



Aplicabilidade da inteligência artificial no aprendizado de máquina para o planejamento territorial

Applicability of artificial intelligence in machine learning for territorial planning

Tiago Henrique da Silva^a, Isaac Pereira Teles^a, Bruno José Oliveira Machado de Araújo^b

^a Universidade Federal Rural de Pernambuco-UFRPE, Departamento de Tecnologia Rural, Centro de Inovações Tecnológicas Aplicadas aos Recursos Naturais-CIPAR. Rua Dom Manuel de Medeiros, s/n, Dois Irmãos, Recife, Pernambuco, Brasil. CEP:52171-900. E-mail: tiagoh99@gmail.com, isaacteles2016@gmail.com.

^b Universidade Católica de Pernambuco-UNICAP, Laboratório de Solos e Tecnologias Sustentáveis. Rua do Príncipe, 526, Boa Vista, Recife, Pernambuco, Brasil. CEP: 50050-900. E-mail: brunomachado15999@gmail.com.

ARTICLE INFO

Recebido 14 Fev 2025

Aceito 11 Jul 2025

Publicado 01 Ago 2025

ABSTRACT

In its broadest sense, remote sensing has been employed in research covering virtually all branches of science and technology, including environmental engineering. It plays a crucial role in understanding and monitoring environmental changes, enabling data collection without direct human intervention. This study aimed to classify land use and land cover in the Toritama region, located in the Agreste of Pernambuco, to assess the effectiveness of the three previously mentioned classifiers in obtaining the best Kappa accuracy index and overall accuracy. CBERS 4A satellite images were utilized, combining multispectral and panchromatic data for classification purposes. Three machine learning methods were employed for classification: *Random Forest*, *Shark Random Forests*, and *Decision Tree*. An object-based image analysis (OBIA) strategy was applied, which involved image fusion, attribute segmentation, attribute extraction, segment sampling, segment classification, and accuracy analysis. The results indicated that the *Shark Random Forest* method performed best in delineating urban areas, while the *Random Forest* method excelled in classifying water areas. The *Decision Tree* method showed inferior results in both cases. Considering the metrics, including overall accuracy and the Kappa index, the *Random Forest* classifier was slightly superior to the *Shark Random Forest* and 5.4 percentage points above the *Decision Tree* method. The qualitative analysis demonstrated excellent agreement with the delimited classes, indicating that the methods have high overall accuracy and an excellent Kappa index.

Keywords: Remote sensing, land use and land cover, algorithms.

RESUMO

O sensoriamento remoto, no seu sentido mais extenso, vem sendo empregado em pesquisas abrangendo, praticamente, todos os ramos da ciência e tecnologia, dentre os quais a engenharia ambiental. Executando um papel importante na compreensão e monitoramento das mudanças ambientais, permitindo assim, a coleta de dados sem mediação direta. O estudo objetivou classificar o uso e cobertura do solo na região de Toritama, agreste pernambucano, visando avaliar a eficácia dos três classificadores citados anteriormente, de forma a obter o melhor índice de exatidão Kappa e acurácia global. Utilizando imagens do satélite CBERS 4A, conciliando dados multiespectrais e pancromáticos, para classificar. Três métodos de aprendizado de máquina foram utilizados para classificação: *Random Forest*, *Shark Random Forests* e *Decision Tree*. Foi aplicada a classificação baseada em objetos (OBIA), que envolveu a fusão de imagens, a segmentação de atributos, a extração de atributos, a amostragem de segmentos, a classificação dos segmentos e a análise de acurácia. Os resultados indicaram que o método *Shark Random Forest* mostrou desempenho superior na delimitação de área urbana. O método *Random Forest* se destacou na



Journal of Environmental Analysis and Progress © 2016
is licensed under CC BY-NC-SA 4.0

classificação das áreas de água. Considerando o método *Decision Tree*, os resultados foram inferiores em ambos os sentidos. Considerando as métricas, a acurácia global e o índice Kappa, o classificador *Random Forest* foi levemente superior ao *Shark Random Forest* e 5,4 pontos percentuais acima do método *Decision Tree*. A análise qualitativa, portanto, evidenciou ótima conformidade com as classes delimitadas, indicando que os métodos possuem alta acurácia global e excelente índice Kappa.

Palavras-Chave: Sensoriamento remoto, uso e cobertura do solo, algoritmos.

Introdução

A intensa ocupação humana dos espaços, associada às diferentes formas de uso e ocupação do solo, gerou uma variedade de impactos ambientais e sociais. Variadas ferramentas de análise do uso e cobertura do solo foram criadas, utilizando sensoriamento remoto, Sistemas de Informação Geográfica (SIG) e o aprendizado de máquina, para uma melhor aplicabilidade e estratégia no apoio ao planejamento territorial (Correa et al., 2024).

O sensoriamento remoto (SR) tem se tornado cada vez mais popular na ciência, principalmente por ser uma ferramenta de baixo custo e por permitir análises precisas no espaço e no tempo. A aplicação de suas técnicas permite que o SR viabilize o planejamento ambiental e urbano, incluindo o uso de índices de vegetação, que ajudam a analisar o comportamento espectral e a diferenciar as coberturas e usos da terra. Isso otimiza uma detecção quantitativa e qualitativa dos alvos observados (Guedes & Silva, 2018).

Com o desenvolvimento da ferramenta computacional em sistema de Informações Geográficas (SIG), popularmente conhecido como *Geographic Information System* (GIS), concentrada em assimilar, agrupar, examinar e exibir dados geoespaciais em informações que descrevem a localização e as características de elementos espaciais (Xia, Efremochkina & Lin, 2022). Nesse sentido, o aprendizado de máquina, em apoio ao SR e ao SIG, é subárea da inteligência artificial (IA), que executa algoritmos capazes de se autoaperfeiçoar com base em dados empíricos.

A aplicabilidade do aprendizado em máquina e classificadores vem impulsionando o desenvolvimento de sistemas computacionais capazes de analisar grandes quantidades de dados rapidamente e gerar relatórios para auxiliar na tomada de decisões (Randhawa et al., 2018). Esse desenvolvimento de metodologias que auxiliem na automação do processo de mapeamento, tanto da supressão da vegetação nativa, quanto das mudanças de uso e cobertura da terra, torna-se essencial entre os classificadores baseados em aprendizado de máquina. As ferramentas *Random Forest*, *Shark Random Forest* (algoritmos) (Amulu, Manjusha & Ashok, 2024; Goez, Gameiro & Miziara, 2024) e a *Decision Tree* (fluxograma

classificador) são amplamente utilizadas com diferentes aplicações (Veloso et al., 2021).

O aprendizado de máquina (*machine learning*) refere-se a um campo da inteligência artificial que desenvolve algoritmos capazes de aprender padrões a partir de dados e fazer previsões ou decisões com base em novas informações. No contexto do sensoriamento remoto, esses algoritmos têm sido amplamente empregados para a classificação de imagens, especialmente dentro de abordagens como a Object-Based Image Analysis (OBIA), onde o objetivo é identificar classes temáticas (ex: floresta, urbano, água) com base em atributos espetrais, texturais, espaciais e contextuais (Lu & Weng, 2007; Belgiu & Drăguț, 2016).

