

Identificação de Depósitos Aluviais em Regiões Semiáridas Utilizando Técnicas de Machine Learning

Daniel Baptista Vio¹, Gustavo Henrique Matos Bezerra Motta¹, Jonas Otaviano
Praça de Souza², Leandro Carlos de Souza¹

¹Centro de Informática – Universidade Federal da Paraíba (UFPB)
Campus I – Cidade Universitária – João Pessoa – PB – Brasil

²Centro de Ciências Exatas e da Natureza Universidade Federal da Paraíba (UFPB)
Campus I – Cidade Universitária – João Pessoa – PB – Brasil.

daniel.vio@academico.ufpb.br, gustavo@ci.ufpb.br,
jonas.souza@academico.ufpb.br, leandro@ci.ufpb.br

Abstract. *This study proposes a method for detecting alluvial areas in semi-arid regions using machine learning techniques. The research applied K-Nearest Neighbours, Decision Tree and Random Forest algorithms to high-resolution geospatial data, pre-processed with instance reduction. The Random Forest algorithm demonstrated the best performance, achieving an F1-score of 89.8%, precision of 91.0% and recall of 88.7%. The results confirm the effectiveness of this approach for the accurate mapping of these geological formations in semi-arid environments.*

Resumo. *Este estudo propõe um método para detecção de áreas aluvionares em regiões semiáridas utilizando técnicas de aprendizado de máquina. A pesquisa aplicou os algoritmos K-Vizinhos Mais Próximos, Árvore de Decisão e Floresta Aleatória a dados geoespaciais de alta resolução, pré-processados com redução de instâncias. A Floresta Aleatória apresentou o melhor desempenho, com F1-score de 89,8%, precisão de 91,0% e recall de 88,7%. Os resultados demonstram a eficácia da abordagem para o mapeamento preciso dessas formações geológicas em ambientes semiáridos.*

1. Introdução

A preocupação global com a preservação dos recursos hídricos tem se intensificado nas últimas décadas, diante das crescentes crises causadas pelas mudanças climáticas (Tayer et al., 2023). Estudos recentes, como o conduzido por Jasechko et al. (2024), analisaram milhares de medições em poços e aquíferos localizados em regiões áridas e semiáridas, evidenciando um declínio acelerado das reservas de água subterrânea em escala global. Esse cenário reforça a importância de estratégias para a gestão sustentável dos recursos hídricos, sobretudo em áreas particularmente vulneráveis às transformações climáticas.

No semiárido da Região Nordeste do Brasil a irregularidade dos regimes pluviométricos agrava os desafios relacionados à disponibilidade hídrica. O déficit de chuvas nesta região contribui para a predominância de rios efêmeros ou intermitentes, cujo fluxo ocorre apenas durante a estação chuvosa (McLeod et al., 2024). Além disso, eventos de seca recorrentes impactam o desenvolvimento socioeconômico, provocando o esgotamento de reservatórios superficiais, como lagos e açudes (Búrquez et al., 2024).

Nesse contexto, os aquíferos aluviais destacam-se como fontes estratégicas de água subterrânea, devido à sua elevada capacidade de infiltração e proteção contra a evaporação, oferecendo uma alternativa viável para mitigar os efeitos das secas prolongadas (Silva e Souza, 2023).

Os aquíferos aluviais são formações geológicas constituídas por sedimentos como cascalho, areia, silte e argila, que armazenam e permitem o fluxo de água subterrânea. Esses aquíferos são comumente encontrados em planícies fluviais e representam uma fonte importante de água para abastecimento. Já as áreas aluvionares referem-se às regiões superficiais onde ocorrem depósitos de sedimentos transportados e acumulados por rios ao longo do tempo. É importante ressaltar que nem todas as áreas aluvionares contêm aquíferos produtivos, o que torna sua identificação e caracterização relevantes (Freeze e Cherry, 1979). Embora este trabalho seja aplicado a regiões semiáridas as técnicas de identificação de aluviões não se limitam a esse contexto, assim, a metodologia aqui proposta pode ser adaptada a outros ambientes, desde que ajustados os parâmetros morfológicos e hidrológicos.

O mapeamento de áreas aluvionares são complexos devido à sua distribuição irregular e à ausência de uma forma geométrica definida (Cervi; Tazioli, 2021). Métodos tradicionais, baseados em trabalho de campo e análises manuais, são demorados, onerosos e muitas vezes insuficientes para cobrir grandes áreas, especialmente em regiões de difícil acesso. Além disso, a complexidade dos processos geomorfológicos envolvidos na formação e evolução desses depósitos dificulta a aplicação de abordagens convencionais.

