados obtidos em um ambiente SIG pode ser tratado utilizando técnicas de estatística multivariada, tais como análise de componentes principais e análise de agrupamentos. As observações no banco de dados contêm informações relacionadas a fatores condicionantes para deslizamentos. Os fatores escolhidos foram elevação, declividade, setorização litológica e setorização pedológica. A análise de componentes principais permitiu a redução da dimensionalidade do banco de dados para 2 variáveis, que mantiveram 83,25% da variância dos dados originais. A análise de agrupamentos revelou a presença de 3 grupos distintos. Comparando a disposição espacial de eventos registrados na ferrovia, os grupos 1 e 2 contiveram cerca de 73% das ocorrências. A validação dos agrupamentos pelo coeficiente de silhueta demonstrou baixa coesão e alta separação para todos os agrupamentos, com coeficiente médio de 0,58, demonstrando agrupamento natural dos dados analisados.

PALAVRAS-CHAVE: Estatística Multivariada, SIG, ferrovias, rodovias.

ABSTRACT: This study aims to analyze whether a dataset obtained in a GIS environment can be treated using multivariate statistical techniques, such as principal component analysis and cluster analysis. The observations in the database contain information related to conditioning factors for landslides. The chosen factors were elevation, slope, lithological zoning, and pedological zoning. Principal component analysis allowed the reduction of the database dimensionality to 2 variables, which retained 83.25% of the original data variance. Cluster analysis revealed the presence of 3 distinct groups. Comparing the spatial distribution of events recorded on the railway, groups 1 and 2 contained approximately 73% of the occurrences. Validation of the clusters by the silhouette coefficient showed low cohesion and high separation for all clusters, with an average coefficient of 0.58, demonstrating a natural clustering of the analyzed data.

KEYWORDS: Multivariate Statistics, GIS, railways, highways.

1 INTRODUÇÃO

O uso e aplicação de ferramentas de geoprocessamento tem crescido substancialmente em diversos campos da ciência nos últimos anos, tendo como exemplos a modelagem de recarga em aquíferos (Souei & Zouaghi, 2021), investigação de contaminações do solo por metais pesados (Hou *et al.*, 2017), ameaças de incêndios florestais (Cetin *et al.*, 2023) entre outros. A aplicação de SIG está fortemente presente nos ramos de ciências da terra e as engenharias (Feng *et al.* 2020, Wei *et al.* 2021, Taloor *et al.* 2021), incluindo a geotecnia (Khan *et al.*, 2022), pois têm como uma das premissas fundamentais a demarcação no espaço das características observados em campo.

A ocorrência de eventos geológicos-geotécnicos como deslizamentos de encostas, tombamento de blocos e ruptura de taludes podem ocorrer pelas intervenções humanas sobre o terreno, seja pela ocupação indiscriminada de locais sujeitos a estas ocorrências, seja por ações planejadas, tais como a implantação de rodovias e ferrovias (Yamusa, Ismail & Tella, 2022; Khusulio & Kumar, 2023). Entretanto, não é simples a escolha de quais variáveis serão avaliadas nestes projetos e estabelecer relações significativas entre elas (Khusulio & Kumar, 2023).

Uma ferramenta essencial na avaliação de áreas sujeitas a episódios relacionados a movimentos de massa é o mapa de suscetibilidade a deslizamentos (Ma *et al.*, 2019). Este é um instrumento que não reduz o fenômeno relacionado ao perigo, mas que auxilia na redução da extensão dos impactos causados por deslizamentos (Ali *et al.*, 2021). Várias técnicas foram desenvolvidas para auxiliar na elaboração de mapas de suscetibilidade nas últimas décadas (Leonardi *et al.*, 2021). Tais métodos podem ser divididos em 2 grupos: métodos qualitativos (baseados em inventários e orientados por conhecimento prévio) e métodos quantitativos (baseados em modelos físicos e orientados por dados) (Leonardi *et al.*, 2021; Ma *et al.*, 2019).

