

Computação em Nuvem e Aprendizado de Máquina: Comparação de imagens OLI/Landsat-8 e MSI/Sentinel-2A no Mapeamento de Uso e Cobertura da Terra no Semiárido Nordestino

Alfíbia Deysi Guedes da Silva¹, Sara Fernandes de Flor de Souza², Rebecca Luna Lucena³

¹(Universidade Federal do Rio Grande do Norte - Programa de Pós-Graduação em Geografia - GEOCERES, Caicó, RN, Brasil, albiadeysi.prof@gmail.com), ²(Universidade Federal do Rio Grande do Norte - Programa de Pós-Graduação em Geografia - GEOCERES, Caicó, RN, Brasil, sara.flor@ufrn.br), ³(Universidade Federal do Rio Grande do Norte - Programa de Pós-Graduação em Geografia - GEOCERES, Caicó, RN, Brasil, rebeccaosvaldo@yahoo.com.br).

Received 10 November; accepted 19 December

Resumo

O processamento em nuvem de imagem de multisensores de média e alta resolução tem impulsionado o crescimento de análises ambientais baseadas em algoritmos de aprendizado de máquina. Logo, o objetivo deste estudo foi utilizar a plataforma Google Earth Engine em conjunto com técnicas de machine learning para realizar uma análise comparativa das respostas obtidas pelos sensores OLI/Landsat-8 e MSI/Sentinel-2A na delimitação de classes de uso e cobertura da terra do município de Jardim de Piranhas/RN no ano de 2022. Foram utilizados o NDVI e o método Random Forest para classificação de 5 classes: Vegetação Densa, Vegetação Aberta, Área Descoberta, Agricultura e Corpos Hídricos. Para ambos os sensores, os NDVIs foram similares com correlação forte ($r = 0.89$), classificações com índice Kappa "excelente", e Exatidão Global de 89% para o OLI e 92% para o MSI. O sensor MSI/Sentinel resultou em uma melhor delimitação dos alvos, sendo a acurácia do produtor e do usuário satisfatórias. A metodologia aplicada e os resultados obtidos poderão contribuir para o desenvolvimento de políticas públicas de planejamento ambiental do município de Jardim de Piranhas e a mesma pode ser reproduzida para outras áreas do semiárido brasileiro. O estudo de dados espaciais, mediante geotecnologias, torna-se fundamental para o ordenamento territorial, visando o equilíbrio da proteção dos ecossistemas, a segurança hídrica e o desenvolvimento urbano.

Palavras-chave: Google Earth Engine, Random Forest, NDVI, Índice Kappa.

Cloud Computing and Machine Learning: Comparison of OLI/Landsat-8 and MSI/Sentinel-2A images in Land Use and Land Cover Mapping in the Semi-arid Northeast of Brazil

Abstract

Cloud processing of medium- and high-resolution multisensor images has boosted the growth of environmental analysis based on machine learning algorithms. Therefore, the aim of this study was to use the Google Earth Engine platform in conjunction with machine learning techniques to carry out a comparative analysis of the responses obtained by the OLI/Landsat-8 and MSI/Sentinel-2A sensors in the delimitation of land use and land cover classes in the municipality of Jardim de Piranhas/RN in the year 2022. NDVI and the Random Forest method were used to classify 5 classes: Dense Vegetation, Open Vegetation, Open Area, Agriculture and Water Bodies. For both sensors, the NDVIs were similar with a strong correlation ($r = 0.89$), classifications with an "excellent" Kappa index, and Global Accuracy of 89% for OLI and 92% for MSI. The MSI/Sentinel sensor resulted in better target delimitation, with satisfactory accuracy for the producer and the user. The methodology applied and the results obtained could contribute to the development of public policies for environmental planning in the municipality of Jardim de Piranhas and could be reproduced in other areas of the Brazilian semi-arid region. The study of spatial data, using geotechnologies, is essential for territorial planning, with a view to balancing the protection of ecosystems, water security and urban development.

Keywords: Google Earth Engine, Random Forest, NDVI, Kappa index.

1 Introdução

O Sensoriamento Remoto (SR) é uma geotecnologia que possibilita a obtenção de informações de grandes áreas da superfície terrestre através da captação remota da energia refletida ou emitida pela superfície pelos alvos terrestres (Florenzano, 2002; Lorenzetti, 2015). Os avanços nas ferramentas e algoritmos de

Processamento Digital de Imagens (PDI) têm impulsionado a análise ambiental por meio de dados provenientes de imagens ópticas (Yang *et al.*, 2013). Nesse contexto, a plataforma Google Earth Engine (GEE) combina o processamento em rede e os algoritmos baseados em aprendizado de máquina (machine learning), viabilizando uma

análise rápida a partir da identificação dos padrões e parâmetros espectrais (Chang; Bai, 2018).

O GEE é uma ferramenta de computação em nuvem (cloud-based computing) lançada pela Google em 2010 com acesso direto pelo navegador de pesquisa, projetada para lidar com big data de SR. Sua eficiência em executar rapidamente tarefas de geoprocessamento em grandes conjuntos de dados eliminou a necessidade de baixar ou instalar programas adicionais, superando as metodologias tradicionais (Gorelick *et al.*, 2017; Amani *et al.*, 2020). Machine Learning (ML) representa uma subárea da Inteligência Artificial, conferindo aos computadores a habilidade de aprender tarefas específicas (Chang; Bai, 2018). No âmbito do SR, essa capacidade de autoaprendizagem destaca-se pela sua eficiência e precisão na classificação de imagens, pois permite integrar diversos tipos de dados, aumentando a combinação automática de variáveis para mapeamento, como localização geográfica, topografia, meteorologia (Shwartz-Shalev; Ben-David, 2014).