A identificação de depósitos aluviais tem evoluído ao longo do tempo, desde métodos tradicionais baseados em trabalho de campo até abordagens modernas que utilizam inteligência artificial. Inicialmente, aluviões eram reconhecidos com base em descrições visuais e análises manuais de amostras de solo, um processo demorado e oneroso que exigia inspeções extensivas e levantamentos topográficos (Amit et al., 1996). Com o avanço da tecnologia, o sensoriamento remoto multiespectral emergiu como uma alternativa promissora, permitindo análises em larga escala com maior precisão e reduzindo a necessidade de trabalhos de campo exaustivos (Gillespie; Kahle; Palluconi, 1984).

A introdução do radar de abertura sintética (SAR) e do LiDAR (Light Detection and Ranging) transformou os modos de classificação e identificação de áreas de aluvião. Essas tecnologias permitiram a análise detalhada da topografia e da estrutura de aluviões, possibilitando a identificação de características geomorfológicas com uma precisão sem precedentes (Cavalli et al., 2008; Frankel e Dolan, 2007). Mais recentemente, o uso de técnicas de aprendizado de máquina potencializou a análise de grandes volumes de dados geoespaciais, ampliando a capacidade de detecção e classificação de áreas aluvionares (Muñoz-Carpena et al., 2023).

Embora a aplicação de aprendizado de máquina tenha sido extensivamente explorada em estudos sobre águas subterrâneas, a identificação e classificação específicas de formações aluvionares ainda carecem de investigação aprofundada. O levantamento bibliográfico revela uma escassez de trabalhos dedicados diretamente a essa temática. No entanto, algumas pesquisas relacionadas oferecem pontos de partida. Pipaud e Lehmkuhl (2017) apresentaram um método para delimitação e classificação de leques aluviais utilizando Modelos Digitais de Elevação (MDEs) combinados com clusterização mean-shift e Máquina de Vetor de Suporte (SVM), apresentando resultados promissores com a

Análise Morfométrica Baseada em Objetos (OBMA). Babič et al. (2021) também modelaram e classificaram leques aluviais com MDEs valendo-se de diversas técnicas de aprendizado de máquina.

Em uma análise de larga escala, Rabanaque et al. (2022) aplicaram SVM e Random Forest (RF) para segmentar e classificar canais fluviais e formas associadas em riachos efêmeros. A avaliação, demonstrou alta performance para ambos os algoritmos, com o SVM alcançando uma precisão média de 0,87 e o RF 0,85. Esse estudo, embora não focado exclusivamente nas formações aluvionares, mostra o potencial do aprendizado de máquina para a análise geomorfológica e hidrológica detalhada.

A crescente demanda por uma compreensão detalhada dos processos geomorfológicos, crucial para o aprimoramento de políticas públicas, ocorre em um cenário de grande disponibilidade de dados geoespaciais, impulsionado por avanços tecnológicos como imagens de satélite de alta resolução e LiDAR. Contudo, esse volume de dados revela uma lacuna significativa na literatura. Os métodos tradicionais de mapeamento de áreas aluvionares, baseados em interpretação visual e análises manuais, demonstram-se ineficazes e impraticáveis para o processamento e a extração de informações relevantes de conjuntos de dados tão extensos. Além disso, as abordagens computacionais existentes podem esbarrar em desafios de escalabilidade e eficiência computacional ao lidar com grandes volumes de dados, resultando em tempos de processamento elevados e, por vezes, na necessidade de infraestrutura computacional inacessível.

Diante dessas limitações, este trabalho propõe uma abordagem inovadora para a identificação de áreas aluvionares na bacia hidrográfica do Riacho Grande, no semiárido brasileiro, utilizando técnicas de aprendizado de máquina integradas a um método de redução de instâncias para o processamento eficiente de grandes volumes de dados. A inovação principal reside na aplicação estratégica dessa combinação, visando superar as limitações de escalabilidade e eficiência dos métodos existentes e oferecer uma solução mais prática e eficaz. Este artigo está organizado da seguinte forma: a seção 2 detalha os Materiais e Métodos, apresentando a área de estudo e os dados utilizados; a seção 3 apresenta os Resultados e Discussão, com análise do desempenho dos algoritmos K-vizinho mais próximos (KNN), Árvore de Decisão (DT) e RF em diferentes níveis de redução de dados. A Seção 4 traz as Conclusões, destacando as contribuições do estudo, suas limitações e perspectivas para pesquisas futuras.