O processo de aprendizado de máquina envolve, geralmente, duas fases principais: treinamento e validação/teste. Na etapa de treinamento, o algoritmo aprende a associar um conjunto de atributos a uma classe de saída, com base em exemplos previamente rotulados. Posteriormente, na fase de validação ou teste, avalia-se a capacidade do modelo de generalizar esse conhecimento para novos dados (Mitchell, 1997).

O *Random Forest* é um algoritmo que tem se destacado por sua ampla adoção em tomadas de decisão, que visam maior precisão na sua aplicação (Santana-Santos et al., 2022; Rhodes, Cutler & Moon, 2023; Della-Silva et al., 2024). A partir do seu uso, é possível obter classificações confiáveis, com previsões geradas a partir de um conjunto de árvores de decisão, produzindo resultados estatísticos excelentes e com processamento rápido (Gustavo et al., 2024).

Adicionalmente, o *Shark Random Forest* é uma biblioteca modular em C++, desenvolvida para o *design* e a otimização de sistemas adaptativos, versátil e útil, tanto para aplicações práticas, quanto para pesquisas em inteligência computacional e aprendizado de máquina. Essa biblioteca oferece métodos baseados em gradiente, aprendizado por *kernel*, redes neurais e outras técnicas de aprendizado de máquina. Seus principais objetivos de design envolvem proporcionar flexibilidade e alto desempenho (Igel et al., 2008).

O classificador *Decision Tree* prevê o valor de uma variável-alvo, a partir de regras de decisão

simples, extraídas das variáveis de entrada. Ele constrói um modelo em forma de árvore para representar decisões e suas possíveis consequências (Priyanka, 2020).

A avaliação comparativa da eficiência de classificadores como *Random Forest*, *Shark Random Forest* e *Decision Tree* na classificação do uso e ocupação do solo é motivada por diversos desafios técnicos e científicos. A heterogeneidade espectral e espacial dos dados oriundos de sensores remotos dificulta a discriminação precisa entre classes distintas de uso do solo, demandando algoritmos capazes de lidar com essa variabilidade (Belgiu & Drăguț, 2016). Embora os métodos baseados em árvores de decisão, como o *Decision Tree*, sejam intuitivos e de fácil interpretação, eles tendem a sofrer com sobreajuste e desempenho inferior em cenários com alta complexidade de dados (Lary et al., 2016).

Em contrapartida, classificadores como o *Random Forest* têm se destacado por sua robustez, resistência ao *overfitting* (sobreajuste) e alto desempenho em tarefas de classificação geoespacial (Rodriguez-Galiano et al., 2012). Apesar disso, novas variantes, como o *Shark Random Forest* - que incorpora melhorias em eficiência computacional e capacidade de generalização - ainda são pouco exploradas na literatura científica, o que evidencia uma lacuna importante a ser preenchida. Além disso, a maioria dos estudos existentes concentra-se em regiões temperadas e ecossistemas amplamente estudados, havendo escassez de análises aplicadas a áreas tropicais ou urbanas em rápida transformação, como as encontradas no Brasil (Souza et al., 2020).

Nesse contexto, uma análise comparativa entre esses algoritmos permitirá identificar os métodos mais adequados conforme o tipo de cobertura do solo e características dos dados, contribuindo para a padronização de boas práticas metodológicas em estudos de mapeamento geoespacial. Tal abordagem pode oferecer subsídios técnicos para aplicações práticas em planejamento urbano, gestão ambiental e políticas públicas territoriais (Maxwell et al., 2018).

Assim, o estudo visa suprir lacunas metodológicas e tecnológicas, fornecendo diretrizes atualizadas sobre a escolha de classificadores em projetos de classificação do uso e ocupação do solo, de forma a obter o melhor índice de exatidão Kappa e acurácia global.

O índice Kappa e a acurácia global, ambos obtidos a partir da matriz de confusão, são amplamente utilizados na avaliação da qualidade de classificações temáticas, como mapas de uso e ocupação do solo. No entanto, é importante destacar que essas métricas não são classificadores estatísticos, mas sim medidas de exatidão utilizadas para quantificar o desempenho de um classificador. Em outras palavras, enquanto os classificadores (Máquinas de Vetores de Suporte, *Random Forest*, *Shark Random Forest*, Árvores de Decisão ou Redes Neurais) são responsáveis por atribuir classes aos dados, a acurácia global e o índice Kappa são utilizados para verificar a exatidão do resultado gerado pelos classificadores (Congalton & Green, 2009).

A acurácia global expressa a proporção total de amostras corretamente classificadas em relação ao total de amostras avaliadas. O índice Kappa é uma métrica estatística que mede a concordância entre os dados classificados e os dados de referência, ajustando os resultados para considerar os acertos esperados ao acaso (Landis & Koch, 1977).

Portanto, a utilização conjunta dessas métricas permite uma avaliação mais robusta da qualidade dos mapas temáticos, o que é fundamental para aplicações que envolvem planejamento territorial, gestão ambiental e políticas públicas baseadas em dados espaciais.

Material e Métodos

Área de estudo

O estudo foi desenvolvido no município de Toritama, localizado a aproximadamente 170 km da capital de Pernambuco, Recife, em uma altitude de 349 m, nas coordenadas 8°0'24" de latitude sul e 36°3'24" de longitude oeste (Figura 1). O relevo é complexo, marcado por vales profundos e estreitos. Nas áreas mais altas e nas cristas residuais predominam solos litólicos. Nos topo e encostas das ondulações encontram-se solos brunos não cárnicos, pouco profundos e com fertilidade variando entre média e alta (Beltrão et al., 2005). Toritama está localizada na Província Borborema; sua formação geológica característica apresenta um solo composto principalmente por rochas das suítes Serra de Taquaritinga e Shoshonítica (Lei N° 1.452/2015, 2015).