A disponibilidade de multisensores de média e alta resolução com características espectrais e espaciais distintas cujas capacidades atendem a diferentes aplicações e escalabilidade, tem expandido as possibilidades de análise espacial. Com destaque para o sensor OLI do satélite Landsat-8 e para o sensor MSI do satélite Sentinel-2 que são amplamente usados para estudos ambientais, pois oferecem bandas multiespectrais semelhantes e multiresoluções. A missão LANDSAT (Land Remote Sensing Satellite) foi desenvolvida pela NASA (National Aeronautics and Space Administration) em colaboração com o USGS (United States Geological Survey), atualmente conhecida por sua série temporal padronizada de mais de 40 anos, lançada em 1972. O satélite Landsat-8, lançado em 2013, é equipado com dois sensores: OLI (Operational Land Imager) e o TIRS (Thermal Infrared Sensor). O OLI é composto por 8 bandas multiespectrais de resolução espacial de 30 metros (Tabela 1) e resolução radiométrica de 16 bits. A altitude do satélite é de 705 km, a faixa imageadora é de 185 km e o período de revisita é de 16 dias. Além disso, possui uma banda denominada PAN de 15 metros de resolução espacial (USGS).

A missão Sentinel-2 possui dois satélites, o Sentinel-2A e Sentinel-2B, equipados com o sensor MultiSpectral Instrument (MSI) capaz de gerar imagem multiespectral de alta resolução, são parte do programa Copernicus da Agência Espacial Europeia (ESA). O MSI possui uma faixa imageadora de 290km e conta com 13 bandas espectrais (Tabela 1), resolução radiométrica de 12

bits e temporal de 5 dias. As imagens abrangem 4 bandas espectrais nas regiões do visível e 6 bandas no infravermelho próximo (NIR). As bandas no visível 2, 3, 4 e 8 possuem resolução espacial de 10 metros, enquanto as 6 bandas no red edge e infravermelho de ondas curtas (SWIR) 20 metros (bandas 5, 6, 7, 8A, 11 e 12). Além disso, existem 3 bandas para correções atmosféricas de 60 metros (Drusch, 2012).

A coleção de imagens dos sensores OLI e MSI estão disponíveis na plataforma do GEE, onde é possível aplicar algoritmos de aprendizado de máquina para identificar padrões, processar relações complexas e detectar tendências em grandes conjuntos de dados. O método Random Forest combina o conjunto de árvores aleatórias de decisão, tornando-se confiável para classificação de uso e cobertura da terra (Naushad; KAUR; Ghaderpour, 2021). Além disso, a utilização de Índices Espectrais, como o Índice de Vegetação da Diferença Normalizada (NDVI) derivado da combinação de bandas específicas, constitui técnicas poderosas para a análise da saúde da vegetação, ressaltando seu comportamento espectral e possibilitando a distinção de diferentes coberturas da terra (Moreira, 2012). O NDVI é um índice sensível à densidade da vegetação e ao teor de clorofila das folhas, logo, é indicado para mensurar a quantidade vegetal na superfície, bem como, a saúde das plantas. Utilizando as bandas do infravermelho próximo e do vermelho é possível quantificar a vegetação, identificar padrões e mudanças na sua cobertura (MOREIRA, 2012).

O município de Jardim de Piranhas está localizado na bacia hidrográfica do rio Piancó-Piranhas-Açu, especificamente da sub-bacia do Rio Piranhas-Açu, ao sul do estado do Rio Grande do Norte no semiárido nordestino. Segundo os dados do MapBiomass, o setor agropecuário saltou de 39% em 1985 para 76% da área total do município em 2021. Em paralelo, a classe floresta regrediu cerca de 26% (de 46% para 20%) no mesmo período. A remoção indiscriminada da vegetação provoca o aumento de perda dos solos, assim como, afeta a qualidade da água, causando retorno às comunidades humanas e faunísticas, sobretudo, em ambientes de bacias hidrográficas submetidas às condições de semiaridez, pois afeta a disponibilidade hídrica potável (Refati *et al.*, 2023; Rocha, 2014).

O semiárido brasileiro é predominantemente localizado na região Nordeste, caracterizando-se por baixas precipitações, elevadas temperaturas, rios intermitentes e uma vegetação adaptada às condições de escassez hídrica. Destaca-se nessa área uma das maiores

florestas secas do mundo, o bioma Caatinga (Silva; Leal; Tabarelli, 2018). Essa peculiaridade climática e ambiental abrange aproximadamente 93% do território do estado do Rio Grande do Norte (Vale *et al.*, 2020). As condições naturais do semiárido brasileiro restringem a economia da região, propiciando o desenvolvimento de atividades diretamente relacionadas à disponibilidade de biomassa, como o pastoreio extensivo, agricultura de sequeiro e extração de lenha. (Silva; Leal; Tabarelli, 2018).

De acordo com estudos realizados pelo Painel Intergovernamental sobre Mudanças Climáticas, as regiões semiáridas do Nordeste brasileiro poderão ser atingidas pelo aumento de secas ecológicas e agrícolas, que resultarão em situações de desastre socioambiental (IPCC, 2021). Associada a esse prognóstico climático, tem-se a pressão antrópica que se dá através dos usos do solo e da água que, realizados sem planejamento, geram a degradação ambiental (Chueh, 2005). Segundo Sun e Scanlon (2019) as informações espacializadas são capazes de apresentar a organização territorial, os graus de conservação e artificialização, contribuindo para a elaboração e implementação de projetos de planejamento e gestão territorial para a conservação e uso sustentável dos recursos naturais.

O objetivo desta pesquisa é utilizar a plataforma Google Earth Engine em conjunto com técnicas de machine learning para realizar uma análise comparativa das respostas obtidas pelos sensores OLI/Landsat-8 e MSI/Sentinel-2A, no contexto da delimitação de diferentes classes de uso e cobertura da terra do município de Jardim de Piranhas localizado na bacia hidrográfica Piancó-Piranhas-Açu no Rio Grande do Norte/RN no ano de 2022.