2. Materiais e Métodos

Esta seção descreve os materiais e métodos utilizados no estudo. O objetivo é apresentar de forma clara e concisa as etapas metodológicas adotadas para a identificação de áreas aluvionares na bacia hidrográfica do Riacho Grande, localizada no semiárido brasileiro.

2.1. Área Objeto da Pesquisa

A área de estudo é a bacia hidrográfica do Riacho Grande, localizada nas cidades de Serra Talhada, Calumbi, Flores e Betânia, no sertão de Pernambuco. A bacia possui uma área de aproximadamente 315 km² e está situada no Planalto da Borborema, caracterizado pela predominância de rochas cristalinas, o que influencia o relevo da região. Os rios da bacia são efêmeros e intermitentes, com fluxo ocorrendo principalmente durante a estação chuvosa. A altitude da bacia varia entre 430 e 870 metros, com as maiores declividades localizadas na área de cabeceira. O clima na região é tropical semiárido, com temperaturas

médias variando entre 25°C e 30°C e precipitação média anual entre 450 e 700 mm, concentrada principalmente entre os meses de novembro e abril. A bacia é uma sub-bacia do Rio Pajeú e apresenta características geomorfológicas típicas de regiões semiáridas, com áreas de produção de sedimentos nas regiões mais altas e áreas de deposição nas regiões mais baixas, onde se formam os depósitos aluvionares.