Métodos qualitativos, ou heurísticos, são mais descritivos e subjetivos, baseando-se no conhecimento de especialistas em mapeamentos de suscetibilidade para deslizamento e zonas de perigo (Senouci *et al.*, 2021). Estes métodos fornecem uma estimativa de “onde” os deslizamentos são esperados (Leonardi *et al.*, 2021). Já os métodos quantitativos podem ser divididos em 2 subgrupos, os determinísticos, nos quais se calcula um fator de segurança relacionado ao talude em estudo, e os estatísticos, em que o objetivo é analisar as correlações entre fatores condicionantes e a distribuição dos deslizamentos no espaço (Senouci *et al.* 2021, Leonardi *et al.* 2021). A escolha de qual método utilizar, irá depender da escala do estudo, da disponibilidade de dados e do próprio conhecimento científico envolvido (Senouci *et al.*, 2021).

Antes da escolha do método a ser empregado para a execução do mapa de suscetibilidade, é primordial elaborar um banco de dados dos fatores desencadeantes nos processos de deslizamentos (Bhat *et al.*, 2023). Alguns exemplos de fatores relevantes utilizados em pesquisas ao redor do mundo consideram elevação (altitude), declividade, orientação de vertentes, solo, geologia, uso e ocupação dos solos, distância de rios, distância de rodovias, NDVI, entre outros (Das, Sarnar & Kanungo 2022; Bhat *et al.* 2023; Li, Wang & Chen 2021, Leonardi *et al.* 2021, Senouci *et al.* 2021).

Os usos extensivos dos Sistemas de Informação Geográfica- SIGs (ou *Geographic Information System - GIS*) e o Sensoriamento Remoto permitiram que eles se estabelecessem como parte integral na avaliação de suscetibilidade a deslizamentos graças às capacidades de gerenciamento dos fatores envolvidos, bem como permitirem estabelecer correlações e realizar estimativas entre os fatores e as áreas propensas aos deslizamentos (Leonardi *et al.* 2021; Das, Das, Sarnar & Kanungo 2022).

O objetivo do presente trabalho pode ser sintetizado como a verificação de tendência de agrupamentos para um conjunto de observações extraídas num trecho da ferrovia Estrada de Ferro Vitória-Minas por meio do qual será realizada análises de estatística multivariada (análise de componentes principais e análise de agrupamentos) de fatores condicionantes escolhidos (elevação, declividade, litologia e pedologia), extraídos usando um software SIG. Os agrupamentos serão validados e comparados com os registros de eventos ocorridos no passado para o mesmo trecho.

2 MATERIAIS E MÉTODOS

2.1 Área de estudo

O trecho em análise é denominado “Ramal de Fábrica”, uma parte da ferrovia Estrada de Ferro Vitória-Minas, cuja concessão atual é da mineradora Vale do Rio Doce.

Este trecho possui cerca de 165 km de extensão e está completamente inserido no estado de Minas Gerais. Tem seu ponto inicial próximo a cidade de Nova Era e do entroncamento com o “Ramal de Itabira”. Seu ponto final é nas imediações da mina de Fábrica Nova, também pertencente à mineradora Vale do Rio Doce e localizado próximo à cidade de Congonhas.

O propósito da operação deste ramal é o escoamento das produções de minério de ferro da concessionária e outras cargas gerais, como combustível e celulose. Além disto, numa porção deste segmento também trafega o comboio de vagões do Trem Turístico de Passageiros, que parte de Belo Horizonte (MG) com destino a Vitória (ES) e vice-versa.

Foi estipulado um *buffer* de 120 m em cada lado do traçado da linha, a fim de delimitar a área de influência das camadas temáticas condicionantes, as quais constituíram o mapa de suscetibilidade a deslizamentos.

2.2 Fatores condicionantes de deslizamentos

2.2.1 Elevação

A elevação afeta características topográficas, geomorfológicas e geológicas, as quais conduzem a uma variação espacial dos diversos processos de deslizamentos e também influencia na distribuição vegetal do local (Nachappa *et al.*, 2020). Grandes elevações também são propensas a experimentar maiores índices pluviométricos, os quais podem provocar um aumento na saturação e reduzir a estabilidade do solo (Bhat *et al.*, 2023).

Os dados de elevação foram extraídos do Modelo Digital de Elevação NASADEM, disponibilizado pela *National Aeronautics and Space Administration, NASA*. A resolução do arquivo é 30 m x 30 m.