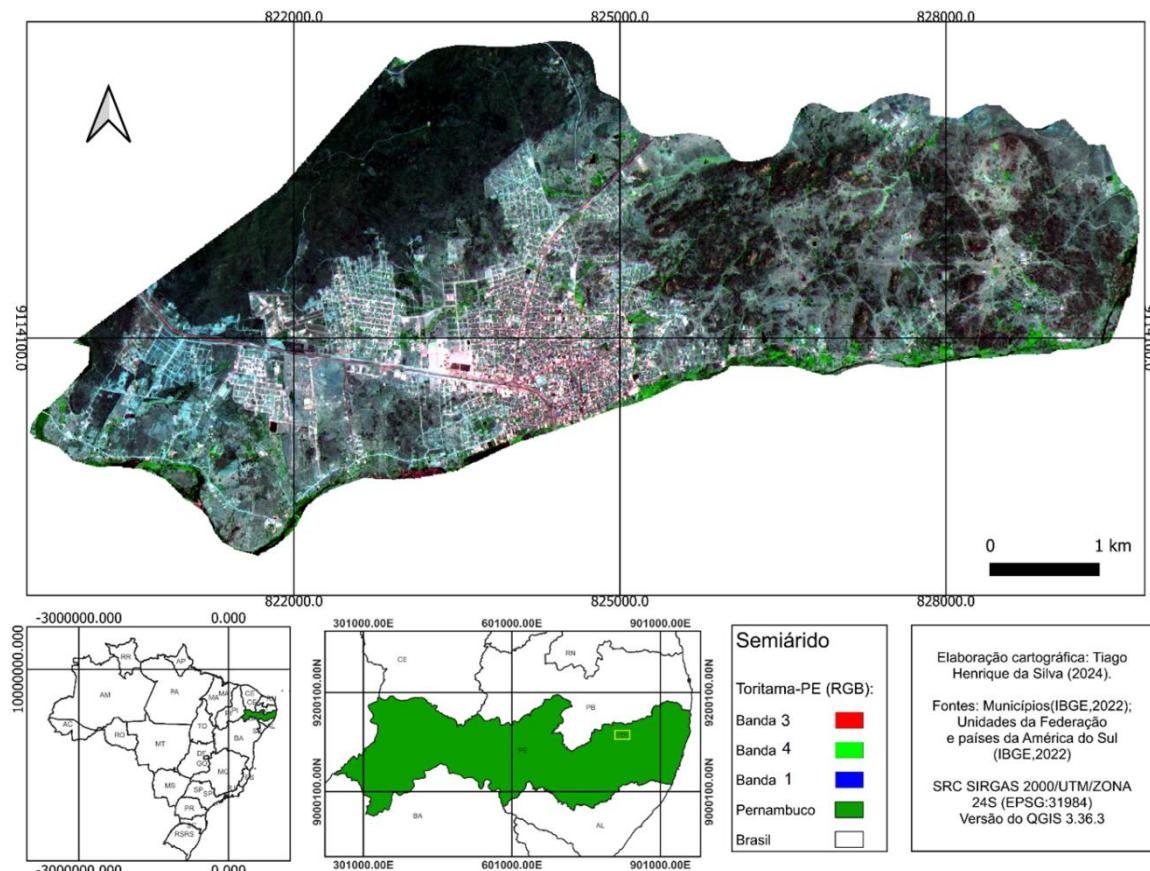


Figura 1. Mapa da área de estudo (imagem CBERS 4A WPM composição colorida R3G4B1), no município de Toritama, Pernambuco. Banda 1 = Utilizada para estudos de corpos d'água, identificação de vegetação saudável e diferenciação entre solos e áreas urbanas; essencial para análises ambientais e monitoramento de recursos hídricos; Banda 3 = Sensível à clorofila, é ideal para análise da vegetação; usada na classificação de uso do solo e na identificação de áreas agrícolas e estudos ambientais.; Banda 4 = Essencial para a avaliação da vegetação, diferenciando plantas saudáveis e estressadas; permitindo distinguir entre diferentes tipos de cobertura do solo, incluindo água, vegetação e áreas urbanas. Fonte: Silva, Machado & Teles (2025).

De acordo com a Agência Pernambucana de Águas e Climas (APAC, 2018), o clima é semiárido, quente, com chuvas no outono e inverno, apresentando os índices pluviométricos mais baixos da região do Agreste, com precipitação variando de 17 mm a 81 mm, com média anual de 38,6 mm. A vegetação predominante é uma caatinga hipoxerófica arbustiva (Beltrão et al., 2005). Toritama possui uma população de 47.088 hab., distribuída em uma área de 25,7 km², com uma densidade demográfica de 1.759,2 hab. km⁻² (IBGE, 2019).

Atualmente, Toritama, ao lado de Caruaru e Santa Cruz do Capibaribe, é uma das principais cidades que compõem o Polo de Confecções do Agreste. Esse conglomerado, considerado o maior do Nordeste e o segundo maior do Brasil, movimenta bilhões de reais e gera emprego para 130 mil pessoas, conforme dados do Estudo Econômico do Arranjo Produtivo Local (Sebrae, 2019). O polo têxtil do Agreste comercializa seus produtos no mercado nacional e no internacional (Sebrae, 2019).

Coleta e análise de dados

Os dados foram coletados em imagens do satélite sino-brasileiro CBERS 4A, produzido em cooperação entre o Brasil e a China, obtidas pela câmera multiespectral e pancromática de vasta varredura (WPM), referentes ao dia 25 de outubro de 2021. Uma composição colorida RGB-341 foi produzida com o SRC SIRGAS 2000/UTM/ZONA 24S (EPSG:31984).

A estratégia de classificação foi baseada em objetos (*Object Based Image Analysis - OBIA*), com a decorrente sucessão de fases: fusão, segmentação, extração de atributos, amostragem, classificação e análise de acurácia, na perspectiva de segmentação das imagens (Leão et al., 2018). O método de análise utilizado usou grupos de pixels que apresentam características semelhante, tratando-os como unidades básicas de análise (Bernardi et al., 2007). Diferentemente dos métodos tradicionais baseados em pixels, que analisam apenas os valores espectrais, esse método utiliza textura, forma e estrutura espacial e outras características multidimensionais dos pixels

vizinhos, ampliando assim, o leque de verificações entre pixels (Filippi et al., 2022).

Em auxílio à abordagem OBIA, utilizamos também o modelo Intensidade-Matiz-Saturação (IHS), que é uma representação alternativa das cores *Red*, *Green*, *Blue* (RGB), utilizada para separar a informação cromática (matiz e saturação) da informação de brilho (intensidade). Isso facilita a análise visual e computacional de imagens, sendo amplamente utilizado em técnicas de sensoriamento remoto e fusão de dados (Carper, Lillesand & Kiefer, 1990). Os componentes IHS, quando combinados com outros atributos (como forma, textura e NDVI), aumentam o número de variáveis discriminantes durante a classificação, o que eleva a acurácia dos modelos de classificação (Hay & Castilla, 2008).

Os softwares livres *Quantum Geographic Information System* (QGIS 3.36.3), Orfeo Toolbox e o plugin *Semi-Automatic Classification* foram aplicados para os diversos processamentos da imagem e confecção dos mapas. A definição para escolha dos parâmetros de segmentação foi baseada em uma análise estritamente visual, levando em conta a agregação de regiões com pixels de características semelhantes, juntamente com uma avaliação quantitativa do número total de segmentos gerados e seguindo a importância do Manual Técnico em Geociência proposto pelo Instituto Brasileiro de Geografia e Estatística em 2013, que propõem as seguintes definições e características para água, vegetação densa, vegetação rasteira, área urbana e solo exposto.

A classe “água” compreende todas as superfícies recobertas por corpos hídricos, sejam eles naturais, como rios e lagos, ou artificiais, como reservatórios e açudes. Esses elementos desempenham papel essencial nos ciclos naturais e no abastecimento humano, e sua detecção por sensoriamento remoto é facilitada pela baixa reflectância espectral da água, sobretudo no infravermelho próximo (IBGE, 2013).

Em estudos de uso e ocupação do solo, a vegetação densa é caracterizada por uma cobertura vegetal contínua, com alta biomassa e pouca ou nenhuma exposição do solo, geralmente associada a florestas tropicais, matas ciliares preservadas ou formações do tipo cerradão. Essas áreas apresentam altos valores de índices de vegetação, como o NDVI, e dificultam a visualização direta do solo em imagens de sensoriamento remoto (Anderson et al., 1976; IBGE, 2013).