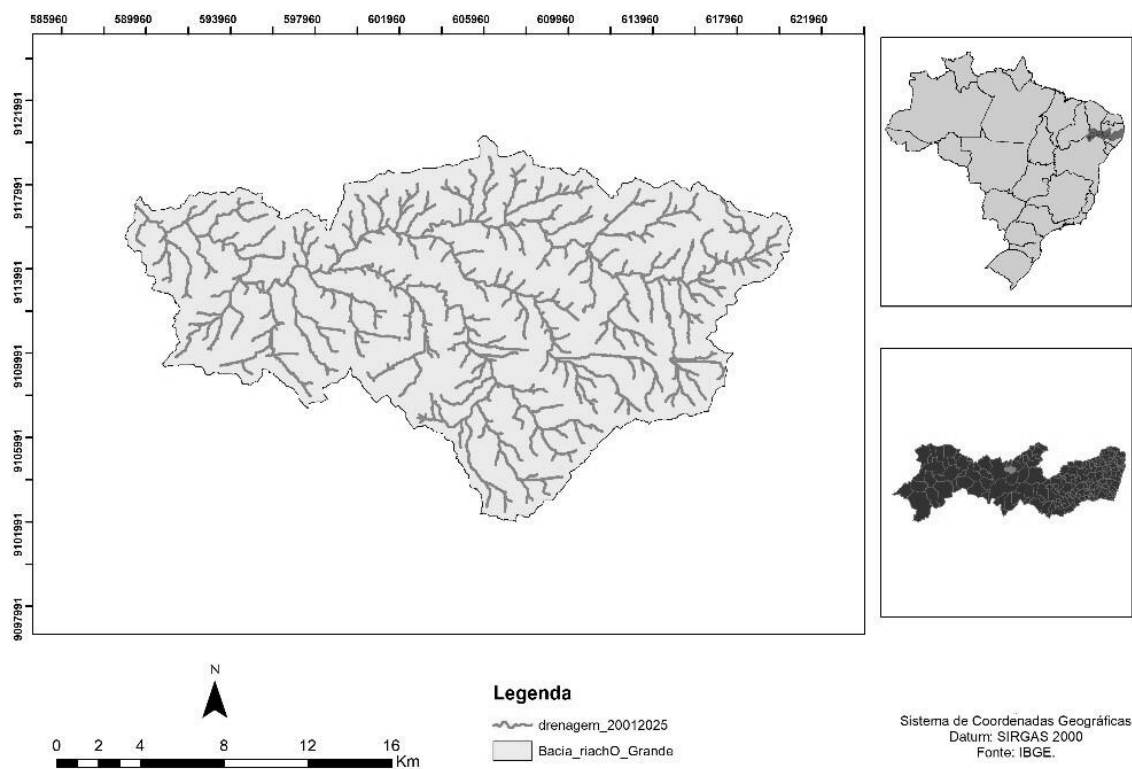


Figura 1. Localização e Drenagem da Bacia do Riacho Grande/PE

2.2. Obtenção dos Dados

Os dados utilizados neste estudo foram obtidos a partir de Modelos Digitais de Elevação (MDEs) gerados pelo Programa Pernambuco Tridimensional (PE3D), que utilizou tecnologias de aerofotogrametria e LiDAR (Light Detection and Ranging) para mapeamento do estado de Pernambuco. Os MDEs possuem uma resolução espacial de 1 metro quadrado (1x1m), o que representa uma precisão significativamente superior aos modelos gratuitos disponíveis globalmente, como o Shuttle Radar Topography Mission (SRTM) que possui resolução de 900 metros quadrados (30x30m). Essa alta resolução permitiu a extração detalhada de variáveis topográficas e hidrológicas, aumentando a precisão da análise das áreas aluvionares.

Para a caracterização geomorfológica e hidrológica das áreas aluvionares, foram coletadas onze variáveis relevantes para a formação e identificação desses depósitos. Estas incluíram atributos topográficos como altitude (metros), área de acumulação (quilômetros quadrados), conectividade, curvatura total, e declividade (graus), além de medidas de distância a elementos fluviais (distância até o centro do rio em metros, distância até a cabeceira em quilômetros) e índices hidrológicos (SPI, TWI). A precipitação, crucial para a dinâmica hídrica da bacia, foi estimada como a média dos últimos 30 anos a partir de

dados de cinco estações pluviométricas da Agência Pernambucana de Águas e Clima (APAC): Betânia, Calumbi, Flores, Serra Talhada e Custódia (milímetros). A variável de saída binária (Aluvião: 1 para presença, 0 para ausência) foi definida a partir de trabalhos de campo e análise detalhada de imagens do Google Earth, garantindo a representatividade das áreas mapeadas para a modelagem.

A obtenção das variáveis foi realizada utilizando o software ArcMap, versão 10.8, uma ferramenta de Sistema de Informação Geográfica (SIG). Através dessa ferramenta, foram realizadas análises espaciais e geográficas fundamentais para a coleta das variáveis. O processo de trabalho para obtenção dos dados está retratado na Figura 2, que apresenta as etapas desde a aquisição dos dados brutos até a geração do conjunto final de informações utilizadas para a modelagem.

