

MÉTODO RÁPIDO DE CLASSIFICAÇÃO GEOMORFOLÓGICA POR ALGORITMO DE APRENDIZADO DE MÁQUINA APLICADO EM SUB-BACIAS DO RIO SÃO FRANCISCO

Cristiano Marcelo Pereira Souza¹
Lucas Augusto Pereira da Silva²
Sirius Oliveira Souza³
Kleber Carvalho Lima⁴
Marcos Esdras Leite⁵

INTRODUÇÃO

O mapeamento geomorfológico é essencial para a compreensão e gestão do ambiente natural, exigindo uma abordagem multidisciplinar que integra Geologia, Geografia e Sensoriamento Remoto (SREENIVASAN; JHA, 2022; FAYE et al., 2024). Técnicas de sensoriamento remoto e geoprocessamento são indispensáveis, mas a subjetividade na definição dos limites das unidades de relevo pode levar a variações nos resultados (GIACCONE et al., 2022; VAN DER MEIJ et al., 2022)

Um dos principais desafios do mapeamento geomorfológico é conciliar abordagens metodológicas com técnicas objetivas para definir formas do relevo. Ferramentas como lógica matemática mostram-se promissoras, mas não se deve perder os critérios teóricos geomorfológicos para definir unidades (HOUSER et al., 2022). No Brasil, esses estudos ainda são iniciais (SILVEIRA et al., 2014; LANA et al., 2022).

Avanços na coleta e processamento de dados espaciais oferecem oportunidades para melhorar os mapas geomorfológicos (SREENIVASAN; JHA, 2022). Técnicas semiautomáticas, como algoritmos de aprendizado de máquina, têm sido aplicadas em mapeamentos de solos, litologias e relevo (GOMES et al., 2019; FAYE et al., 2024). No contexto de dados, o modelo digital de elevação é uma fonte valiosa de covariáveis para capturar a distribuição espacial das unidades geomorfológicas, e a seleção criteriosa dessas covariáveis pode aumentar a robustez da modelagem (SOFIA, 2020). Neste contexto, o objetivo desse estudo foi aplicar o algoritmo Random Forest (RF) para a predição de unidades do relevo.

¹ Professor Colegiado de Geologia, Universidade Federal do Vale do São Francisco - UNIVASF, cristiano.souza@univasf.edu.br

² Professor Departamento de Geociências, Universidade Estadual de Montes Claros, lucaskaio1605@gmail.com

³ Professor Colegiado de Geografia, Universidade Federal do Vale do São Francisco - UNIVASF, sirius.souza@univasf.edu.br

⁴ Professor adjunto do Departamento de Geografia da Universidade de Pernambuco – UPE, kleber.carvalho@upe.br

⁵ Professor Departamento de Geociências, PPGE – Universidade Estadual de Montes Claros, marcos.leite@unimontes.br

MATERIAIS E MÉTODOS

Área de estudo

O estudo foi conduzido na região do Submédio São Francisco, abrangendo áreas dos estados da Bahia e Pernambuco, no Nordeste brasileiro (Figura 1). Esta região apresenta uma diversidade geológica significativa, resultando em variadas unidades de relevo. A área é dominada por uma grande superfície rebaixada, formando uma depressão geológica. Na porção jusante da bacia, o afloramento de material geológico cristalino é recorrente, enquanto a porção montante é mais predominantemente recoberta por sedimentos e metassedimentos de diferentes origens e idades.