A vegetação rasteira é composta predominantemente por espécies herbáceas e subarbustivas de pequeno porte, que formam uma cobertura vegetal de baixa altura e densidade, com presença variável de solo exposto. Esse tipo de

vegetação pode ocorrer de forma natural, como nos campos nativos, ou resultar de processos antrópicos, como o manejo de pastagens ou regeneração secundária (IBGE, 2013).

As áreas urbanas são definidas como porções do território ocupadas predominantemente por edificações e infraestrutura destinadas às atividades humanas, como moradia, comércio, serviços e transporte. Essas áreas apresentam intensa modificação do meio físico e são caracterizadas pela baixa cobertura vegetal e pela presença de estruturas regulares, como ruas e edificações (IBGE, 2013).

O solo exposto é caracterizado pela ausência de cobertura vegetal ou de construções, apresentando-se como superfície nua, muitas vezes associada a processos de desmatamento, erosão, obras civis ou preparo agrícola. Essas áreas podem representar situações de instabilidade ambiental ou uso transitório do solo, e são facilmente detectadas por sensoriamento remoto em função da elevada reflectância nos canais do espectro visível (IBGE, 2013).

Em seguida, foram extraídos atributos estatísticos das cinco bandas por camada por meio de um modelo gráfico, incluindo a camada do índice de Vegetação da Diferença Normalizada (NDVI) (Labib & Harris, 2018), gerando as estatísticas dos pixels por segmento, como: valor máximo, mínimo, desvio padrão, variância, média, mediana, contagem e soma. Posteriormente, foi realizado o treinamento do classificador, composto por 4.294 amostras e, consequentemente, a validação, composta por 1.840 amostras, que consiste em definir polígonos de amostras, totalizando 6.134 polígonos amostrais. Foram coletadas amostras de cada uma das seguintes classes: água = 838, vegetação densa = 540, vegetação rasteira = 811, área urbana = 3469 e solo exposto = 475. Foi incluído um quantitativo superior na classe urbana devido à imagem conter trechos com alta semelhança com solo exposto. Dessa forma, com a quantificação de mais pixels de solo urbano, o índice de erro deve diminuir consideravelmente.

Finalizado o treinamento, iniciou-se a etapa de aplicação dos métodos de classificação supervisionada: *Random Forest*, *Shark Random Forest* e *Decision Tree*. Na avaliação dos resultados foram calculadas as métricas de acurácia global e índice Kappa visando a eficiência máxima dos classificadores.

Resultados e Discussão

O processamento de imagens de sensoriamento remoto e o uso dos sistemas de informações geográficas (GIS) melhoram a

interpretação dos resultados da extração de informações e facilitam a integração entre as fases da classificação do uso e ocupação do solo (Ye et al., 2023). Nesse sentido, as bandas multiespectrais foram fundidas com a banda pancromática, resultando em uma imagem com resolução espacial de 2 m x 2 m.

Os melhores parâmetros para a segmentação foram: diferença de limiar: 0,500000, método de segmentação: *region_growing*, método de cálculo da similaridade: *Manhattan*, número mínimo de segmento: 1, quantidade de memória a usar em mb: 300 e número máximo de iterações: 20. Considerando os parâmetros avançados, foram habilitadas as funções: vizinhança (3x3) e entrada ponderada, sem desempenho e escaneamento padrão. Estas configurações resultaram em um total de 80.418 segmentos para a imagem.

Logo após a análise visual dos resultados de acurácia global e índice Kappa da (Tabela 1), pode-se concluir que, no entanto, eles são bem semelhantes.

Tabela 1. Resultados de acurácia global e índice Kappa. Fonte: Silva, Machado & Teles (2025).

Método	Acurácia global (%)	Índice Kappa
Random Forest	89,89	0,8616
Shark		
Random Forest	89,73	0,8593
Decision Tree	84,45	0,7901

Em princípio, os métodos *Random Forest* (Figura 2) e o *Shark Random Forest* apresentaram resultados semelhantes. O método *Decision Tree* gerou efeito levemente inferior, em relação aos outros dois relacionados. Algumas adversidades foram apontadas na classificação, como a presença de pontos da classe Urbana, dispersos pela área de estudo, sem sinais evidentes de ocupação. Essa situação foi observada em todas as classificações. Ao verificar a ocupação atual, constatou-se que esses pontos correspondem a classe de solo exposto.

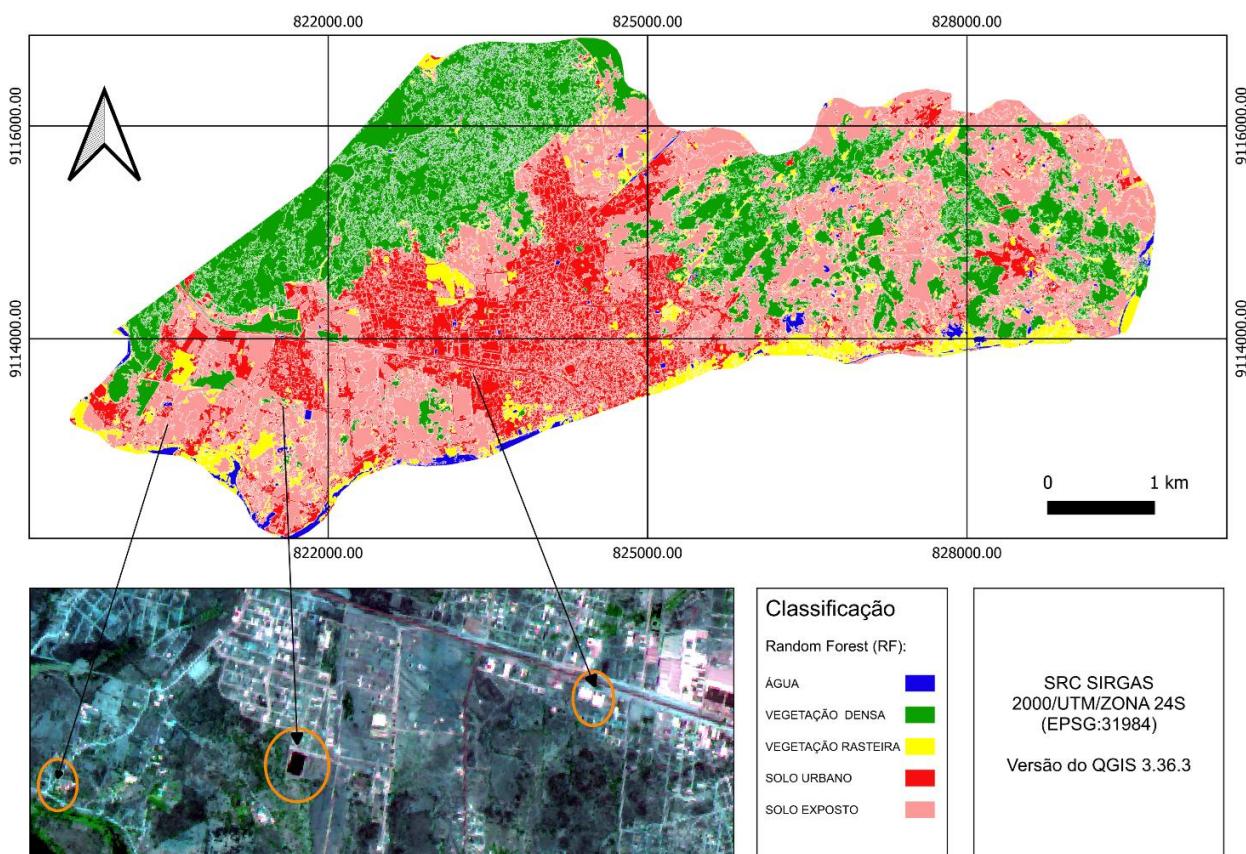


Figura 2. Resultado da classificação do uso e cobertura do solo, no município de Toritama, Pernambuco, utilizando o método *Random Forest*. Fonte: Silva, Machado & Teles (2025).