2.2.2 Declividade

Quanto maior a declividade, maior será a tensão de cisalhamento no pé do talude e, conseqüentemente, mais propenso será o talude à deformação e ruptura (Li, Wang & Chen, 2021). Portanto, a declividade pode ser considerada uma variável condicional crucial na análise de suscetibilidade a deslizamentos, visto que é empregada na maioria das análises.

A declividade é um produto derivado da elevação e, portanto, os dados foram obtidos do MDE NASADEM. Na elaboração do mapa temático, os intervalos foram definidos como 0-10°, 10-20°, 20-30°, 30-40°, 40-50° e > 50°.

2.2.3 Litologia

Características como a resistência e permeabilidade das rochas são capazes de serem previstas pela litologia. Do mesmo modo, propriedades *in-situ* do solo que influenciam processos de deslizamentos também podem ser indicadas pela litologia (Das, Das, Sarnar & Kanungo 2022).

Na área de estudo, a litologia foi dividida em 3 setores pela equipe de geologia da concessionária, tendo como base o Mapa Geológico do Estado de Minas Gerais (Silva *et al.*, 2020).

A litologia ficou dividida como: Xistos – Quartzitos Ferruginosos, Filitos – Xistos Quartzitos e Itabiritos, Granitos – Xistos Carbonáticos e Quartzitos.

2.2.4 – Pedologia

As características dos solos podem influenciar a ocorrência de deslizamentos, especialmente se uma classe pedológica possuir um alto grau de retenção para água (Tešić, 2021).

As classes de solos neste estudo foram extraídas do Mapa de Solos do Brasil (Santos *et al.*, 2011) e ajustadas pela equipe de geologia da concessionária. As classes foram Latossolo Vermelho e Cambissolo RFA.

2.3 – Componentes Principais

A técnica de componentes principais foi proposta por Karl Pearson, em 1901, e fundamentada por Hotteling em 1933. De acordo com Santos (2019), o objetivo desta técnica é explicar como se estruturam a variância e a covariância do conjunto de dados em um vetor aleatório contendo p variáveis, por meio da construção de combinações lineares das variáveis originais, de maneira que estas combinações lineares não apresentem correlações entre si. Por não exigir normalidade multivariada, esta técnica é aplicável a qualquer conjunto de variáveis, obrigando-se apenas que estas sejam correlacionadas entre si. Outro uso dessa técnica é a quantificação de variáveis. Essa técnica foi empregada nos dados com essa finalidade.

2.4 – Análise de Agrupamentos

Esta técnica buscar analisar eventos ou variáveis baseados nas suas próprias características. Os cálculos estatísticos agrupam elementos de forma indistinta, considerando apenas similaridades ou dissimilaridades, não considerando critérios prévios para inserção em qualquer grupo (Rafael, 2019).

A similaridade (ou a dissimilaridade - sua inversa) entre os elementos é uma medida para determinar as semelhanças entre as observações a serem agrupadas. As características utilizadas para expressar a similaridade são combinadas em um cálculo para todos os pares de elementos, de maneira que todos os objetos sejam comparados a todos os demais e que após essa comparação, os agrupamentos serão formados (Rafael, 2019).

Após obter a matriz das medidas de similaridade ou dissimilaridade, escolhe-se um método de agrupamento. Neste trabalho o método do *k-means* é adequado, pois ele busca minimizar a distância quadrática média dos pontos presentes no mesmo agrupamento, de maneira a encontrar uma configuração que diminua a variância interna dos grupos e maximize a variância entre os grupos (Santos, 2021). A medida de distância utilizada foi a euclidiana.

2.4 – Banco de Dados

O banco de dados é constituído por 42.831 pontos, que são consideradas as ocorrências avaliadas. Cada ponto é o centróide de um pixel de resolução 30 m x 30 m, cujo conjunto deles formam as imagens *raster* que recobrem a área de influência do trecho da ferrovia. Cada ponto contém as observações das 4 variáveis (Elevação, Declividade, Litologia, Pedologia). A variável água não foi avaliada neste artigo. Utilizando o R (R Core Team, 2024), foi aplicada a técnica de componentes principais no banco de dados e logo em seguida, realizada análise de agrupamentos.

3 RESULTADOS

3.1 – Mapas Temáticos

Os mapas temáticos relacionados aos fatores condicionantes foram confeccionados no *software* QGIS (QGIS.org, 2024) e são apresentados na Figura 1.