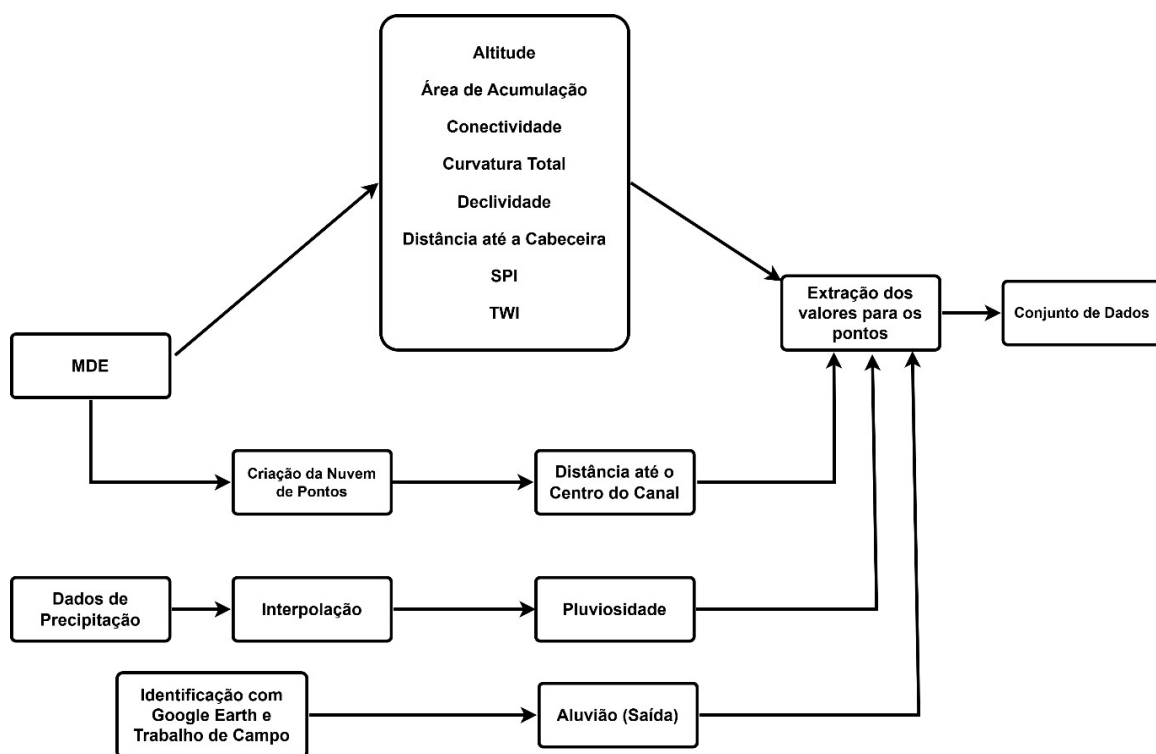


Figura 2. Etapas de trabalho para obtenção dos dados

O conjunto de dados para a modelagem apresentava um desbalanceamento significativo, com aproximadamente 92,5% de instâncias correspondendo à classe não aluvionar (0) e apenas 7,5% representando áreas aluvionares (1). Para otimizar o desempenho dos algoritmos de aprendizado de máquina, realizou-se uma busca de hiperparâmetros (grid search) para cada modelo. Os hiperparâmetros ótimos identificados foram: para o KNN, $n_neighbors=7$, $weights='distance'$ e $metric='manhattan'$; para a Árvore de Decisão, $max_depth=20$, $min_samples_split=5$, $min_samples_leaf=2$ e $class_weight=6$; e para o Random Forest, $max_depth=25$, $min_samples_split=2$, $min_samples_leaf=2$, $class_weight=5$ e $n_estimators=50$. Essa otimização foi aplicada em todos os bancos resultantes da etapa de redução descrita a seguir.

2.3. Redução de Instâncias

Devido ao grande volume de dados, foi necessário aplicar um método de redução de instâncias para viabilizar o processamento e a aplicação de técnicas de aprendizado de máquina. O método escolhido foi o Difusão Geométrica Markoviana estendido (DGM-e), uma abordagem validada na literatura para a redução de grandes conjuntos de dados, preservando a representatividade das informações originais (Silva, 2012; Silva et al., 2016). O DGM-e foi aplicado para obtenção de diferentes níveis de redução, permitindo a criação de bancos de dados menores, mas ainda representativos.

O processamento foi realizado utilizando um computador com processador Intel Core i7 de 13ª geração, 64 GB de memória RAM e uma placa gráfica dedicada com 8 GB de memória. A redução foi essencial para a realização do treinamento e dos testes para a aplicação das técnicas de aprendizado de máquina. Os patamares de redução alcançados (90%, 95%, 99% e 99,9%) foram os máximos possíveis dentro dos recursos computacionais disponíveis, garantindo que o processamento pudesse ser realizado de forma eficiente.

O método apresentado buscou, dentro desses recursos, otimizar o tempo necessário para a conclusão de todas as etapas de redução. O conjunto de dados original, composto por mais de 315 milhões de pontos, foi dividido em 16 partes para viabilizar o processamento. Cada parte foi reduzida individualmente, seguindo uma abordagem em três etapas: (1) redução inicial de cada parte do banco de dados, (2) reunião das partes reduzidas e nova redução, e (3) unificação final e redução do conjunto completo.

