AVALIAÇÃO DO USO E COBERTURA DO SOLO UTILIZANDO IMAGENS DO SATÉLITE SENTINEL-2 NO ESTUDO DE DESASTRES AMBIENTAIS: O CASO DO ROMPIMENTO DA BARRAGEM DE BRUMADINHO-MG

Kariny Moreira Rocha[1], Lídia Raiza Sousa Lima Chaves Trindade[2], Romário Oliveira de Santana[3]

[1]Instituto Federal da Bahia, Vitória da Conquista, BA, Brasil, karinymoreirarocha@gmail.com; [2]Universidade de São Paulo, Instituto de Matemática e Estatística, Butantã, SP, Brasil, lidiaraiza@hotmail.com; [3]Empresa Brasileira de Pesquisa Agropecuária, Embrapa Recursos Genéticos e Biotecnologia (CENARGEM), Brasília, DF, Brasil, romarioambiental1@gmail.com

RESUMO

O rompimento da barragem de rejeitos B1 em Brumadinho-MG, em 2019, resultou em um dos maiores desastres ambientais do Brasil, devastando aproximadamente 297 hectares e causando impactos significativos no ecossistema local. O estudo utilizou imagens do satélite Sentinel-2, geoprocessamento e o software RStudio para avaliar as mudanças no uso e cobertura do solo antes e após o desastre. Algoritmos de classificação supervisionada, como Random Forest (RF) e Support Vector Machine (SVM), foram aplicados para mapear a área afetada. Ambos os algoritmos apresentaram acurácia global e coeficiente Kappa superiores a 90%, com o RF demonstrando desempenho ligeiramente superior, especialmente na classificação de vegetação e áreas de mineração. Foi identificada uma redução na cobertura vegetal entre 2019 e 2021, acompanhada de um aumento em áreas de solo exposto e mineração, destacando a importância do uso de aprendizado de máquina e sensoriamento remoto para identificar mudanças ambientais e apoiar ações de recuperação em desastres.

Palavras-chave — Classificação supervisionada, algoritmos de aprendizado de máquina, mineração, sensoriamento remoto.

ABSTRACT

The rupture of the B1 tailings dam in Brumadinho-MG, in 2019, resulted in one of the largest environmental disasters in Brazil, devastating approximately 297 hectares and causing significant impacts on the local ecosystem. This study used Sentinel-2 satellite imagery, geoprocessing, and RStudio software to assess changes in land use and land cover before and after the disaster. Supervised classification algorithms, such as Random Forest (RF) and Support Vector Machine (SVM), were applied to map the affected area. Both algorithms demonstrated overall accuracy and Kappa coefficient values above 90%, with RF showing slightly superior performance, particularly in classifying vegetation and mining areas. A reduction in vegetation cover was identified between 2019 and 2021, accompanied by an increase in exposed soil and mining areas, highlighting the importance of using machine learning and remote sensing to identify environmental changes and support recovery actions in disaster scenarios.

Key words — *Supervised classification, machine learning algorithms, mining, remote sensing.*

1. INTRODUÇÃO

A mineração envolve a extração de minerais e recursos naturais do subsolo, essencial para várias indústrias e o desenvolvimento econômico de diversos países. No Brasil, essa prática tem raízes na busca por ouro no período colonial, quando bandeirantes desbravaram novos territórios, levando à ocupação de regiões ricas em recursos, como Minas Gerais [1]. Contudo, a mineração gera resíduos que precisam ser armazenados, comumente em barragens de rejeitos, estruturas que apresentam riscos ambientais, especialmente em casos de falhas. Exemplos notórios são os desastres de Mariana e Brumadinho, em 2015 e 2019, respectivamente, que causaram graves impactos ambientais, afetando extensas áreas e contaminando ecossistemas, incluindo rios e vegetação da Mata Atlântica [2].

Para reduzir a necessidade de visitas a campo e custos, o sensoriamento remoto tem se mostrado uma ferramenta eficaz na análise de desastres ambientais, permitindo a obtenção de informações a partir de imagens que medem a radiação refletida ou emitida pela superfície terrestre [3]. Desde a década de 70, o desenvolvimento da tecnologia de sensoriamento remoto tem permitido o mapeamento do uso e cobertura do solo, com avanço significativo após o lançamento da série de satélites Sentinel em 2014, dentro do Programa Copernicus da Agência Espacial Europeia e da Comissão Europeia [4].