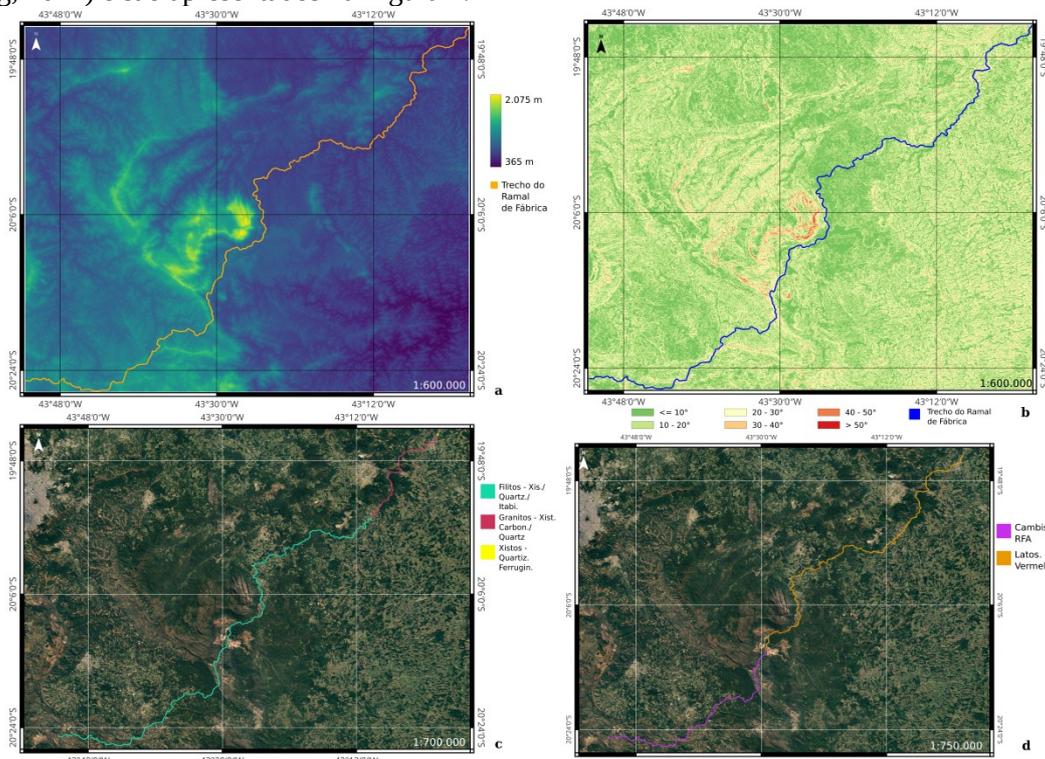


Figura 1. **a:** Elevação, **b:** Declividade, **c:** Setorização litológica, **d:** Setorização pedológica. O datum dos mapas é WGS 84 (EPSG: 4326). Imagem de fundo nos itens **c** e **d** - Maxar Technologies/Google.

3.2 – Análise de Componentes Principais (ACP)

A ACP permitiu a quantificação e redução da dimensionalidade do banco de dados original para as duas primeiras componentes. A primeira componente principal reteve 58,63% da variância original, enquanto a segunda componente reteve 25,22% da variância. Desta forma, as duas primeiras componentes foram capazes de explicar 83,85% da variância total do banco de dados. Este é um bom indicativo de empregabilidade da técnica, já que usualmente são aceitos valores acima de 70% para retenção das componentes.

Na Figura 2, os valores das componentes principais relacionadas à cada observação são representados em gráfico. Nela é possível observar um agrupamento natural dos trechos analisados em três grupos distintos.