2 Material e métodos

O município de Jardim de Piranhas está localizado na sub-bacia do Rio Piranhas-Açu, ao sul do estado do Rio Grande do Norte (Figura 1). O município apresenta um clima característico de semiárido quente e seco (Bsh) na classificação de Köppen, sendo a vegetação predominante a Caatinga. A cidade possui uma rede industrial têxtil que tem influenciado a organização espacial e a economia da cidade, sendo conhecida pela produção de tecidos, principal, de panos de prato, mas também se fabricam redes de dormir, toalhas e cobertores (Santos, 2021). Com as atividades agrícolas se localizando nas margens do rio Piranhas, o que vem provocando alterações na cobertura de vegetação (Pereira, 2021; Santos, 2019).

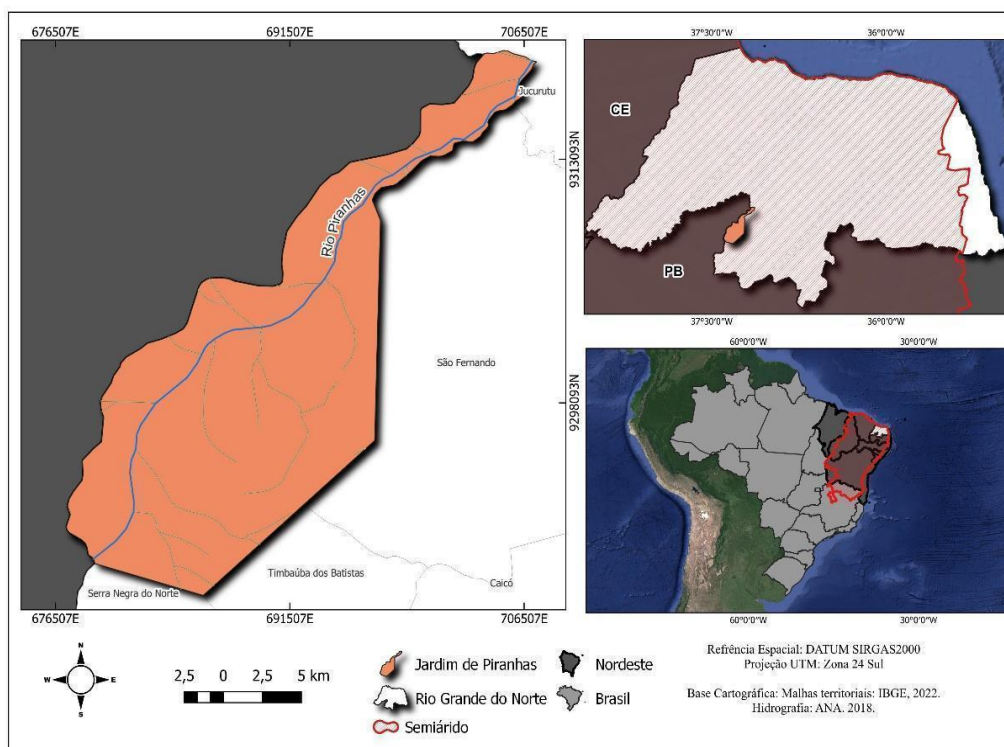


Figura 1. Localização do Município de Jardim de Piranhas/RN.

A utilização do GEE exige habilidades em linguagens de programação, como JavaScript ou Python, entretanto, a plataforma oferece uma interface intuitiva, conhecida como API (Application Programming Interface), que disponibiliza diversos scripts prontos em bibliotecas (Mutanga; Kumar, 2019). Além disso, a ferramenta está em constante atualização, incorporando técnicas de machine learning avançadas. Essa particularidade permite aos

pesquisadores aplicarem algoritmos diretamente nos dados de SR disponíveis, explorando de forma eficaz padrões complexos nos dados geoespaciais. Logo, todos os procedimentos de processamento de imagens orbitais, cálculo do NDVI, classificação supervisionada, cálculos estatísticos e geração de gráficos foram realizados em computação em nuvem no Google Earth Engine na linguagem de programação JavaScript, conforme Figura 2.