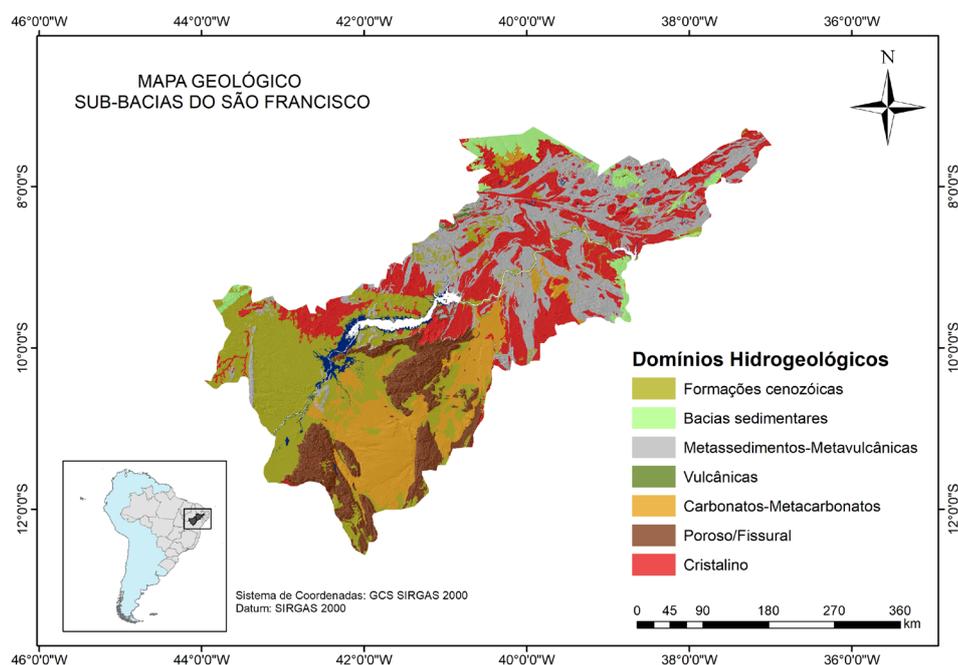


Figura 1. Localização da área de estudo com principais domínios hidrogeológicos das sub-bacias do rio São Francisco.

Rochas sedimentares do Proterozoico e Mesoproterozoico formam planaltos com altitudes de até 1400 metros, principalmente na porção sudeste da bacia. As serras, ocasionalmente isoladas, são compostas por rochas mais resistentes, como metassedimentos ou rochas cristalinas. Chapadas, com presença de escarpas ou patamares erodidos na parte frontal, são predominantes na porção norte da bacia. Essas formações são constituídas pela Cobertura Detrito-Laterítica Paleogênica, marcando distintamente a paisagem. O eixo central da bacia é caracterizado por planícies fluviais com sedimentos do Holoceno.

Procedimentos Metodológicos

A metodologia foi estruturada em três etapas principais: definição de amostras, obtenção das covariáveis preditivas e treinamento e teste da predição. Todas as etapas foram realizadas com o software R 2023.

Definição de Amostras – Foi criado um arquivo de pontos com as classes de relevo, baseado em mapas geomorfológicos anteriores (IBGE, 2023). Utilizamos o MDE Copernicus (30 x 30 m), imagens Landsat 8, imagens de alta resolução e dados geológicos para alocação dos pontos, totalizando 5.700 pontos amostrais. As amostras corresponderam à variável dependente (y).

Tabela 1. Lista de covariáveis usadas para predição do relevo.