Visando evidenciar os erros dos três algoritmos, circunferências foram utilizadas em destaque. Com relação aos dados de área urbana

obtidos entre os três classificadores, esse parâmetro alcançou 24,84% de recobrimento, representando 6,39 km² de extensão da área estudada. Segundo o

Estudo Econômico das Indústrias de Confecções, realizado pelo Sebrae em 2019, o município é reconhecido em todo o país como a capital do jeans; a cidade se destaca como um importante polo têxtil e comercial. A fabricação e venda de roupas em jeans representam a principal atividade econômica de Toritama, que abriga mais de 3 mil empresas de confecção e mais de 50 lavanderias industriais. Essas indústrias são responsáveis por, aproximadamente, 15% da produção nacional de confecções em jeans, totalizando 60 milhões de peças por ano.

A classe de solo exposto apresenta uma área total de 32,01%. Isto significa que, em média, 8,24 km² de solo do município (litólicos e brunos não cárnicos) são de baixa qualidade e não servem para a agricultura ou a pecuária. Além disto, por não terem sido realizadas técnicas de realce e contraste nas bandas espectrais das imagens do satélite CBERS-04A e por terem sido capturadas em outubro, período mais seco na região, visualmente a área de solo exposto pode ter sido superestimada com vegetação gramínea ou herbácea arbustiva.

Logo, após uma análise visual dos resultados de acurácia global e índice Kappa da Tabela 1, pode-se concluir que o melhor resultado foi do classificador *Random Forest*, seguido pelo *Shark Random Forest*. Ressalta-se que a acurácia global de 89,8924%, atingida pelo método *Random Forest* foi bastante elevada, referindo-se aos resultados obtidos com as amostras de validação e, dessa forma, indicando que o resultado da imagem classificada como um todo é correto.

Analizando a matriz de confusão, em todas as classes, o método *Random Forest* (Tabela 2) foi superior aos seus comparados. Em ambos os classificadores, algumas áreas de solo exposto foram erroneamente categorizadas como vegetação rasteira ou como área urbana, erro cometido, possivelmente, pelo quantitativo de amostragem ou pelo fato da existência de sombreamento. Como também, isso se deve ao comportamento espectral parecido entre essas classes. De todo modo, isso demonstra como o método *Random Forest* é poderoso e eficaz.

Tabela 2. Dados da classificação de uso e ocupação da terra do município de Toritama, em Pernambuco, utilizando o algoritmo *Random Forest*. Fonte: Silva, Machado & Teles (2025).

Classe	Recobrimento (%)	Área (km ²)
Água	2,73	0,7

Vegetação Densa	31,12	8,02
Vegetação Rasteira	8,01	2,06
Urbano	25,98	6,69
Solo Exposto	32,16	8,28
Total		25,75

Na faixa do visível, os corpos hídricos apresentam padrão único de resposta espectral; a água limpa reflete, principalmente, o azul no infravermelho é totalmente absorvido.

No entanto, os corpos hídricos contêm não apenas água, mas também sedimentos, provenientes da erosão do solo, além de matéria orgânica, vegetação e microrganismos aquáticos. A presença desses materiais na água gera variações de coloração em imagens de composição falsa-cor, dificultando sua interpretação. Exemplo disso é a própria água, com substâncias dissolvidas (substância amarela ou matéria orgânica), organismos de efluentes (fitoplâncton e macrófitas quando presentes) e elementos não vivos (fragmento orgânico mortal e partículas inorgânicas) (Jensen & Epiphanio, 2009).

Desse modo, os estudos sobre o comportamento espectral da água são fundamentados na coleta simultânea de dados espetrorradiométricos e informações in situ relacionadas aos parâmetros de qualidade da água, com foco principal na faixa do visível, conforme à radiação fotossinteticamente ativa, que abrange de 350 nm a 700 nm.

Quanto à classificação de áreas com vegetação densa e solo exposto (Tabela 3), a análise dos dados revelou que não houve diferença significativa entre os três métodos avaliados, tanto sob o ponto de vista estatístico quanto visual (Figura 3).

Tabela 3. Dados da classificação de uso e ocupação da terra do município de Toritama, em Pernambuco, utilizando o algoritmo *Shark Random Forest*. Fonte: Silva, Machado & Teles (2025).

Classe	Recobrimento (%)	Área (km ²)
Água	3,01	0,78
Vegetação Densa	31,02	7,99
Vegetação Rasteira	7,57	1,95
Urbano	25,79	6,64
Solo Exposto	32,61	8,40
Total		25,76