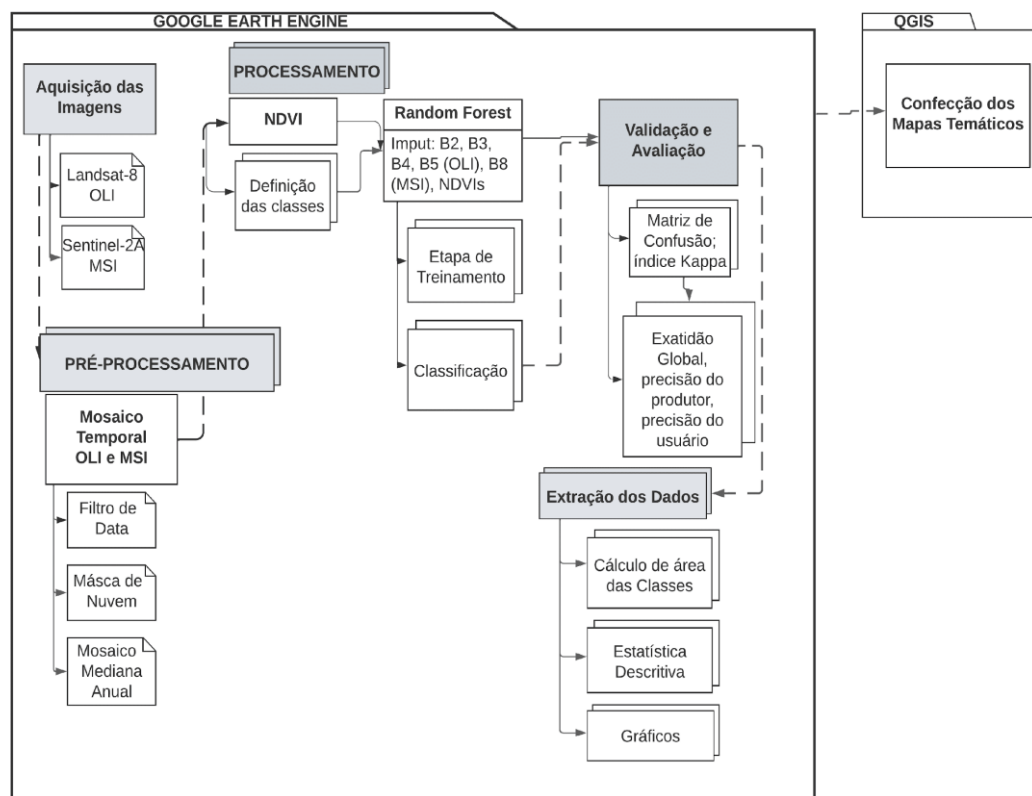


Figura 2. Fluxograma das etapas metodológicas executadas no trabalho.

A primeira etapa correspondeu aos processos iniciais de aquisição e seleção das imagens, em que foram adquiridos os produtos dos sensores OLI/Landsat-8 SR e MSI/Sentinel-2A. Foi realizado o corte para a área de interesse e filtragem de data: ano completo de 2022. Além disso, aplicou-se uma máscara de nuvem para remover interferências, sombras e ruídos. A partir disso, toda a coleção foi filtrada para um mosaico da mediana (median()) dos pixels, resultando em um único produto.

Seguiu-se, então, para a segunda etapa, que consistiu em aplicar técnicas de tratamento das imagens. Inicialmente foi calculado o índice de vegetação NDVI integrado ao GEE (NormalizedDifference()) a partir das bandas do infravermelho próximo (NIR) e do vermelho (RED) do Landsat (B5 e B4) e do Sentinel (B8 e

B4), conforme a equação elaborada por Rouse *et al* (1974): $(NIR-RED)/(NIR+RED)$.

O passo seguinte foi a definição das classes e a coleta de amostras de treinamento. A chave de interpretação elaborada de uso e cobertura da terra estabeleceu 5 classes: Vegetação Densa, Vegetação Aberta, Agricultura, Área Descoberta e Corpos Hídricos (Quadro 1). A análise das imagens considerou a diferenciação das estruturas das feições vegetais, como a densidade, e classificou-as conforme o padrão de comportamento espectral dos alvos. Fracionou-se as áreas de vegetação natural em caatinga densa e aberta.

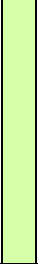
O processo de coleta das amostras foi realizado manualmente por meio da geometria ponto, sendo gerados cerca de 600 pontos ao longo do município de Jardim de Piranhas, divididos entre os dois produtos e 5 classes. O método de classificação supervisionada utilizado foi baseado

em machine learning a partir do classificador Random Forest (RF) com 100 ramificações da árvore de decisão: “`ee.Classifier.smileRandomForest(100)`”. Os parâmetros de entrada (*input*) para a classificação foram as medianas das bandas espectrais do visível (B2, B3 e B4, ambos os sensores), a banda do infravermelho próximo (B5/OLI e B8/MSI) e os NDVIs.

O Random Forest foi desenvolvido por Breiman (2001) como um algoritmo de machine

learning que consiste em uma combinação de um conjunto de árvores de decisão, aleatoriamente, para realizar classificações. A combinação é feita por meio da divisão desse em dois: um de treinamento e outro de teste, através do método bagging. O subconjunto de treinamento é para treinar o modelo e, em seguida, realizar as classificações no subconjunto de teste. A classificação final é feita considerando a média aritmética entre os subconjuntos treinamento e teste (Zhang; Yang, 2020).

Quadro 1. Chaves de interpretação

Classes		Amostras (OLI) RGB543	Google Earth Pro	Descrição
Vegetação Densa				Área com vegetação pouco esparsa, entre adensada e semiadensada. Caracteriza-se pelos tons de vermelho escuro.
Vegetação Aberta				Área de vegetação esparsamente distribuída.
Agricultura				Áreas com predomínio de algum tipo de cultivo de plantas. Apresenta forma regular e tons de vermelho claro.
Área Descoberta				Área sem nenhuma cobertura vegetal e áreas construídas.
Corpos Hídricos				Áreas de corpos d'água continentais.

As etapas posteriores são de tratamentos estatísticos aos produtos gerados. Com relação aos NDVIs foram extraídas informações estatísticas, também na plataforma do GEE, onde é preciso utilizar a biblioteca de Reducers. A priori foi realizada a estatística descritiva, como os valores máximos, médios, mínimos e desvio padrão de cada NDVI resultante (`ee.Reducer.max()`; `ee.Reducer.mean()`; `ee.Reducer.min()`; `ee.Reducer.stdDev()`). Além disso, foi aplicado o Coeficiente de Correlação de Pearson (r) para avaliar o grau de correspondência linear entre o $NDVI_{OLI}$ e $NDVI_{MSI}$ (`ee.Reducer.pearsonsCorrelation()`), o qual varia de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita) (Mukaka, 2012). Entretanto, antes de aplicar a correlação, foi necessário converter as imagens para dados numéricos a partir de listas (arrays).

Para entender o comportamento espacial da taxa de cobertura da vegetação em Jardim de Piranhas nos dois sensores, o resultado do NDVI foi dividido em 5 intervalos: valores negativos; 0-0.2; 0.2-0.5; 0.5-0.7; e acima de 0.7. Desse modo, foram feitas análises a partir do cálculo de área das classes pela função `ee.Image.pixelArea()`. Ao produto de classificação gerado foi extraído o cálculo de área de cada classe a partir da combinação da função `ee.Image.pixelArea()` e `reduceRegion()`. Os resultados também foram calculados em porcentagem para melhor visualização em gráficos de pizza (`ui.Chart`) gerados no GEE.