Variável	Definição
Aspect	Representa a direção da inclinação da superfície do terreno em relação ao norte.
Convexity	Mede a convexidade local de uma superfície, indicando se a área é côncava ou convexa.
Cross Curvature	Curvatura transversal à direção do fluxo, utilizada para identificar formas de relevo como vales e cristas.
MDE Copernicus	Modelo Digital de Elevação com resolução de 30m, utilizado para diversas aplicações em planejamento urbano e gestão de recursos naturais.
Flow Line	Linhas que representam o caminho de fluxo da água sobre a superfície do terreno.
General Curvature	Curvatura geral da superfície, combinando curvatura longitudinal e transversal.
Geomorphons	Mapas de formas de relevo gerados a partir da análise de padrões geomorfológicos.
Hillshade	Sombreado do terreno, utilizado para visualização tridimensional das formas de relevo.
Landform	Classificação das formas de relevo em diferentes categorias morfológicas.
Landsat 8 Bandas 1 a 7	Imagens de satélite que capturam diferentes comprimentos de onda da luz, utilizadas para análise da cobertura do solo e vegetação.
Longitudinal Curvature	Curvatura na direção do fluxo da água, influenciando a velocidade e o padrão de fluxo.
Maximal Curvature	A maior curvatura presente na superfície em uma direção específica.
Minimal Curvature	A menor curvatura presente na superfície em uma direção específica.
MRRTF	Multi-resolution Ridge Top Flatness, um índice que identifica áreas planas e topos de cristas.
MRVBF	Multi-resolution Valley Bottom Flatness, um índice que identifica áreas planas nos fundos de vales.
Planície Fluvial	Covariável binária que diferencia vales fluviais de outras áreas, baseada em depósitos fluviais.
Plan Curvature	Curvatura plana, relacionada à convergência ou divergência do fluxo sobre a superfície do terreno.
Profile Curvature	Curvatura do perfil, influenciando a aceleração ou desaceleração do fluxo ao longo da superfície do terreno.
Ridge Level	Nível da crista, uma medida utilizada para identificar cristas e cumes no terreno.
Slope	Inclinação da superfície do terreno, expressa em graus ou porcentagem.
Surface	Representação da superfície topográfica em termos de elevação.
Tangential Curvature	Curvatura tangencial, similar à curvatura plana, mas focada na forma da superfície perpendicular ao fluxo principal.
TPI	Topographic Position Index, um índice que classifica a posição topográfica de um ponto em relação às suas vizinhanças.
Valley Depth	Profundidade do vale, medida da diferença de elevação entre o fundo do vale e os cumes adjacentes.

Covariáveis Preditivas (Tabela 1) – Extraímos atributos do terreno do MDE Copernicus, uma fonte global e atualizada ideal para diversas aplicações. Usando o R (RCORE, 2024) com os pacotes "rsaga" e "raster", obtivemos 22 atributos do terreno. Incluímos também sete covariáveis das bandas do Landsat 8, úteis para identificar variações de cobertura do solo e vegetação, e uma covariável geológica binária para diferenciar vales fluviais. Ao final, uma matriz de dados com a variável dependente (y) e variável independente (x) de 30 covariáveis foi criada para procedimentos posteriores

Treinamento e Teste da Predição – Para mitigar problemas com excesso de covariáveis, usamos a função FindCorrelation para eliminar covariáveis com alta correlação (coeficiente de Pearson superior a 0,8) (SOUZA et al., 2018). As amostras foram divididas em 75% para treinamento e 25% para validação. Utilizamos o modelo Random Forest (RF), um algoritmo baseado no método Bootstrap, que constrói várias árvores de decisão, reduzindo a correlação entre elas e diminuindo as chances de overfitting (BREIMAN, 2001). O treinamento e ajuste de hiperparâmetros foram realizados por validação cruzada com cinco grupos (k-folds) e 10 repetições. A importância das covariáveis foi determinada pela função VarImp, e a performance do RF foi avaliada com os 25% das amostras do treinamento, calculando-se o índice Kappa e a Acurácia Global.

RESULTADOS E DISCUSSÃO

Desempenho do Modelo

O modelo RF mostrou-se eficiente e robusto, fundamentado em diversos dados ambientais para definir as unidades de relevo, confirmando sua eficácia em lidar com grandes conjuntos de dados e variáveis complexas (GOMES et al., 2019; FAYE et al., 2024). A aplicação da função FindCorrelation resultou em um modelo parcimonioso, conforme recomendado em estudo com aprendizado de máquina (SOUZA et al., 2018; GOMES et al., 2019; ZHANG et al., 2023).

Entre as covariáveis mais importantes, destacaram-se as derivadas do MDE Copernicus (Figura 2), especialmente a altitude, crucial para distinguir formas de relevo em regiões com variação altimétrica significativa (PELECH et al., 2019). A segunda covariável mais importante foi uma variável binária que delimita planícies fluviais, baseada nos depósitos do rio São Francisco. Esta covariável é importante para identificar padrões geomorfológicos específicos associados a processos fluviais. Algumas

covariáveis das bandas do Landsat 8 também foram importantes, pois elas captam distinções na paisagem, estratificando ambientes pedogeomorfológicos a partir de aspectos da vegetação (XU et al., 2024).