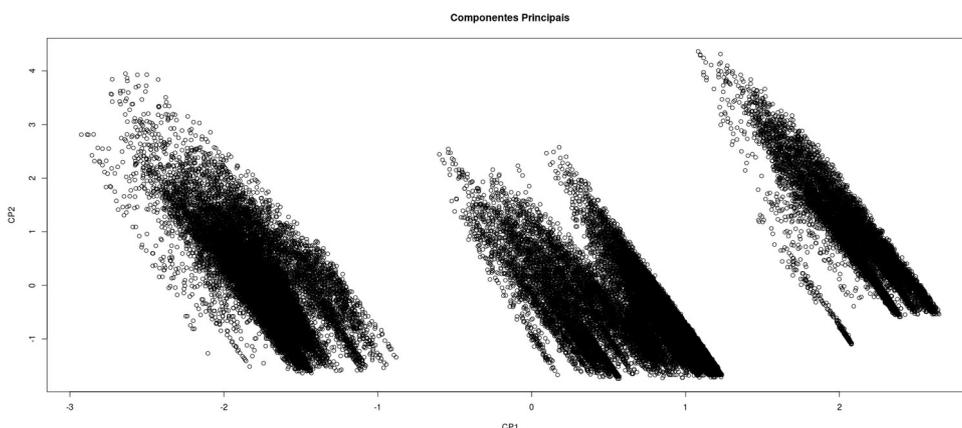


Figura 2. Componentes Principais

3.3 – Análise de Agrupamentos (*Clustering*)

O agrupamento dos trechos foram conduzidos utilizando a técnica de agrupamentos *k-means*, considerando a distância euclidiana. No agrupamento 1, foram alocados 17.476 pontos, no agrupamento 2, 16.614 pontos e no agrupamento 3, 8.741 pontos. A Figura 3 demonstra a alocação dos dados nos respectivos grupos.

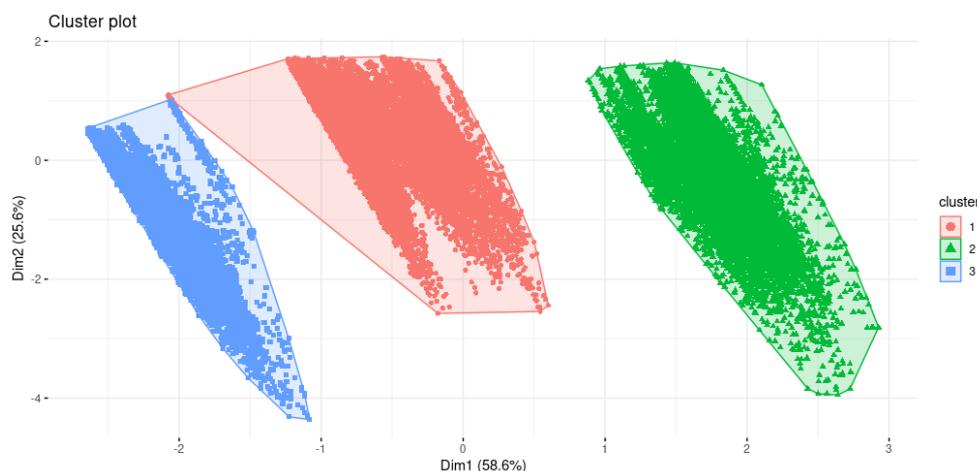


Figura 3. Alocação dos pontos conforme agrupamento.

A concessionária da ferrovia mantém um banco de registros contendo todos os eventos relacionados a deslizamentos que ocorreram ao longo do trecho viário. Entretanto, o método empregado neste artigo não prevê quais eventos ocorrerá, pois o método apenas divide os pontos em *clusters*. A Tabela 1 lista o número de eventos que ocorreram considerando a segmentação dos agrupamentos neste trabalho.

Tabela 1. Quantitativo de eventos ocorridos no trecho.

Cluster	Nº de Eventos	% do Total de Eventos
1	31	36
2	32	37
3	24	27

Os agrupamentos 1 e 2 concentram a maioria de eventos por também possuírem a maior quantidade de pontos em cada agrupamento.

3.4 – Validação dos Agrupamentos

Deseja-se que agrupamentos em algoritmos de classificação não-supervisionados apresentem grupos com baixa coesão (distância dentro dos grupos) e alta separação (distância entre grupos).

O coeficiente de silhueta (CS) foi utilizado neste trabalho como parâmetros de validação dos agrupamentos. Sua representação está ilustrada na Figura 4.