Na avaliação estatística da classificação supervisionada, utilizou-se as amostras de teste para avaliar a acurácia do algoritmo de machine learning. A matriz de confusão foi a primeira etapa (`ee.ConfusionMatrix()`), que consiste em comparar a classificação obtida com o conjunto de validação, e permite avaliar o desempenho da classificação. De modo aleatório 70% foram destinados para o treinamento e 30% para o teste de acurácia do classificador.

Em seguida, foram calculados a exatidão global (Overall Accuracy) e o índice Kappa, ambos implementados no GEE. O primeiro indica a porcentagem de pixels ou amostras classificadas corretamente em relação ao total, o segundo fornece uma métrica de concordância entre as classificações obtidas e as observações reais a

partir da matriz de confusão (Congalton & Green 1999). De acordo com Landis e Koch (1977), os resultados Kappa são qualificados em 6 valores de desempenho: péssimo ($< 0,000$); ruim (0,00-0,200); regular (0,201-0,400); bom (0,401-0,600); muito bom (0,601-0,800); e excelente ($> 0,801$). Sendo, duas medidas de precisão calculadas: A Acurácia do Usuário que é uma indicativa de probabilidade referente à proporção de instâncias positivas corretamente classificadas pelo modelo em relação ao total de instâncias que o usuário considera como positivas, ou seja, a amostragem; E Acurácia do Produtor que se refere à proporção de instâncias positivas corretamente classificadas pelo modelo em relação ao total de instâncias que o modelo classificou como positivas (Congalton, 1991).

3 Resultados e discussão

A plataforma Google Earth Engine disponibiliza todos os produtos da USGS que incluem as imagens TOA (Top-of-Atmosphere Reflectance), SR (Surface Reflectance) e LST (Land Surface Temperature). Os produtos SR do Landsat-8 disponíveis no GEE são gerados pelo algoritmo LaSRC que considera os efeitos atmosféricos, como aerossóis, melhorando a visualização das informações visíveis. Os produtos do sensor MSI/Sentinel-2A, também disponíveis no GEE desde março de 2017, apresentam reflectância de superfície corrigida e ortorretificada. A partir de janeiro de 2022, a coleção passou por uma melhoria na linha de base de processamento, neste caso, no GEE às imagens foram adicionadas o código “harmonized”.

A discussão dos resultados se baseou em análises comparativas realizadas entre os produtos do NDVI e as classificações digitais obtidas pelos diferentes sensores OLI e MSI acoplados nos satélites Landsat-8 e Sentinel-2A, respectivamente utilizando o processamento em nuvem na plataforma Google Earth Engine, e o algoritmo Random Forest com scripts gerados em JavaScript. As análises foram pautadas a partir de métricas de Exatidão Global, Índice Kappa, Acurácia do Usuário e Acurácia do Produtor, assim como, análise dos padrões visuais e cálculo de áreas.

Tabela 1. Parâmetros espectrais dos sensores OLI-TIRS/Landsat-8 e MSI/Sentinel-2A.

Bandas OLI	Comprimento de Onda (µm)	Resolução Espacial	Bandas MSI	Comprimento de Onda (µm)	Resolução Espacial
1 - Aerosol	0,443	30 m	1 - Aerosol	0,443	60 m
2 - Azul	0,483	30 m	2 - Azul	0,490	10 m
3 - Verde	0,561	30 m	3 - Verde	0,560	10 m
4 - Vermelho	0,654	30 m	4 - Vermelho	0,665	10 m
5 - NIR	0,864	30 m	5 - Red Edge 1	0,705	20 m
6 - SWIR 1	1,609	30 m	6 - Red Edge 2	0,740	20 m
7 - SWIR 2	2,201	30 m	7 - Red Edge 3	0,783	20 m
8 - Pancromática	0,589	15 m	8A - Red Edge 4	0,865	25 m
9 - Cirrus	1,364	30 m	8 - NIR	0,842	10 m
10 - TIR 1	10,895	100 m	9 - Vapor D'água	0,940	60 m
11 - TIR 2	12,005	100 m	10 - Cirrus	1,375	60 m
			11 - SWIR 1	1,610	20 m
			12 - SWIR 2	2,190	20 m

Fonte: Adaptado de USGS (2023) e ESA (2023).

3.1 Índice de Vegetação da Diferença Normalizada - NDVI

O Coeficiente de Correlação de Pearson (r) calculado entre os produtos de NDVI_{OLI} e NDVI_{MSI} gerados na plataforma do GEE, resultou em 0.89 e indica uma correlação positiva moderadamente forte entre os dois resultados. Nesse caso, os valores de NDVI obtidos nos dois sensores tendem a aumentar ou diminuir juntos de forma consistente. Isso sugere que os NDVIs apresentam uma relação positiva, ou seja, quando um dos métodos indica uma maior densidade ou saúde da

De acordo com Jensen (2011), os valores negativos atendem aos corpos hídricos, já os valores positivos próximos a zero são áreas descobertas ou com baixa cobertura vegetal, como areia, solos, rochas expostas e área urbana. Quanto mais se aproximar de 1, mais saudável é a

vegetação, o outro também tende a indicar uma condição análoga.

Na tabela 2 destaca-se os resultados das médias que apresentam vegetação com baixa a moderada atividade fotossintética no município de Jardim de Piranhas, neste caso, o NDVI_{MSI} foi ligeiramente inferior. Isso sugere que, no caso do MSI, a vegetação na região manifesta um desempenho fotossintético um pouco menor que o OLI.

vegetação, neste caso, áreas com vegetação com alta atividade fotossintética correspondem a valores maiores que 0.7. A Figura 3 apresenta os NDVIs resultantes e também os gráficos histogramas dos pixels dos sensores.