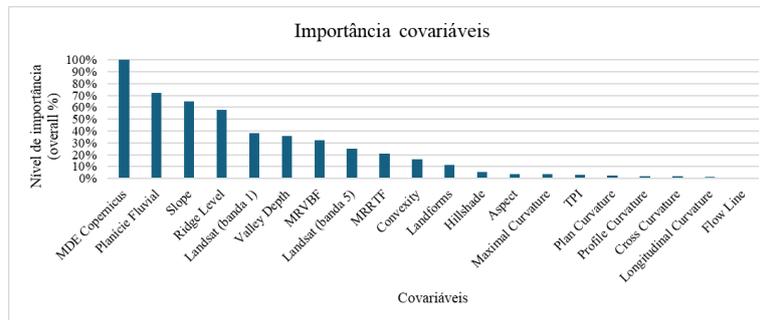


Figura 2. Ranking de importância (Overall%) das covariáveis usadas para treinamento do algoritmo Random Forest

O modelo RF obteve uma acurácia global de 95% e um índice Kappa de 0,86, indicando excelente performance (Rwanga & Ndambuki, 2017). Os erros de omissão variaram entre 4,70% e 20%, enquanto os erros de comissão variaram de 5,36% a 19%. Em geral, quanto menores forem os valores de erros de comissão e erros de omissão, maior será a acurácia e confiabilidade do mapeamento (RWANGA; NDAMBUKI, 2017). As classes mais problemáticas foram serras e planaltos, devido à semelhança morfológica.

Tabela 2. Matriz de confusão obtida a partir de classificação de unidades do relevo pelo algoritmo Random Forest. Valores acurácia global e índice Kappa.

Classes	Chapadas	Chapadas	Chapadas	Chapadas	Chapadas	Total	E.C (%)
Chapadas	251	19	5	0	7	282	8.06
Depressão	5	527	6	9	6	553	6.39
Serras	10	6	208	0	34	258	19.07
Planícies	0	4	0	159	0	163	5.36
Planaltos	7	7	38	0	275	327	14.60
Total	273	563	257	168	322	1583	
E.C (%)	8.1	6.4	19.1	5.4	14.6		
Acurácia Global	0.90						
Kappa	0.87						

Análise Espacial

A análise espacial evidenciou a eficiência do aprendizado de máquina na identificação e delineamento das unidades de relevo (Figura 3 e Figura 4). Contudo, algumas unidades requereram reclassificação manual e aplicação de filtros para eliminar pequenas áreas que poluem visualmente o mapa, destacando a importância da expertise técnica (VAN DER MEIJ et al., 2022). As planícies fluviais foram bem delimitadas,

seguindo os limites do rio São Francisco, com uma contribuição significativa da covariável vales fluviais, segunda covariável mais importante (Figura 3 e Figura 4). A unidade depressão foi mapeada extensivamente, consistente com mapeamentos prévios. Os planaltos foram refinadamente delimitados, obedecendo a aspectos altimétricos bem definidos, uma vantagem das técnicas automatizadas em comparação aos métodos manuais.