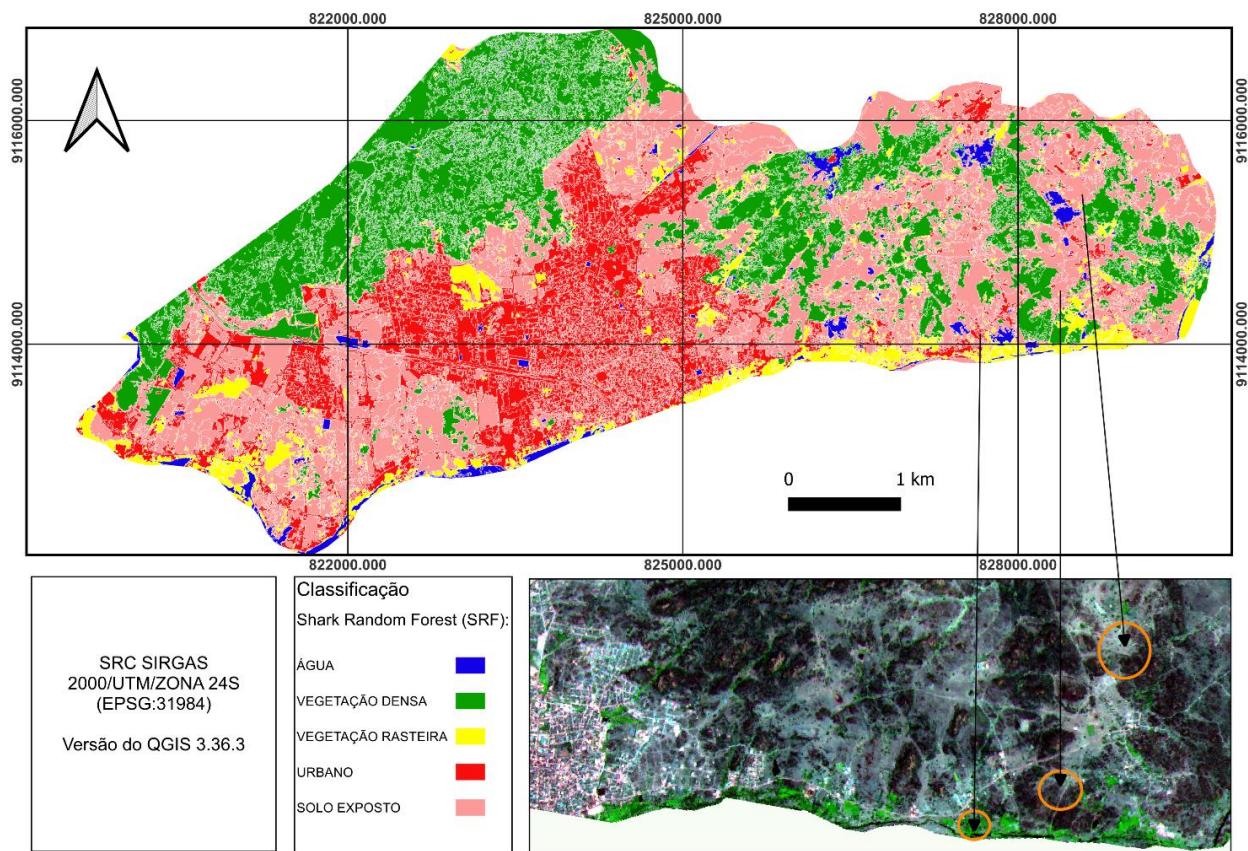


Figura 3. Resultado da classificação do uso e cobertura do solo, no município de Toritama, Pernambuco, utilizando o método *Shark Random Forest*. Fonte: Silva, Machado & Teles (2025).

Na categorização pelo método *Shark Random Forest*, ambas as classes morfológicas e espectrais tornam a identificação da superfície mais fácil; no solo exposto ocorre um aumento na reflectância, à medida em que o comprimento de onda cresce (Jensen & Epiphanio, 2009), adquirindo uma coloração quase branca, em combinação com o infravermelho próximo.

A vegetação densa destaca-se pela distribuição quase geométrica e alta densidade, facilitando sua diferenciação em relação às outras culturas perenes e às áreas de vegetação rasteira. Este cenário de cobertura vegetal destaca a expansão das áreas de ocupação humana sobre a vegetação densa, convertendo-a em espaços modificados pela própria ação humana (Albano & Galvanin, 2024). Para reduzir os impactos negativos no ambiente exposto, a posição sobre o uso e a ocupação do solo torna-se essencial, pois fornece ideias valiosas para a proteção, preservação e exploração sustentável de uma área.

Nesse contexto, a análise temporal do uso e ocupação do solo ressalta-se como um objeto vital para apoiar e orientar a gestão dos recursos

naturais, permitindo o monitoramento eficaz dessa dinâmica (Lima et al., 2021).

Nesse sentido, ficaram caracterizadas confusões significativas entre a classe água e a classe vegetação rasteira. É importante salientar, considerando o quantitativo de sombreamento na área estudada, que esse fator ajuda o algoritmo a cometer erros. Consequentemente, a classificação pelo método *Shark Random Forest* apresentou resultados satisfatórios, em termos de uma correta discriminação visual das classes estudadas.

Finalmente, o *Shark Random Forest* disponibiliza diversos algoritmos de diferentes áreas do aprendizado de máquina e inteligência computacional. Sua interface intuitiva permite que qualquer usuário combine dados com facilidade, revelando relações, padrões e tendências valiosas que, de outra forma, permaneceriam ocultas.

Entre os diversos algoritmos de *machine learning*, a árvore de decisão se destaca por sua estrutura visual intuitiva (Figura 4), além de proporcionar uma visão mais profunda do funcionamento do aprendizado de máquina.

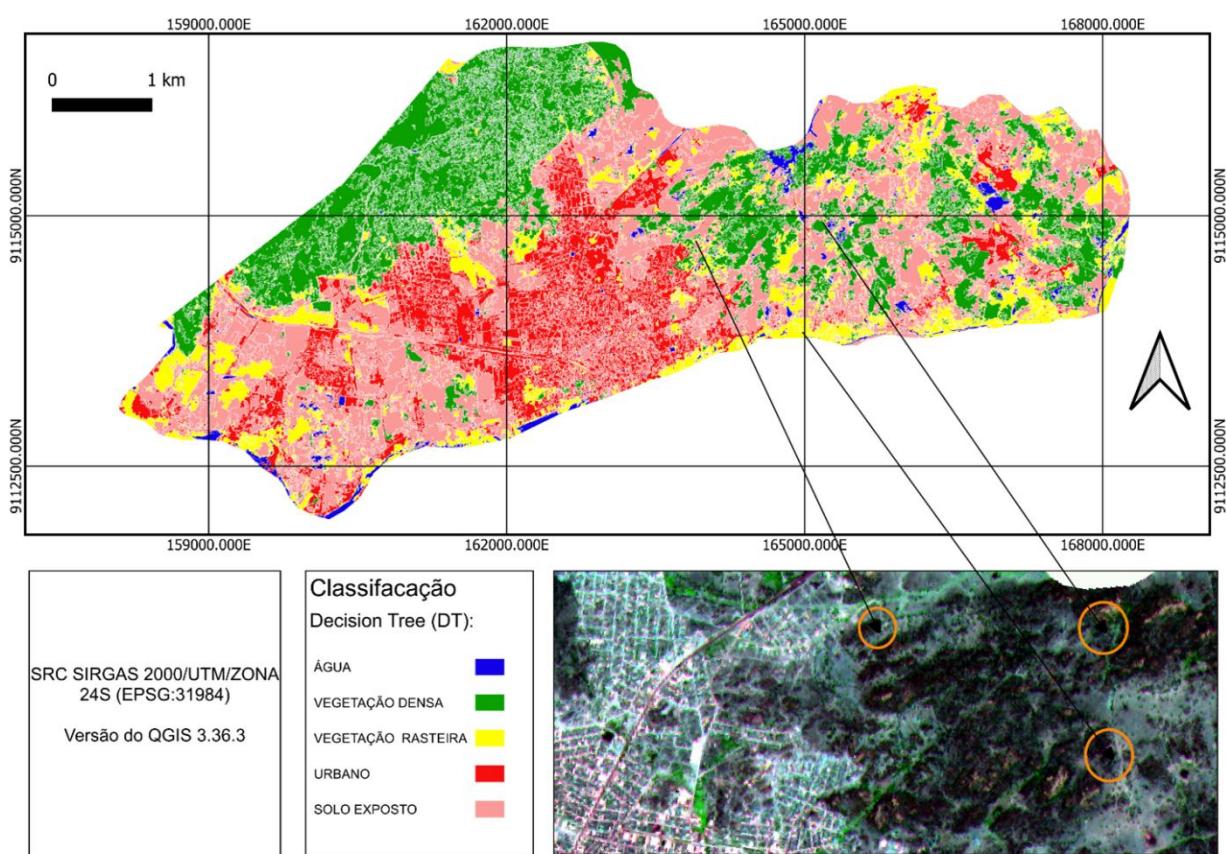


Figura 4. Resultado da classificação do uso e cobertura do solo, no município de Toritama, Pernambuco, utilizando o método *Decision Tree*. Fonte: Silva, Machado & Teles (2025).