Tabela 2. Valores estatísticos dos sensores OLI e MSI

Sensores	Mínimo	Média	Máximo	Desvio Padrão
OLI/Landsat	-0,68	0,43	0,88	0,12
MSI/Sentinel	-0,42	0,38	0,89	0,14

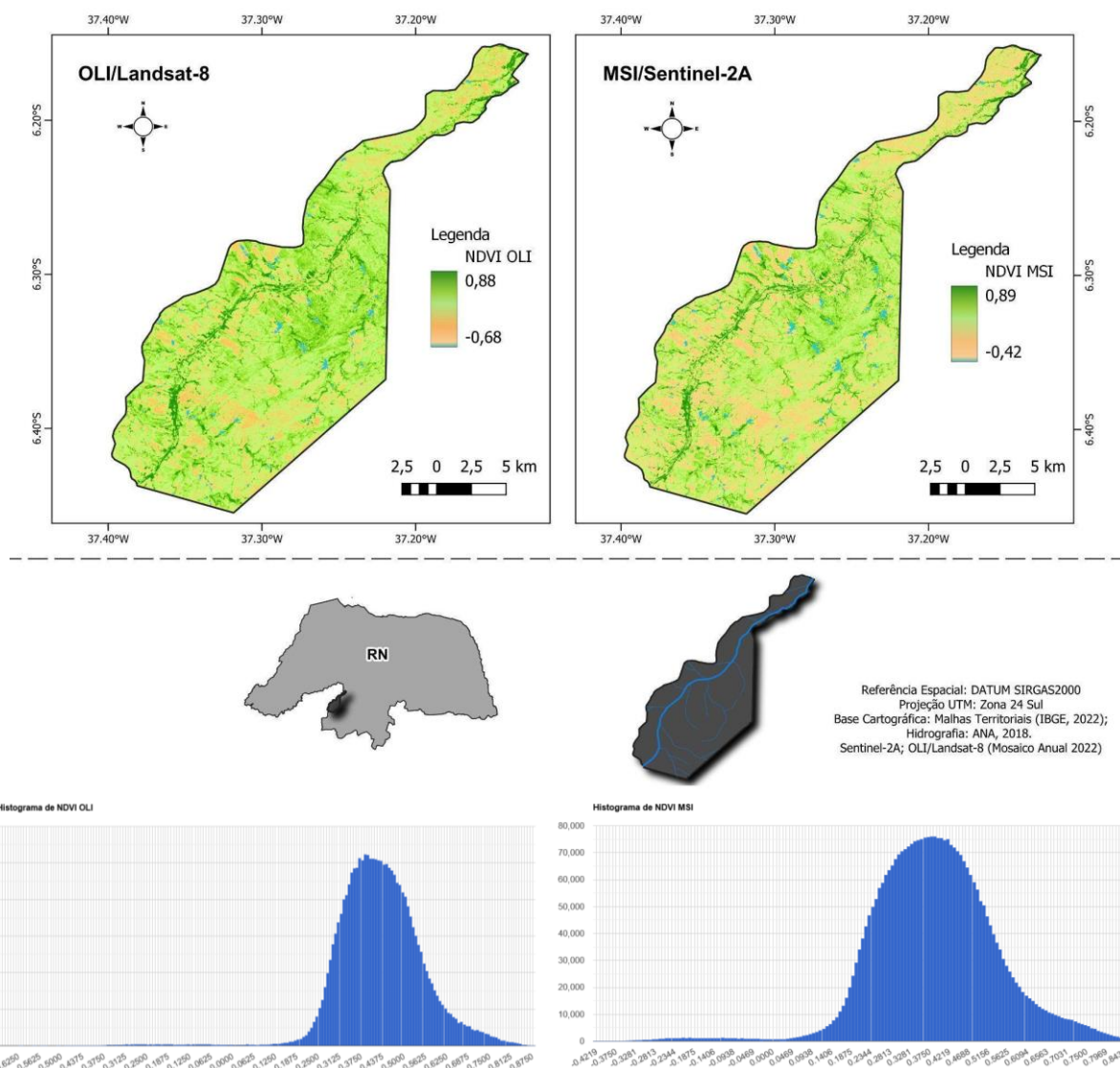


Figura 3. Resultados dos NDVIs e gráfico histograma.

A Tabela 3 apresenta a área das classes em km². A maior discrepância entre os sensores MSI e OLI ocorreu na classe 2. O histograma (Figura 3) indica que no sensor MSI o início da curvatura crescente ocorre a partir de 0.05, enquanto o OLI

somente a partir de 0.18, isso explica a classe 2 ser 12,8 km² maior no NDVI_{MSI}. Este intervalo inclui áreas com baixa ou nenhuma cobertura vegetal, representada por areia, solo exposto e áreas construídas.

Tabela 3. Métricas das classes de NDVI em km².

NDVI	OLI/Landsat	MSI/Sentinel
Classe 1 (-1 a 0)	3	4,9
Classe 2 (0 a 0.2)	2,7	15,5
Classe 3 (0.2 a 0.5)	239,3	245,2
Classe 4 (0.5 a 0.7)	77,7	56,5
Classe 5 (0.7 a 1)	7,6	8,3

Fonte: Elaboração da pesquisa (2023).

O histograma revelou que os sensores MSI e OLI têm concentrações maiores de frequência do NDVI entre 0.336 e 0.414, abrangendo a faixa de vegetação de baixo a médio vigor, o que corrobora a Classe 3 ser a maior em extensão (Tabela 3).