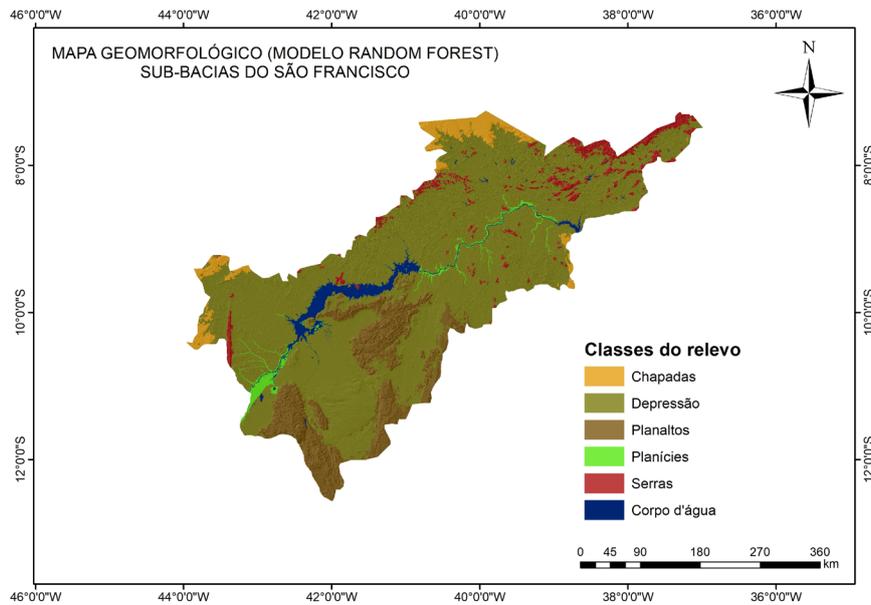


Figura 3. Mapa geomorfológico da área de estudo predito com algoritmo Random Forest.

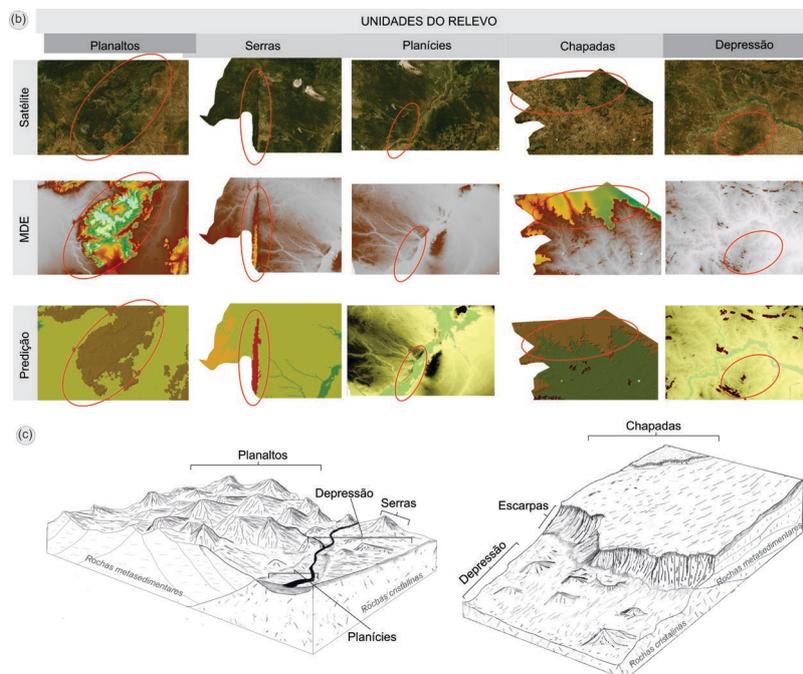


Figura 4. (a) Imagem de satélite, MDE e formas de relevo predita pelo Random Forest para cada unidade de relevo da área de estudo. (b) blocos diagramas representando cada unidade do relevo.

CONSIDERAÇÕES FINAIS

O uso do algoritmo Random Forest (RF) mostrou-se um método eficiente e rápido para classificar unidades de relevo, apoiado em dados robustos e na indicação de especialistas. Embora os métodos automatizados sejam eficazes, ajustes manuais baseados em expertise técnica são necessários para refinar os resultados.

A falta de padronização metodológica na definição das unidades de relevo é um desafio significativo. A obtenção de amostras precisas, especialmente em mapas de escala nacional, continua problemática. A diversidade de abordagens e a necessidade de ajustes locais indicam que, apesar das melhorias tecnológicas, a contribuição dos especialistas é essencial para garantir a precisão e a confiabilidade dos mapas geomorfológicos.

Este estudo destaca a importância de uma abordagem híbrida, que combine técnicas automatizadas com validação e ajustes realizados por especialistas, assegurando mapas geomorfológicos precisos e úteis para diversas aplicações.

Palavras-chave: Relevo. Random Forest. Classificação Automática. Variáveis Topográficas.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo projeto processo: 408197/2021-8. Também pela implementação das bolsas de Pós-doutorado e e bolsa pesquisador CNPq.