Ademais, com os resultados da Tabela 1 e da classificação (Tabela 4), temos que o índice global de concordância Kappa obtido foi de 79,0%, o qual é considerado ‘substancial’, pois tal resultado está abaixo do valor mínimo e aceitável de 85% para uma classificação quase perfeita (Guptill et al., 1995). Por conseguinte, a acurácia global para a classificação foi de 84,45%.

Tabela 4. Dados da classificação de uso e ocupação da terra do município de Toritama, em Pernambuco, utilizando o algoritmo *Decision Tree*. Fonte: Silva, Machado & Teles (2025).

Classe	Recobrimento (%)	Área (km ²)
Água	3,13	0,81
Vegetação Densa	29,49	7,59
Vegetação Rasteira	13,35	3,44
Urbano	22,76	5,86
Solo Exposto	31,27	8,05
Total	25,75	

A priori, a classe com maior incidência na zona estudada é solo exposto, ocupando 31,27% de sua área total, seguida da classe vegetação densa, que abrange uma extensão correspondente a

29,49% em uma área com forte núcleo têxtil e comercial, no semiárido pernambucano.

Considerando as classes da Tabela 4, ficou evidenciado que a operação superestimou as classes de água e vegetação rasteira e desvalorizou as classes de solo exposto e área urbana. Apesar disso, foi perceptível que o algoritmo cumpriu bem seu objetivo. De todo modo, *Decision Tree* é um algoritmo de *machine learning* amplamente utilizado, com uma estrutura de fácil entendimento e que apresenta bons resultados em suas previsões.

Portanto, as árvores de decisão podem ser utilizadas, tanto para classificação, quanto para regressão, abrangendo ambos os tipos de tarefas de aprendizado supervisionado. Tal flexibilidade contribui fortemente para o sucesso desse algoritmo, especialmente quando comparado a métodos mais simples, como, por exemplo, *Naive-Bayes*.

Conclusão

A análise da relação entre os elementos vetorizados e as classificações automáticas apontou ótima conformidade com as classes delimitadas, indicando que os métodos selecionados possuem alto desenvolvimento no mapeamento de vegetação, quanto às mudanças de uso e cobertura da terra. Em ambos os

classificadores foi possível constatar confusões com relação à classificação de categorias que possuem comportamento espectral semelhantes, o que é esperado neste tipo de aplicação.

Analizando os resultados das métricas de índice Kappa e acurácia global, o *Random Forest* apresentou, levemente, melhores resultados em relação aos seus concorrentes. Como a abordagem de aprendizado de máquina é altamente dependente dos volumes de dados de entrada, são necessários novos testes com áreas maiores e quantidade superior de amostras de treinamento de cada classe.

Considerando o planejamento territorial, a utilização da inteligência artificial no aprendizado de máquina foi eficiente na detecção e na classificação de feições, podendo ser utilizada em diversas áreas e com diferentes finalidades, contribuindo para uma melhor gestão e tomada de decisão envolvendo os recursos naturais.

Agradecimentos

O primeiro autor agradece à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela concessão de bolsa de estudos para o mestrado e o terceiro autor agradece à Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE) pela concessão de bolsa de estudos para o mestrado.

Referências

- Albano, C. H.; Galvanin, E. A. dos S. 2024. Análise da dinâmica temporal de uso e ocupação do solo da área de proteção ambiental Bororé Colônia (2015 a 2020). Observatório de La Economía Latinoamericana, 22, 1975-1988. <https://doi.org/10.55905/oelv22n1-103>
- Anderson, J. R.; Hardy, E. E.; Roach, J. T.; Witmer, R. E. 1976. A land use and land cover classification system for use with remote sensor data. Washington: U.S. Geological Survey. 28p.
- Amulu, L. M.; Manjusha, M.; Ashok, J. 2024. An improved K-means clustering and optimal random forest for scheduling workloads in virtual servers. Evolving Systems, 15, (5), 1801-1815. <https://doi.org/10.1007/s12530-024-09586-5>
- APAC. Agência Pernambucana de Águas e Clima. 2019. Dados do Posto Pluviométrico. Disponível em: <https://www.apac.pe.gov.br/>. Acesso em: 10/01/2025.
- Belgiu, M.; Drăguț, L. 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31.
- <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Beltrão, A. B.; Mascarenhas, J. de C.; Miranda, J. L. de F.; Souza Junior, L. C. de; Galvão, M. J. da T. G.; Pereira, S. N. 2005. Diagnóstico do Município de Toritama Toritama Pernambuco Projeto Cadastro de Fontes de Abastecimento por Água Subterrânea. https://rigeo.sgb.gov.br/jspui/bitstream/doc/16986/1/Rel_Toritama.pdf. Acesso em: 16 dez. 2024.
- Bernardi, H. V. F.; Dzedzej, M.; Carvalho, L. M. T.; Acerbi Júnior, F. W. 2007. Classificação digital do uso do solo comparando os métodos “pixel a pixel” e orientada ao objeto em imagem QuickBird. Simpósio Brasileiro de Sensoriamento Remoto, 13, 5595-5602.
- Carper, W. J.; Lillesand, T. M.; Kiefer, R. W. 1990. The use of intensity-hue-saturation transformations for merging Spot panchromatic and multispectral image data. Photogrammetric Engineering and Remote Sensing, 56, (4), 459-467.
- Correa, L. R. de S.; Dantas, J. A. da S.; Verde, D. dos S. V.; Santana, A. N.; Vasco, A. N. 2024. Sensoriamento Remoto e Geoprocessamento aplicado à Agronomia. Wissen Editora. 117. <https://doi.org/10.52832/wed.117>
- Congalton, R. G.; Green, K. 2009. Assessing the accuracy of remotely sensed data: Principles and practices. 2. ed. Boca Raton: Crc Press. 210p. <https://doi.org/10.1201/9781420055139>
- Della-Silva, J. L.; Pelissari, T. D.; dos Santos, D. H.; Oliveira-Júnior, J. W.; Teodoro, L. P. R.; Teodoro, P. E.; Santana, D. C.; Oliveira, I. C.; Rossi, S. F.; Silva Junior, C. A. 2024. Land use prediction accuracy of different supervised classifiers over agriculture and livestock economy-based municipality in Brazil, Remote Sensing Applications: Society and Environment, 35, 101-257. <https://doi.org/10.1016/j.rsase.2024.101257>
- Filippi, A. M.; Güneralp, İ.; Castillo, C. R.; Ma, A.; Paulus, G.; Anders, K. H. 2022. Comparison of Image Endmember-and Object-Based Classification of Very-High-Spatial-Resolution Unmanned Aircraft System (UAS) Narrow-Band Images for Mapping Riparian Forests and Other Land Covers. Land, 2022, (11), 246. <https://doi.org/10.3390/land11020246>
- Gomez, A. H.; Gameiro, S.; Miziara, F. 2024. O Caso da Fronteira Agrícola na Região do Vale Médio do Rio Magdalena – Colômbia. História Ambiental Latinoamericana y Caribeña (HALAC) Revista de La Solcha, 14,