Segundo Oliveira et al. [5], o uso e a cobertura do solo são indicadores para avaliar o estado ambiental e os efeitos da intervenção humana. Esses dados podem ser sintetizados por meio de mapas obtidos com imagens de satélite, representando a distribuição espacial da ação humana no

ambiente. Para a obtenção desses dados, técnicas de classificação supervisionada são amplamente empregadas, com o uso de amostras de treinamento que identificam classes específicas na imagem [6]. Recentemente, algoritmos de aprendizado de máquina, como o *Random Forest* (RF) e o *Support Vector Machine* (SVM), têm se destacado por sua precisão na classificação de paisagens [7].

Assim, o sensoriamento remoto, aliado a técnicas avançadas de classificação, oferece potencial para avaliar o impacto ambiental e socioeconômico dos desastres de mineração. Este estudo aplicou técnicas de classificação supervisionada com dados do Sentinel-2 para analisar os impactos socioambientais resultantes do desastre de Brumadinho, evidenciando a importância do uso de tecnologias para monitoramento e mitigação de efeitos nocivos em futuras avaliações de impactos.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

O estudo concentra-se na área impactada pela lama de rejeitos resultante do rompimento da barragem B1 da Mina do Córrego do Feijão, localizada em Brumadinho, Minas Gerais. Essa área se estende da barragem até o Rio Paraopeba, abrangendo as coordenadas (20°07'00"S, 44°07'02"W) e (20°09'25"S, 44°09'32"W) [8]. A mina, operada pela Vale, possuía um conjunto de barragens posicionadas acima de Brumadinho, na Região Metropolitana de Belo Horizonte. O desastre causou o deslocamento de cerca de 12 milhões de metros cúbicos de rejeitos [9] (Figura 1).

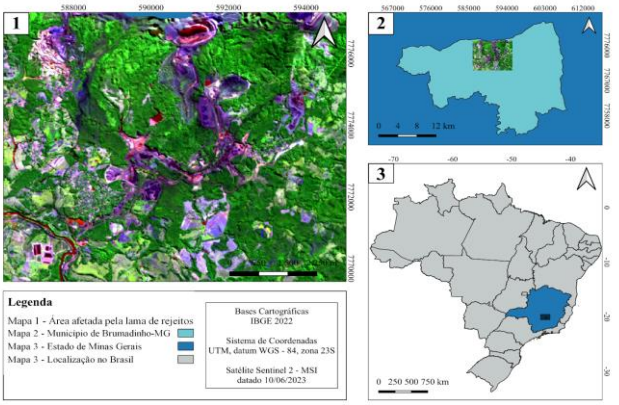


Figura 1. Área afetada pelo rompimento da barragem, em Brumadinho, Minas Gerais.

2.2. Coleta de dados

Para análise da área, foram utilizadas imagens do satélite Sentinel-2, obtidas gratuitamente pelo site Copernicus. As imagens, datadas de 07 de janeiro de 2019 (antes do rompimento) e 31 de janeiro de 2021 (cerca de dois anos após o desastre), foram selecionadas pela qualidade e ausência de nuvens, ambas capturadas durante o verão. As imagens têm resolução espacial de 20 m nas bandas espectrais B02, B03, B04, B05, B06, B07, B8A, B11 e B12, e o Sentinel-2 tem capacidade de revisita a cada 10 dias.

2.3 Processamento digital das imagens

Para este estudo, as imagens do satélite Sentinel-2 foram baixadas do portal Copernicus. Em seguida, os arquivos *raster* foram abertos no software R (versão 4.3.1, RStudio 2023.09.1), utilizando o pacote *raster*. Por meio da função *stack*, todas as bandas espectrais (B02, B03, B04, B05, B06, B07, B8A, B11, B12) foram agrupadas em um único objeto, facilitando o processamento subsequente.

Os pacotes *rgdal* e *sp* foram carregados no RStudio para manipulação de dados vetoriais. O *shapefile* do município de Brumadinho, obtido do IBGE, foi importado usando a função *readOGR* e ajustado para a mesma projeção geográfica das imagens satélite por meio da função *spTransform.* A área de interesse foi delimitada pelo *shapefile*, e a função *crop* foi utilizada para recortar o *raster*.

2.4 Classificação supervisionada por algoritmos de aprendizado de máquina

Para a classificação das imagens, aplicaram-se os algoritmos *Random Forest* (RF) e *Support Vector Machine* (SVM) no RStudio, utilizando os pacotes *caret*, *rgdal*., *raster*, *dplyr*, *caTools*, *randomForest*, e *e1071*.