Além disso, o histograma mostra maior agrupamento de pixels acima de 0.5 no OLI, sendo possível a visualização na Figura 3 de maior cobertura de tons esverdeados. Entretanto, a maior extensão da vegetação saudável foi detectada no

sensor MSI. Os NDVIs evidenciaram que a vegetação saudável correspondeu a mata ciliar próxima ao rio principal e a outros corpos d'água, como também a áreas de agricultura.

3.2 Classificação Supervisionada

A classificação supervisionada envolve o treinamento dos algoritmos com dados de entrada rotulados para que o modelo reconheça os padrões espectrais, texturais e espaciais que caracterizam diferentes tipos de alvos em toda a imagem, como vegetação, solo, água etc. Para a análise de uso e cobertura de Jardim de Piranhas, optou-se por trabalhar com as bandas espectrais equivalentes entre os sensores e com a mesma resolução

espacial, adicionada a banda NDVI. A partir das chaves de interpretação criadas (ver Quadro 1), considerou-se o mesmo número de amostras em posições similares para ambos os sensores, a fim de minimizar interferências ao classificador.

Os mapeamentos de uso e cobertura vegetal da terra estão dispostos na Figura 4. Visualmente, as classificações resultantes dos sensores OLI e MSI não apresentam grandes discrepâncias, e a Tabela 4 afirma que ambos os sensores obtiveram um bom desempenho e qualidade excelente para o índice Kappa na classificação supervisionada. Ainda assim, o MSI apresentou os melhores valores e é perceptível mais delineamento dos alvos na superfície.

Tabela 4. Exatidão global e índice Kappa dos sensores OLI e MSI.

Sensores	Exatidão Global	Índice Kappa	Qualidade
OLI/Landsat	89%	0,86	Excelente
MSI/Sentinel	92%	0,90	Excelente

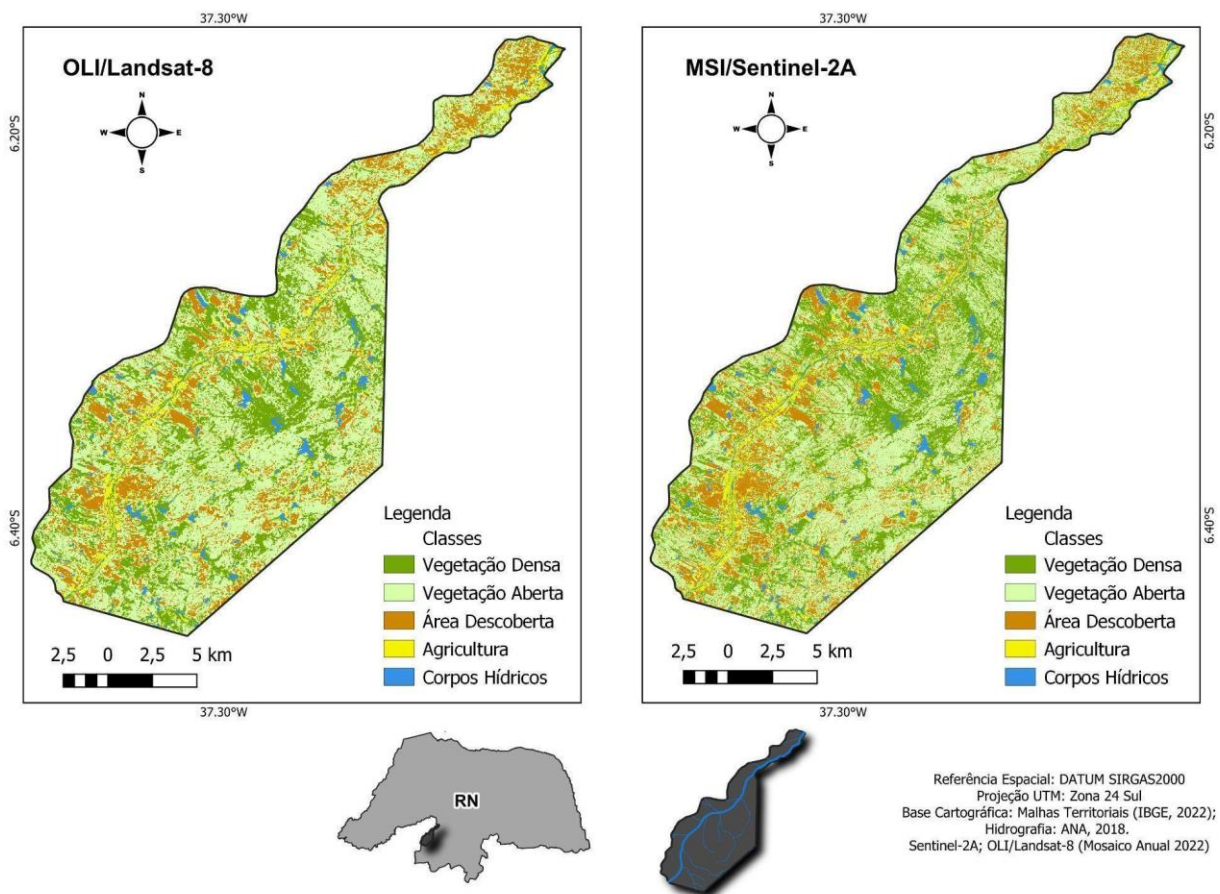


Figura 4. Mapeamentos de uso e cobertura da terra dos sensores OLI e MSI.

As Tabelas 5 e 6 apresentam as matrizes de confusão das classificações supervisionadas

geradas na plataforma GEE dos produtos dos sensores OLI e MSI, respectivamente. Essa matriz

permite uma análise mais detalhada do desempenho do modelo de classificação com base nas amostras de entrada. Verifica-se o baixo resultado de acurácia do usuário para a classe de “Vegetação Densa” no sensor OLI (70%), que indica baixo percentual de acerto da classificação em relação a verdade na superfície, esse percentual é superior no sensor MSI, no qual 83% das amostras foram corretamente classificadas.