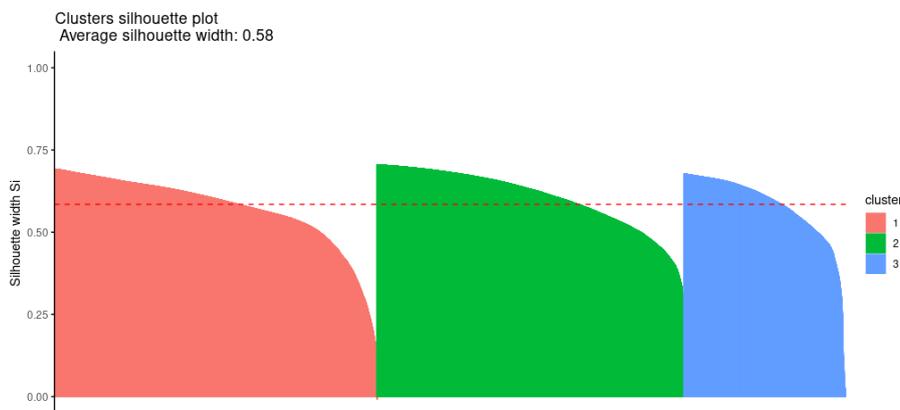


Figura 4. Gráfico de silhueta para os 3 agrupamentos.

O coeficiente de silhueta é definido no intervalo $[-1,1]$ e quanto mais próximo de 1, menos coeso e melhor separado é o agrupamento.

De forma geral, um CS entre 0,7 e 1 pressupõe forte agrupamento; entre 0,5 e 0,7 considerado um resultado aceitável; entre 0,5 e 0,25 revela uma fraca estrutura de agrupamento e finalmente, entre 0 e 0,25 indica que nenhuma estrutura considerável foi assinalada nos dados (Campos, 2019).

Os valores de silhueta obtidos neste estudo foram: *cluster 1* 0,57; *cluster 2* 0,60; *cluster 3* 0,58. Portanto, os agrupamentos revelaram-se aceitáveis.

4 CONCLUSÕES

As técnicas de estatística multivariada aplicada ao conjunto de dados apresentado, cujos fatores condicionantes avaliados (elevação, declividade, setorização litológica e setorização pedológica) demonstrou ser possível e adequada, com clara tendência para formação de 3 agrupamentos distintos.

O percentual dos eventos de deslizamentos ocorridos no passado e monitorados na ferrovia pela concessionária concentrou-se em sua maioria nos agrupamentos 1 e 2. A validação dos agrupamentos revelou que os grupos possuem coesão e separação aceitável. Como trabalho futuro, os grupos serão analisados internamente para que um rótulo possa ser atribuído. Posteriormente, esses grupos serão representados em ambiente SIG.

5 AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil, através do Projeto 402756/2023-1. Além disso, os autores agradecem à Universidade Federal de Ouro Preto (UFOP) e Empresa VALE S/A pelo apoio financeiro, bem como o apoio recebido do iMobilis UFOP (Laboratório de Computação Móvel), na pessoa do Prof. Mateus Coelho Silva.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ali, S.A.; Parvin, F.; Vojteková, J.; Costache, R.; Linh, N. T. T.; Pham, Q. B.; Vojtek, M.; Gigović, L.; Ahmad, A.; Ghorbani, M. A. (2021) GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. *Geoscience Frontiers*, Volume 12 (2), p.857-876.
- Bhat, I.A.; Ahmed, R.; Bhat, W.A.; Ahmed, P. (2023) Application of AHP based geospatial modeling for assessment of landslide hazard zonation along Mughal road in the Pir Panjal Himalayas. *Environ Earth Sci* 82, 336.
- Campos, S. L. S. (2019) *Busca não supervisionada de padrões por técnicas de agrupamento clássica e nebulosa*. 149f. Monografia. Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora.
- Das, S.; Sarkar, S. & Kanungo, D.P. (2022) GIS-based landslide susceptibility zonation mapping using the analytic hierarchy process (AHP) method in parts of Kalimpong Region of Darjeeling Himalaya. *Environ Monit Assess* 194, 234.
- Feng, J.; Feng, L.; Wang, J.; King, C. W. (2020) Evaluation of the onshore wind energy potential in mainland China—Based on GIS modeling and EROI analysis. *Resources, Conservation and Recycling*, 152:104484.
- Khan, F.; Das, B.; Mishra S. R. K.; Awasthy, M. (2022) A review on the Feasibility and Application of Geospatial Techniques in Geotechnical Engineering Field. *Materials Today: Proceedings*, Volume 49, Part 2, p. 311-319.
- Khusulio, K.; Kumar, R. (2023) Feasibility assessment of multi-criteria decision making and quantitative landslide susceptibility methods: A case study of Mao-Maram Manipur. *J Earth Syst Sci* 132, 56.
- Leonardi, G.; Palamara, R.; Manti, F.; Tufano, A. (2022) GIS-Multicriteria Analysis Using AHP to Evaluate the Landslide Risk in Road Lifelines. *Appl. Sci.* 12, 4707.
- Li, B.; Bin, L.; Wang, N.; Chen, J. (2021) GIS-Based Landslide Susceptibility Mapping Using Information, Frequency Ratio, and Artificial Neural Network Methods in Qinghai Province, Northwestern China. *Advances in Civil Engineering*. Vol. 2021:4758062.
- Ma, Z.; Qin, S.; Cao, C.; Lv, J.; Li, G.; Qiao, S.; Hu, X. (2019) The Influence of Different Knowledge-Driven Methods on Landslide Susceptibility Mapping: A Case Study in the Changbai Mountain Area, Northeast China. *Entropy*, 21(4):372.
- Nachappa, T. G.; Kienberger, S.; Meena, S. R.; Hölbling, D.; Blaschke, T. (2020) Comparison and validation of per-pixel and object-based approaches for landslide susceptibility mapping. *Geomatics, Natural Hazards and Risk*, 11:1, p.572-600.
- QGIS.org, 2024. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>
- R Core Team (2023). *_R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rafael, M. C. E. (2019) *Análise estatística multivariada para correlação de variáveis sísmográficas e geotécnicas em uma cavidade natural subterrânea – mina de ferro N4EN, Carajás*. 99 f. Dissertação (Mestrado Profissional em Engenharia Geotécnica) – Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto.