Inicialmente, as bandas espectrais foram agrupadas em um único arquivo *raster* por meio da função *stack*, facilitando o processamento subsequente. O arquivo vetorial contendo as classes amostrais foi utilizado para gerar uma base de dados com os valores de pixel (Números Digitais) correspondentes a cada banda e índice espectral. A associação entre os pixels e as classes temáticas foi realizada com a função *merge*, criando uma base de dados estruturada para o treinamento dos algoritmos.

Para o algoritmo *RF*, foram configurados os seguintes parâmetros: o número de árvores foi definido como 500 (*ntree* = 500), e o número de variáveis preditoras selecionadas aleatoriamente em cada nó foi ajustado automaticamente pelo modelo *mtry*. Já para o algoritmo *SVM*, foi utilizado o kernel radial (Radial Basis Function - RBF), com os parâmetros de custo (*C*) ajustado para 1 e o parâmetro *gamma* ajustado automaticamente por validação cruzada.

Para validação dos algoritmos foi utilizada a técnica de validação cruzada *k-fold*, com 70% das amostras destinadas ao treinamento e 30% para validação. A divisão foi realizada de forma aleatória utilizando a função *sample.split*, garantindo a reprodutibilidade dos resultados por meio da função *set.seed(1234).* Foram extraídas manualmente 350 amostras homogêneas a partir de polígonos vetoriais que representavam as classes temáticas (vegetação, solo exposto, água, áreas urbanas, agrícolas e pastagens). A distribuição espacial das amostras foi planejada para cobrir toda a extensão da área de estudo, assegurando a representatividade

de todas as classes temáticas e minimizando possíveis vieses na classificação.

Após o treinamento, os modelos foram aplicados ao conjunto de validação, e as matrizes de confusão foram geradas com a função *predict.* Essas matrizes permitiram avaliar o desempenho dos algoritmos em termos de erros de omissão e comissão para cada classe temática. A acurácia global foi utilizada como principal métrica de avaliação, complementada pelo coeficiente Kappa, que avalia a concordância entre os resultados da classificação e os dados de referência, considerando o acaso.

3. RESULTADOS E DISCUSSÃO

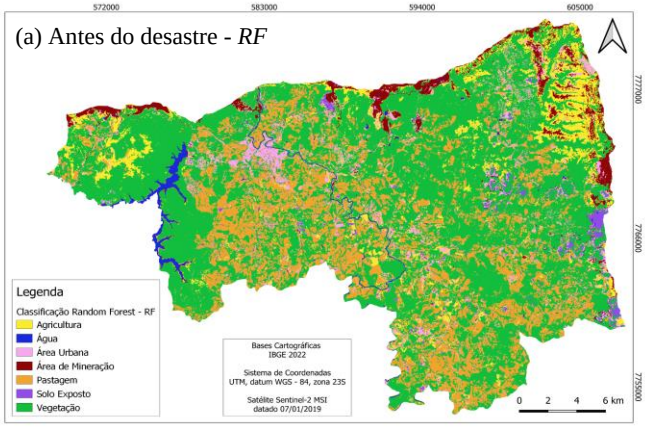
Na análise da imagem anterior ao rompimento da barragem, em 07/01/2019 (Figura 2a), o algoritmo *Random Forest* (RF) alcançou acurácia global de 0,9939 e um Kappa de 0,9895, demonstrando excelente desempenho na classificação do uso do solo. O *Support Vector Machine* (SVM), por sua vez, antes do desastre (Figura 2b) apresentou acurácia global de 0,9749 e Kappa de 0,9562, resultados sarisfatórios, mas inferiores ao RF. Na imagem posterior ao rompimento, em 31/01/2021 (Figura 2c), o RF manteve seu desempenho elevado, com acurácia de 0,9924 e Kappa de 0,984, enquanto o SVM apresentou acurácia de 0,9755 e Kappa de 0,947 (Figura 2d), mas ainda assim ficou atrás do RF.