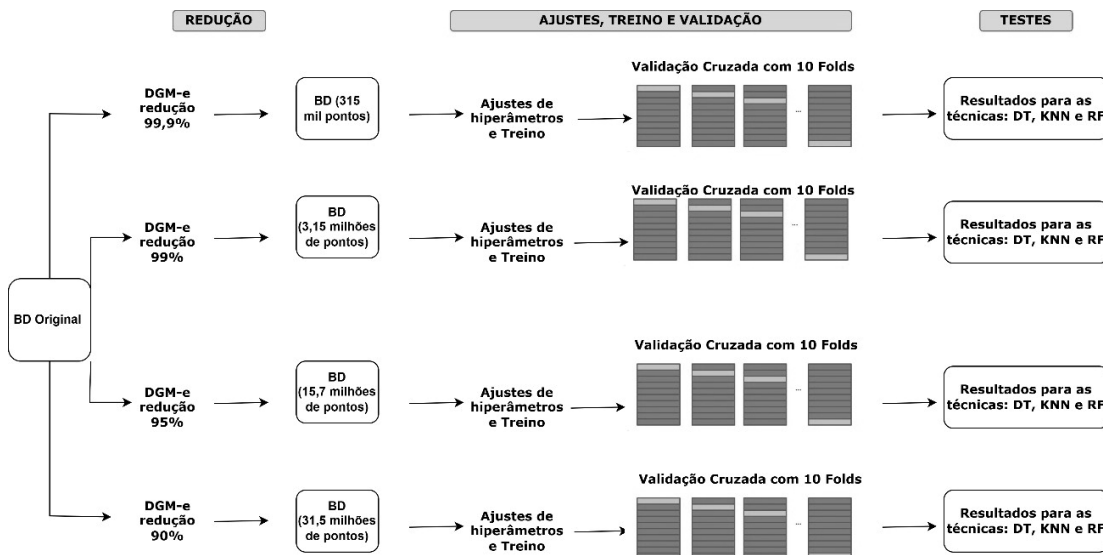


Figura 3. Etapas de trabalho para aplicação das técnicas de aprendizagem de máquina

O processo de trabalho para aplicação das técnicas de aprendizagem de máquina está ilustrado na Figura 3, que apresenta as etapas desde a redução dos dados até a validação dos modelos. Após a redução, os dados foram submetidos a técnicas de aprendizado de máquina, incluindo DT, KNN e RF, para a identificação de áreas aluvionares.

A validação dos modelos foi feita por meio de validação cruzada com 10 folds, garantindo a consistência dos resultados. A escolha do método DGM-e e a aplicação de diferentes

níveis de redução permitiram a análise eficiente dos dados, mantendo a qualidade preditiva dos modelos. Essa abordagem possibilitou a identificação de áreas aluvionares com alta precisão, mesmo em um cenário de grande volume de dados.

3. Resultados e Discussão

A Tabela 1 apresenta os resultados obtidos com a aplicação de três técnicas de classificação (KNN, DT e RF) nos quatro bancos gerados após o processo de redução dos dados. Não foi possível aplicar as técnicas de aprendizado de máquina ao conjunto total de dados, que continha aproximadamente 315 milhões de pontos, devido ao alto custo computacional envolvido. Dessa forma, a redução dos dados tornou-se essencial para viabilizar a modelagem, permitindo que os experimentos fossem conduzidos dentro do limite da capacidade dos recursos disponíveis. Após alguns testes, o percentual de 10% do total de instâncias representou o maior volume de dados que foi possível processar sem comprometer a execução dos modelos, garantindo uma análise consistente sem exceder os limites operacionais.

Tabela 1. Resultados obtidos por cada técnica de aprendizado de máquina nos diferentes níveis de redução.

Técnica Aplicada	Total de Instâncias	F1-score	Precisão	Recall
KNN	0,1%	72,7%	75,1%	70,4%
KNN	1%	66,3%	69,5%	63,5%
KNN	5%	86,1%	87,3%	84,9%
KNN	10%	85,5%	86,7%	84,3%
DT	0,1%	70,3%	73,2%	67,9%
DT	1%	74,4%	76,1%	72,9%
DT	5%	86,3%	88,2%	84,5%
DT	10%	85,3%	87,1%	83,7%
RF	0,1%	75%	78,5%	71,7%
RF	1%	77,4%	80,2%	74,8%
RF	5%	89,8%	91,0%	88,7%
RF	10%	89,0%	90,2%	87,9%

Os resultados obtidos representam quatro diferentes níveis de redução do conjunto original, com proporções de 0,1%, 1%, 5% e 10% do total, permitindo avaliar o impacto da redução na performance dos classificadores. A acurácia não foi utilizada como critério de avaliação, uma vez que o extremo desbalanceamento dos dados tornaria essa métrica pouco representativa para a qualidade real dos modelos

O KNN mostrou-se sensível à quantidade de dados disponíveis, com desempenho inferior nos bancos mais reduzidos (0,1% e 1%), mas melhorando significativamente com 5% e 10% dos dados. Esse comportamento sugere que o KNN é mais dependente do volume de dados para alcançar resultados consistentes. Já o DT apresentou uma melhoria progressiva à medida que o volume de dados aumentou, embora não tenha superado o desempenho do RF em nenhum dos níveis de redução.

O RF destacou-se pela sua capacidade de manter métricas elevadas mesmo com apenas 5% dos dados originais, alcançando um F1-score de 89,8%. Esse resultado reforça a eficácia do RF em problemas de classificação complexos, como a identificação de áreas aluvionares. A validação cruzada com 10 folds confirmou a pertinência do modelo, com métricas consistentes e baixa variabilidade entre os folds. O modelo treinado com 5% dos dados obteve um F1-score de 0,8975, com intervalo de confiança de (0,8972, 0,8978), enquanto o modelo treinado com 10% dos dados apresentou um F1-Score de 0,8897, com intervalo de (0,8895, 0,8899). A ausência de sobreposição entre os intervalos indica que o modelo treinado com 5% dos dados manteve um desempenho levemente superior. Adicionalmente, a análise da curva Precision-Recall para este modelo revelou um bom desempenho na identificação das áreas de aluvião, com uma Average Precision (AP) de 0.97.