- Santos, A. E. M. (2021) *Classificação de maciços rochosos por meio de técnicas da estatística multivariada e inteligência artificial*. 195 f. Tese (Doutorado em Engenharia Mineral) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto.
- Santos, H. G. dos; Junior, W. de C.; Dart, R. de O.; Aglio, M. L. D.; Sousa, J. S. de; Pares, J. G.; Fontana, A.; Martins, A. L. da S.; Oliveira, A. P. de. (2011) O novo mapa de solos do Brasil: legenda atualizada. Rio de Janeiro: *Embrapa Solos*, 67 p. (Embrapa Solos. Documentos, 130.) 1 mapa, color. Escala 1:5.000.000. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/123772/1/DOC-130-O-novo-mapa-de-solos-do-Brasil.pdf>
- Santos, T. B dos. (2019) *Análise de risco geotécnico em taludes rochosos de mina com uso de técnicas estatísticas multivariadas e de aprendizado de máquina*. 106 f. Tese (Doutorado em Engenharia Mineral) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto.
- Senouci, R.; Taibi, N.-E.; Teodoro, A.C.; Duarte, L.; Mansour, H.; Yahia Meddah, R. (2021) GIS-Based Expert Knowledge for Landslide Susceptibility Mapping (LSM): Case of Mostaganem Coast District, West of Algeria. *Sustainability*. 13(2):630.
- Silva, M. A. da; Pinto, C. P.; Pinheiro, M. A. P; Mainho, M. de S.; Lombello, J. C.; Pinho, J. M. M. P.; Goulart, L. E. A.; Magalhães, J. R. (2020) Mapa geológico do estado de Minas Gerais. Belo Horizonte: *CPRM*. 1 mapa, color. Escala 1:1.000.000.
- Tešić, D. (2021) Landslide susceptibility mapping using GIS along the Niš-North Macedonia highway. *Zbornik radova Departmana za geografiju, turizam i hotelijerstvo*. 01. p.1-13.
- Taloor, A. K.; Adimalla, N.; Goswami, A. (2021) Remote Sensing and GIS applications in Geoscience. *Applied Computing and Geosciences*, 11:100065.
- Wei, J.; Chen, G.; Huang, J.; Xu, L.; Yang, Y.; Wang, J.; Sadick, A.-M. (2021) BIM and GIS Applications in Bridge Projects: A Critical Review. *Applied Sciences*, 11(13):6207.
- Yamusa, I.B.; Ismail, M.S.; Tella, A. (2022) Highway Proneness Appraisal to Landslides along Taiping to Ipoh Segment Malaysia, Using MCDM and GIS Techniques. *Sustainability*; 14(15):9096.