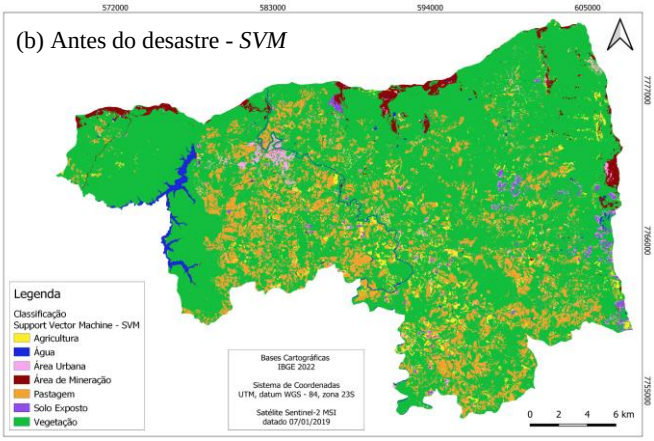
Um teste de significância foi realizado para comparar os coeficientes Kappa dos dois algoritmos, utilizando o método descrito por Congalton e Green [10]. O valor do Kappa calculado foi de 0,847, indicando uma concordância substancial entre os classificadores e os dados de referência. O teste apresentou um valor de $z = 0$ e p-valor = 1, indicando que não há diferença estatisticamente significativa entre os desempenhos dos algoritmos ($p \geq 0,05$).

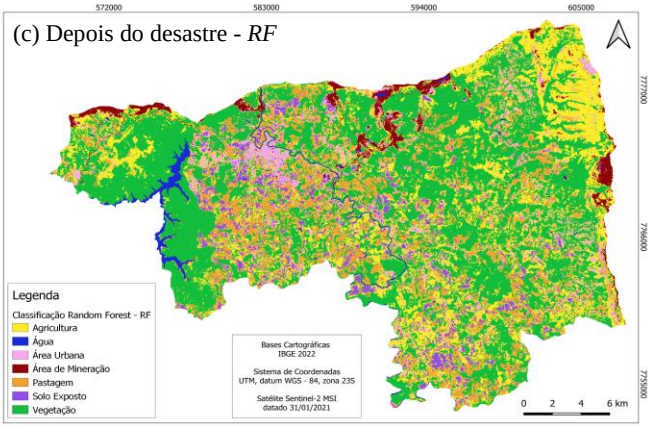
Ambos os algoritmos possuem desempenhos equivalentes na classificação das classes temáticas analisadas, com alta concordância em relação aos dados de referência. No entanto, os algoritmos apresentaram erros de omissão e comissão em diferentes classes temáticas. O RF demonstrou menor quantidade de erros, além de um desempenho ligeiramente superior em termos de acurácia e Kappa, especialmente na classificação de vegetação e áreas de mineração, enquanto o SVM apresentou maior confusão em algumas classes, como áreas urbanas e solo exposto. Essa confusão pode ser atribuída à semelhança espectral entre determinadas classes, como vegetação e agricultura, além de fatores como a resolução espacial das imagens e a heterogeneidade da superfície.

Estudos anteriores, como os de Wolfer et al. [11] e Alba [12], corroboram esses achados, destacando a eficiência dos algoritmos RF e SVM, com o RF frequentemente apresentando resultados superiores em termos de acurácia e Kappa. Alves et al. [13] também identificaram a superioridade do RF em estudos de classificação supervisionada, utilizando imagens de satélites Landsat-8, Sentinel-1 e Sentinel-2.

Os mapas gerados pela classificação supervisionada, ilustrados nas Figuras 2 (a, b, c e d), mostram as sete classes de uso e cobertura da terra mapeadas neste estudo. A análise dos mapas pós-rompimento revelou mudanças significativas na cobertura do solo, incluindo indícios de aumento em áreas de mineração. No entanto, essas mudanças podem estar sujeitas a limitações na classificação, como a semelhança espectral entre solo exposto e áreas de mineração, ressaltando a importância de considerar as nuances e limitações de cada algoritmo na escolha para aplicações específicas.







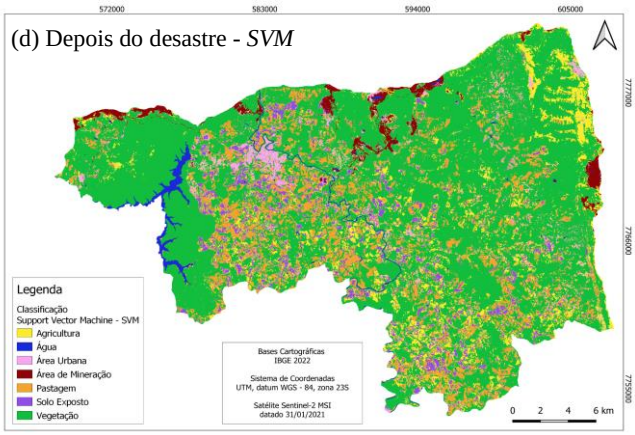


Figura 2 - Resultado da classificação supervisionada para os algoritmos RF (a) e SVM (b) antes do rompimento, e RF (c) e SVM (d), após o rompimento da barragem, em Brumadinho, MG.