Ao comparar os resultados obtidos neste estudo com as métricas apresentadas na literatura, observa-se que as abordagens utilizadas, como RF, alcançam níveis de precisão e F1-score comparáveis aos relatados em estudos similares, como o de Rabanaque et al. (2022) que obteve uma acurácia média de 85% e um coeficiente Kappa de 0,81 na classificação de elementos fluviais, incluindo áreas aluvionares.

Os resultados do presente estudo, onde o RF atingiu um F1-score de 89,8%, além de precisão e recall próximos desse patamar, indicam desempenhos similares, considerando as diferenças metodológicas e os contextos de aplicação. Isso reforça a aplicabilidade do aprendizado de máquina na identificação de áreas aluvionares, indicando que modelos como o RF podem obter resultados consistentes mesmo com diferentes bases de dados e metodologias.

A Figura 4 ilustra a importância relativa das variáveis utilizadas, evidenciando a hierarquia dos fatores mais influentes na classificação, bem como a menor contribuição de atributos geomorfométricos como declividade e curvatura.

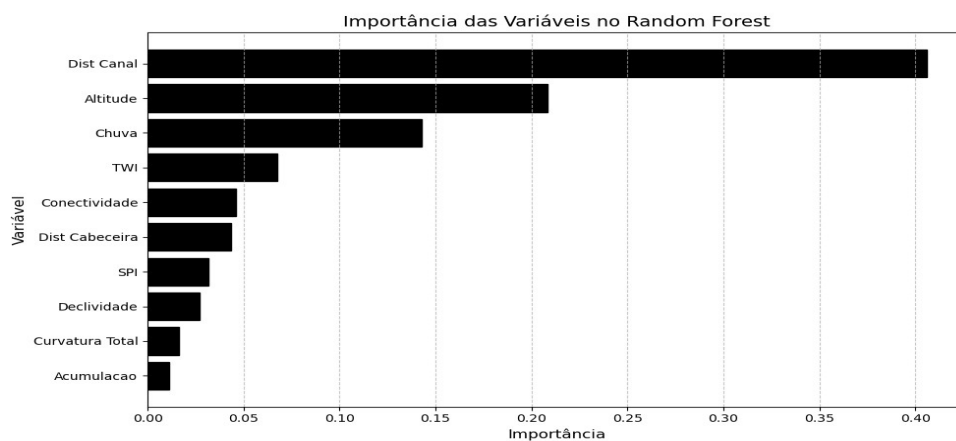


Figura 4. Importância das variáveis utilizadas na predição

Com base nos resultados obtidos utilizando o RF no banco com 5% do total, podemos destacar que as variáveis de maior importância no modelo foram a distância ao canal, a altitude e a precipitação. Esses fatores apresentaram os maiores pesos na classificação das áreas aluvionares. Estudos como os de Pipaud e Lehmkuhl (2017) e Babič et al. (2021) destacam a relevância de variáveis geomorfológicas, como declividade e curvatura, que, embora tenham sido consideradas neste estudo, apresentaram menor impacto na modelagem, com valores de importância inferiores aos das variáveis de maior relevância.

Essa divergência pode estar relacionada às diferenças na resolução espacial dos dados utilizados: enquanto este trabalho empregou MDEs de alta resolução (1x1 m), os estudos citados basearam-se em dados com resoluções mais baixas (ex.: 30x30 m no SRTM). Dados de alta resolução capturam com maior precisão características locais como proximidade a canais e microtopografia, potencialmente elevando a importância de algumas variáveis em detrimento de parâmetros geomorfológicos globais. Essa hipótese sugere que a escala de análise e a resolução dos dados podem influenciar significativamente a hierarquia de variáveis preditivas na modelagem de áreas de aluvião. Futuros estudos podem investigar sistematicamente como diferentes resoluções espaciais afetam a importância relativa das variáveis e, conseqüentemente, a capacidade de discriminar a presença do fenômeno.