- (2), 237-257. <https://doi.org/10.32991/2237-2717.2024v14i2.p237-257>
- Guedes, J.; Silva, S. 2018. Sensoriamento Remoto no Estudo da Vegetação Princípios Físicos, Sensores e Métodos. *Acta Geográfica*, 12, 127-144. <https://doi.org/10.18227/2177-4307.acta.v12i29.4001>
- Guptill, S. C.; Morrison, J. L.; Association, I. C. 1995. Elements of Spatial Data Quality. Elsevier Science. Disponível em: https://books.google.com.br/books?id=i-1_AAAAMAAJ. Acesso em: 15 dez. 2024.
- Gustavo, G. S.; Elvis B. M. M.; Admilson, P. P.; Fabio, C. A.; Henrique, S. F. 2024. Mapping of impermeable surfaces in Western Bahia using Machine Learning Algorithm. *Northeast Geosciences Journal*, 10, (2), 18p. <https://doi.org/10.21680/2447-3359.2024v10n2ID36390>
- Hay, G. J.; Castilla, G. 2008. Geographic Object-Based Image Analysis: A new name for a new discipline. In: Blaschke, T.; Lang, S.; Hay, G. J. (Eds.). *Object-Based Image Analysis*. Berlin: Springer, 75-89 https://link.springer.com/chapter/10.1007/978-3-540-77058-9_4
- IBGE. Instituto Brasileiro de Geografia e Estatística. Manual técnico de uso da terra. 3. ed. Rio de Janeiro: IBGE, 2013. 171p. (Manuais Técnicos em Geociências, n. 7).
- IBGE. Instituto Brasileiro de Geografia e Estatística. 2019. Censo Demográfico 2019. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/pe/toritama.html>. Acesso em: 06 jan. 2025
- Igel, C.; Heidrich-Meisner, V.; Glasmachers, T. 2008. Shark Christian Igel Verena Heidrich-Meisner Tobias Glasmachers. *Journal of Machine Learning Research*, 9, (2008), 993-996. <https://dl.acm.org/doi/10.5555/1390681.1390714>
- Jensen, J. R.; Epiphanio, J. C. N. 2009. Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres. Parêntese Editora. 598p. <https://books.google.com.br/books?id=3dBiP-gAACAAJ>
- Labib, S. M.; Harris, A. 2018. The potentials of Sentinel-2 and LandSat-8 data in green infrastructure extraction, using object-based image analysis (OBIA) method. *European Journal of Remote Sensing*, 1-51, 231-240. <https://doi.org/10.1080/22797254.2017.1419441>
- Lary, D. J.; Alavi, A. H.; Gandomi, A. H.; Walker, A. L. 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7, (1), 3-10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- Leão Junior, E.; Galo, M. de L.; Ivánová, I. 2018. Comparação entre abordagens de classificação por pixel e baseada em objeto para o monitoramento de recursos hídricos. *Revista Brasileira de Cartografia*, 70, (4), 1289-1309. <https://doi.org/10.14393/rbcv70n4-46142>
- Lei N° 1.452/2015, Lei Municipal 2015. Disponível em: https://transparencia.toritama.pe.gov.br/uploads/5404/1/atos-oficiais/2015/plano-municipal-de-educacao/1717094538_plano-municipal-de-educacao7.pdf. Acesso em: 15 dez. 2024.
- Landis, J. R.; Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, (1), 159-174. <https://doi.org/10.2307/2529310>
- Lu, D.; Weng, Q. A. 2007. Survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28, (5), 823-870.
- Lima, S. M. S. A.; Lopes, W. G. R.; Façanha, A. C. 2021. Alterações na cobertura do solo em Teresina, Piauí, Brasil. *Sociedade & Natureza*, 33, e58922.
- Mitchell, T. M. 1997. *Machine Learning*. New York: McGraw-Hill. 421p.
- Maxwell, A. E.; Warner, T. A.; Fang, F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39, (9), 2784-2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Priyanka, K. D. 2020. Decision tree classifier: a detailed survey. *International Journal of Information and Decision Sciences*, 12-3, 246-269.
- Randhawa, K.; Loo, C. K.; Seera, M.; Lim, C.; Nandi, A. 2018. Credit Card Fraud Detection Using AdaBoost and Majority Voting. *IEEEAccess*, 6, (2018), 14277-14284. <https://doi.org/10.1109/ACCESS.2018.2806420>
- Rhodes, J. S.; Cutler, A.; Moon, K. R.; Geometry-and Accuracy-Preserving Random Forest Proximities. 2023. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, (9), 10947-10959.

- <https://ieeexplore.ieee.org/document/10089875>
- Rodriguez-Galiano, V. F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
<https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Santana-Santos, L.; Kwok, L. K.; David, D.; Daniel, B.; Craig, H.; Lawrence, J. J. 2022. Validation of Whole Genome Methylation Profiling Classifier for Central Nervous System Tumors. *The Journal of Molecular Diagnostics*, 24, (8), 924-934.
<https://doi.org/10.1016/j.jmoldx.2022.04.009>
- Sebrae. Serviço Brasileiro de Apoio às Micro e Pequenas Empresas. 2019. Boletim. Acesso em: 10/01/2025. Estudos e Pesquisas. Junho/2019. 2019. Disponível em: <https://sebrae.com.br/Sebrae/Portal%20Sebrae/UFs/PE/Anexos/RELATORIO-TORITAMA-FINAL.pdf>
- Souza, A. A.; Ferreira, K. R.; Câmara, G.; Monteiro, A. M. V. 2020. Mapping land use and land cover changes in the Brazilian Amazon using remote sensing: Opportunities and challenges. *Land Use Policy*, 94, 104313.
<https://link.springer.com/article/10.1007/s10661-024-12424-5>
- Veloso, M. F.; Reis, R. C. T. dos; Magalhães Filho, R. de; Rodrigues, F. M. 2021. Utilização do algoritmo de árvore de decisão para mapeamento do uso e cobertura da terra da área de contribuição de uma barragem localizada na Bacia Hidrográfica do São Francisco. *Caderno de Ciências Agrárias*, 13, 1-9. <https://doi.org/10.35699/2447-6218.2021.29039>
- Xia, H. L. Z.; Liu, Z.; Efremochkina, M. L. X.; Liu, X.; Lin, C. 2022. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society*, 84, 104009.
<https://doi.org/10.1016/j.scs.2022.104009>
- Ye, Z.; Yang, K.; Lin, Y.; Guo, S.; Sun, Y.; Chen, X.; Lai, R.; Zhang, H. 2023. A comparison between Pixel-based deep learning and Object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV Visible-light images. *Computers and Electronics in Agriculture*, 209, 107822.
<https://doi.org/10.1016/j.compag.2023.107822>