4. CONCLUSÕES

Os algoritmos *Random Forest* (RF) e *Support Vector Machine* (SVM) demonstraram desempenhos satisfatórios na classificação do uso e cobertura do solo, alcançando acurácia global e coeficiente Kappa superiores a 90%. Apesar de ambos apresentarem resultados consistentes, o RF destacou-se com um desempenho ligeiramente superior, especialmente na identificação de áreas de vegetação e mineração.

Entre 2019 e 2021, foi constatada uma redução significativa na cobertura vegetal, acompanhada por um aumento em áreas agricultáveis e de solo exposto. Também foi observado um crescimento nas áreas de mineração após o rompimento da barragem.

Os resultados do estudo evidenciaram o impacto ambiental causado pelo desastre de Brumadinho, com a área afetada estimada em aproximadamente 298,55 hectares. Além disso, destacaram o potencial do sensoriamento remoto, aliado a algoritmos de aprendizado de máquina, como uma ferramenta indispensável no monitoramento de desastres ambientais e para subsidiar a tomada de decisões em ações de recuperação e mitigação de impactos.

5. REFERÊNCIAS

[1] Agência Nacional de Mineração (ANM). *Anuário Mineral Brasileiro 2022: Relatório Preliminar*. Brasília, DF, 2022. Available at: https://www.gov.br/anm/pt-br/centrais-de-conteudo/publicacoes/serie-estatisticas-e-economia-mineral/anuario-mineral/anuario-mineral-brasileiro/PreviaAMB2022.pdf. Accessed on: Oct. 05, 2024.

[2] C. Freitas et al. Da Samarco em Mariana à Vale em Brumadinho: desastres em barragens de mineração e Saúde Coletiva. *Cad. Saúde Pública*, 2019.

[3] J. de C. Leopoldino et al. Fernandes. Spatial and temporal analysis of the collapse of the tailings dam in Brumadinho, Brazil. *Revista Agrogeoambiental*, 12(3), 2020.

[4] Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). *SENTINEL - Missão Sentinel*. 2020. Available at: https://www.embrapa.br/satelites-de-monitoramento/missoes/sentinel. Accessed on: Apr. 05, 2024.

[5] G. de C. Oliveira et al. Classificação supervisionada como ferramenta para avaliação da dinâmica do uso e cobertura do solo. *Enciclopédia Biosfera*, Centro Científico Conhecer, Goiânia, 10(18): 924-935, 2014.

[6] Instituto Nacional de Pesquisas Espaciais (INPE). *SPRING: Tutorial de Geoprocessamento*. São José dos Campos, 2006. Available at: http://www.dpi.inpe.br/spring/portugues/tutorial/index.html. Accessed on: Oct. 10, 2024.

[7] C. Li et al. Gong. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sensing*, 6(2): 964-983, 24 jan. 2014.

[8] Google. *Google Earth website*. Available at: https://www.google.com/intl/pt-BR/earth/about/. Accessed on: Oct. 10, 2024.

[9] Governo do Estado Minas Gerais. *Histórico do rompimento das barragens da Vale na Mina Córrego do Feijão*. 20 dez. 2021. Available at: https://www.mg.gov.br/pro-brumadinho/pagina/historico-do-rompimento-das-barragens-da-vale-na-mina-corrego-do-feijao. Accessed on: Oct. 10, 2024.

[10] R. G. Congalton, K. Green. *Assessing the accuracy of remotely sensed data: principles and practices*. 2nd ed. Boca Raton: CRC Press, 2009.

[11] M. F. Wolfer et al. Análise do Uso e Cobertura da Terra Utilizando Imagens Sentinel-2A e Inteligência Artificial. *Anuário do Instituto de Geociências - UFRJ*, 43(4): 395-403, 2020.

[12] E. Al-Ba et al. Comparação entre Algoritmos de Aprendizado de Máquina para a Identificação de Floresta Tropical Sazonalmente Seca. *Anuário do Instituto de Geociências*, 45, 2022, pp. 1-10, Universidade Federal do Rio de Janeiro.

[13] M. de C. Alves et al. Multisensor analysis for environmental targets identification in the region of Funil dam, state of Minas Gerais, Brazil. *Applied Geomatics*, 15(4): 807-827, 19 set. 2023.