4. Conclusões

Este estudo demonstrou a eficácia da combinação da técnica de redução de instâncias DGM-e com o algoritmo Random Forest (RF) na identificação precisa de áreas aluvionares em dados de alta resolução (1x1 m). O desempenho do RF, alcançando um F1-score de 89,8% com apenas 5% dos dados originais, validou a abordagem para lidar eficientemente com grandes volumes geoespaciais. Embora algumas limitações tenham impedido comparações diretas com outros métodos, como o SVM, pesquisas futuras poderão explorar sua aplicação aos dados reduzidos pelo DGM-e, assim como investigar o impacto de diferentes resoluções espaciais no desempenho dos algoritmos.

Do ponto de vista prático, a metodologia desenvolvida oferece uma ferramenta valiosa para o mapeamento de aquíferos aluviais em regiões semiáridas, onde a identificação precisa dessas formações é essencial para a gestão sustentável dos recursos hídricos. Futuros desenvolvimentos, incluindo a incorporação de outros algoritmos e a aplicação em diferentes contextos geográficos, poderão ampliar ainda mais o potencial desta linha de pesquisa.

Referências

- Amit, R., Zilberman, E., and Porat, N. (1996). "Paleoseismic evidence for time dependency of seismic response on a fault scarp in the southern Arava Valley, Israel". *Journal of Geophysical Research: Solid Earth*, 101(B12), 27517–27532.
- Babič, M., Repe, B., and Pavlovec, R. (2021). "Machine learning approaches for alluvial fan classification using digital elevation models". *Geomorphology*, 372, 107451.
- Búrquez, A., Martínez-Yrizar, A., and Felger, R. S. (2024). "Drought impacts on water resources in semi-arid regions: A case study of the Sonoran Desert". *Journal of Arid Environments*, 110, 104-112.

- Cavalli, M., Tarolli, P., and Dalla Fontana, G. (2008). "The effectiveness of LiDAR-derived topographic attributes for identifying alluvial fans". *Geomorphology*, 102(3-4), 555-566.
- Cervi, F., and Tazioli, A. (2021). "Identification of alluvial deposits in semi-arid regions using remote sensing and GIS techniques". *Remote Sensing*, 13(5), 987.
- Frankel, K. L., and Dolan, J. F. (2007). "Characterizing arid region alluvial fan surface roughness with airborne laser swath mapping digital topographic data". *Journal of Geophysical Research: Earth Surface*, 112(F2).
- Freeze, R. A., and Cherry, J. A. (1979). *Groundwater*. Prentice-Hall.
- Gillespie, A., Kahle, A., and Palluconi, F. (1984). "Mapping alluvial fans in Death Valley, California, using multiband thermal infrared images". *Geophysical Research Letters*, 11(11), 1153-1156.
- Jasechko, S., Perrone, D., and Befus, K. M. (2024). "Global groundwater decline: A synthesis of recent well and aquifer measurements". *Nature Geoscience*, 17(2), 123-130.
- McLeod, T., Huntington, J., and Morton, C. (2024). "Hydrological impacts of climate change on ephemeral rivers in semi-arid regions". *Hydrological Processes*, 38(1), 45-60.
- Muñoz-Carpena, R., Fox, G. A., and Sabbagh, G. J. (2023). "Machine learning applications in hydrology and water resources management". *Water Resources Research*, 59(4), e2022WR033456.
- Pipaud, I., and Lehmkuhl, F. (2017). "Object-based morphometric analysis of alluvial fans using digital elevation models and machine learning". *Geomorphology*, 295, 1-14.
- Rabanaque, M. I., Toledo, M., and Colombo, F. (2022). "Large-scale hydromorphological analysis of ephemeral streams using machine learning". *Journal of Hydrology*, 612, 128-145.
- Silva, A. M., and Souza, R. F. (2023). "Alluvial aquifers as strategic water resources in semi-arid regions: A review of their characteristics and management". *Journal of Hydrology*, 605, 127-140.
- Silva, D. A. N. (2012). *Um Método Geral Para Redução de Grandes Conjuntos de Dados Baseados em Difusão Geométrica Markoviana*. Dissertação de Mestrado, Universidade Federal da Paraíba, João Pessoa.
- Silva, D. A. N., Souza, L. C., and Motta, G. H. M. B. (2016). "An instance selection method for large datasets based on Markov geometric diffusion". *Data & Knowledge Engineering*, 101, 24-41.
- Tayer, J., Smith, R., and Johnson, L. (2023). "Climate change impacts on water resources: A global perspective". *Nature Climate Change*, 13(4), 321-330.