REFERÊNCIAS

- BREIMAN, L. Random forests. **Machine Learning**, 45, n. 1, 5-32, 2001.
- FAYE, P. A. L.; BRUNEL, E.; CLAVERIE, T.; MANOU-ABI, S. M.; DABO-NIANG, S. Automatic geomorphological mapping using ground truth data with coverage sampling and random forest algorithms. **Earth Science Informatics**, 2024.
- GIACCONE, E.; ORIANI, F.; TONINI, M.; LAMBIEL, C.; MARIÉTHOZ, G. Using data-driven algorithms for semi-automated geomorphological mapping. **Stochastic Environmental Research and Risk Assessment**, 36, n. 8, 2115-2131, 2022.
- GOMES, L. C.; FARIA, R. M.; SOUZA, E.; VELOSO, G. V.; SCHAEFER, C. E. G. R.; FERNANDES FILHO, E. I. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, 340, 337-350, 2019.
- HOUSER, C.; LEHNER, J.; SMITH, A. The field geomorphologist in a time of artificial intelligence and machine learning. **Annals of the American Association of Geographers**, 112, n. 5, 1260-1277, 2022.

IBGE. **Unidades de relevo** Rio de Janeiro, 2023. Disponível em:

<https://www.ibge.gov.br/geociencias/informacoes-ambientais/geomorfologia/10870-geomorfologia.html>. Acesso em: 14 de junho de 2024.

LANA, J. C.; CASTRO, P. D. T. A.; LANA, C. E. Assessing gully erosion susceptibility and its conditioning factors in southeastern Brazil using machine learning algorithms and bivariate statistical methods: A regional approach. **Geomorphology**, 402, 108159, 2022.

PELECH, A. S.; NUNES, B. T. D. A.; GATTO, L. C. S.; BOTELHO, R. G. M. Considerações sobre o mapeamento geomorfológico do território brasileiro: Algumas abordagens na representação regional. **Revista Brasileira de Geomorfologia**, 20, n. 3, 2019.

RWANGA, S. S.; NDAMBUKI, J. M. Accuracy assessment of land use/land cover classification using remote sensing and GIS. **International Journal of Geosciences**, Vol.08No.04, 12, 2017.

SILVEIRA, R. M. P.; SILVEIRA, C. T.; OKA-FIORI, C. Emprego de técnicas de inferência espacial para identificação de unidades de relevo apoiado em atributos topográficos e árvore de decisão. **Revista Brasileira de Geomorfologia**, 15, n. 1, 2014.

SOFIA, G. Combining geomorphometry, feature extraction techniques and earth-surface processes research: The way forward. **Geomorphology**, 355, 107055, 2020.

SOUZA, C. M. P.; THOMAZINI, A.; SCHAEFER, C. E. G. R.; VELOSO, G. V.; MOREIRA, G. M.; FERNANDES FILHO, E. I. Multivariate analysis and machine learning in properties of ultisols (argissolos) of Brazilian Amazon. **Revista Brasileira de Ciência do Solo**, 42, 1 - 20, 2018.

SREENIVASAN, G.; JHA, C. S. Geospatial technology for geomorphology mapping and its applications. In: JHA, C. S.; PANDEY, A., *et al* (Ed.). **Geospatial technologies for resources planning and management**. Cham: Springer International Publishing, 2022. p. 1-47.

VAN DER MEIJ, W. M.; MEIJLES, E. W.; MARCOS, D.; HARKEMA, T. T. L.; CANDEL, J. H. J.; MAAS, G. J. Comparing geomorphological maps made manually and by deep learning. **Earth Surface Processes and Landforms**, 47, n. 4, 1089-1107, 2022.

XU, H.; CHENG, W.; WANG, B.; SONG, K.; ZHANG, Y.; WANG, R.; BAO, A. Effects of geomorphic spatial differentiation on vegetation distribution based on remote sensing and geomorphic regionalization. **Remote Sensing**, 16, n. 6, 1062, 2024.

ZHANG, X.; CHEN, S.; XUE, J.; WANG, N.; XIAO, Y.; CHEN, Q.; HONG, Y.; ZHOU, Y.; TENG, H.; HU, B.; ZHUO, Z.; JI, W.; HUANG, Y.; GOU, Y.; RICHER-DE-FORGES, A. C.; ARROUAYS, D.; SHI, Z. Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping. **Geoderma**, 432, 116383, 2023.