Esses dados podem estar relacionados com a dificuldade que o usuário apresenta ao identificar e diferenciar pequenas áreas que intercalam vegetação semi-densa e aberta na resolução espacial de 30m do OLI/Landsat-8. Esse fator

também resultou em acurácia de usuário e de produtor de 80% na classe de “Vegetação Aberta” no OLI. As classes de vegetação densa e aberta apresentaram as maiores confusões em ambos os sensores, mas no MSI foi mais discreto. O MSI/Sentinel-2A obteve 84% de acurácia do produtor na vegetação esparsa, que se refere à probabilidade de um pixel ser classificado corretamente, e 90% para a vegetação densa. São resultados esperados, visto que a literatura científica ainda não definiu um limite de cobertura para essas classes, além disso, a atividade humana de uso da terra é o principal fator de mudança dessa cobertura.

Tabela 5. Matriz de confusão e acurácia do produtor e do usuário do sensor OLI.

	Amostras OLI					Acurácia Usuário
	Vegetação Densa	Vegetação Aberta	Área Descoberta	Agricultura	Água	
V. Densa	14	3	0	0	0	0,70
V. Aberta	3	16	1	0	0	0,80
Área Descoberta	0	1	13	0	0	0,93
Agricultura	2	0	0	17	0	1
Água	1	0	0	0	28	1
Acurácia Produtor	0,82	0,80	0,93	0,90	0,96	

Tabela 6. Matriz de confusão e acurácia do produtor e do usuário do sensor MSI.

	Amostras MSI					Acurácia Usuário
	Vegetação Densa	Vegetação Aberta	Área Descoberta	Agricultura	Água	
V. Densa	19	1	0	1	0	0,83
V. Aberta	3	21	1	0	0	0,91
Área Descoberta	0	1	18	0	0	0,95
Agricultura	1	0	0	25	0	0,96
Água	0	0	0	0	18	1
Acurácia Produtor	0,90	0,84	0,95	0,96	1	

A classe denominada de área descoberta atingiu bons resultados de precisão do produtor e do usuário, devido ao comportamento espectral do solo desmatado em áreas extensas e área urbana que diferem das demais classes. No entanto, em ambos os sensores, ocorreu pequena confusão com a área de vegetação aberta. Espectralmente, as áreas desmatadas apresentam alto índice de reflectância, tornando-se mais expressivas que as respostas da vegetação esparsa, o que pode interferir na distinção dos alvos (Fiuza; Barros, 2009).

As classes de uso e cobertura vegetal da terra classificadas neste estudo obtiveram relativas variações em suas extensões territoriais, calculados a partir da contagem de área em km². Os gráficos da Figura 5 mostram que as diferenças correspondem às classes de vegetação e de área descoberta, entretanto, são resultados com certa similaridade, no qual a classe de maior proporção é a de vegetação aberta, seguida de vegetação densa e área descoberta.

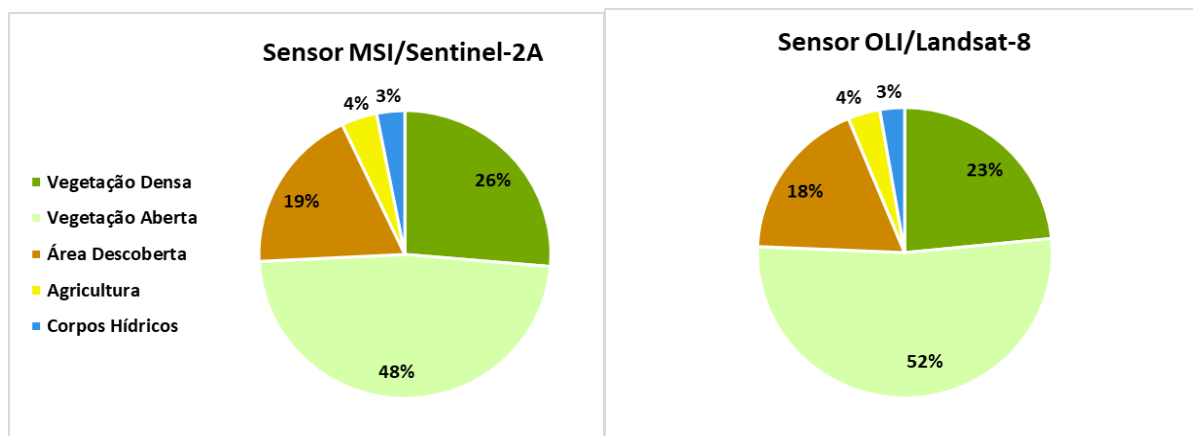


Figura 5. Gráfico das métricas das classes em porcentagem.

As áreas de vegetação denominadas abertas ou esparsas intercalam com afloramento rochoso, vegetação rala, muitas vezes, com cortes de solo exposto. Essa característica favorece a atividade pecuarista e a exploração da madeira. São áreas aplainadas onde as práticas humanas impulsionam o desgaste vegetal, expondo o solo. Ambos os sensores revelaram a maior taxa de cobertura dessa vegetação, sendo maior que 50% no sensor OLI.

A resolução espacial de 10 metros do MSI apresenta maior identificação dos alvos em ambientes heterogêneos. Locais com rochas

expostas e faixas de vegetação foram mais facilmente identificadas no respectivo sensor, como mostra a Figura 6. Ademais, locais de cultivo temporário, em que a terra é preparada para o plantio e toda a vegetação é removida, o MSI foi mais sensível (Figura 6), corroborando com os valores percentuais postos na extensão espacial (Figura 5). Esses ambientes podem apresentar diferentes graus de degradação, proporcionando diferentes reflectâncias nas imagens de satélites, podendo ocasionar confusões na classificação (Moreira, 2012).

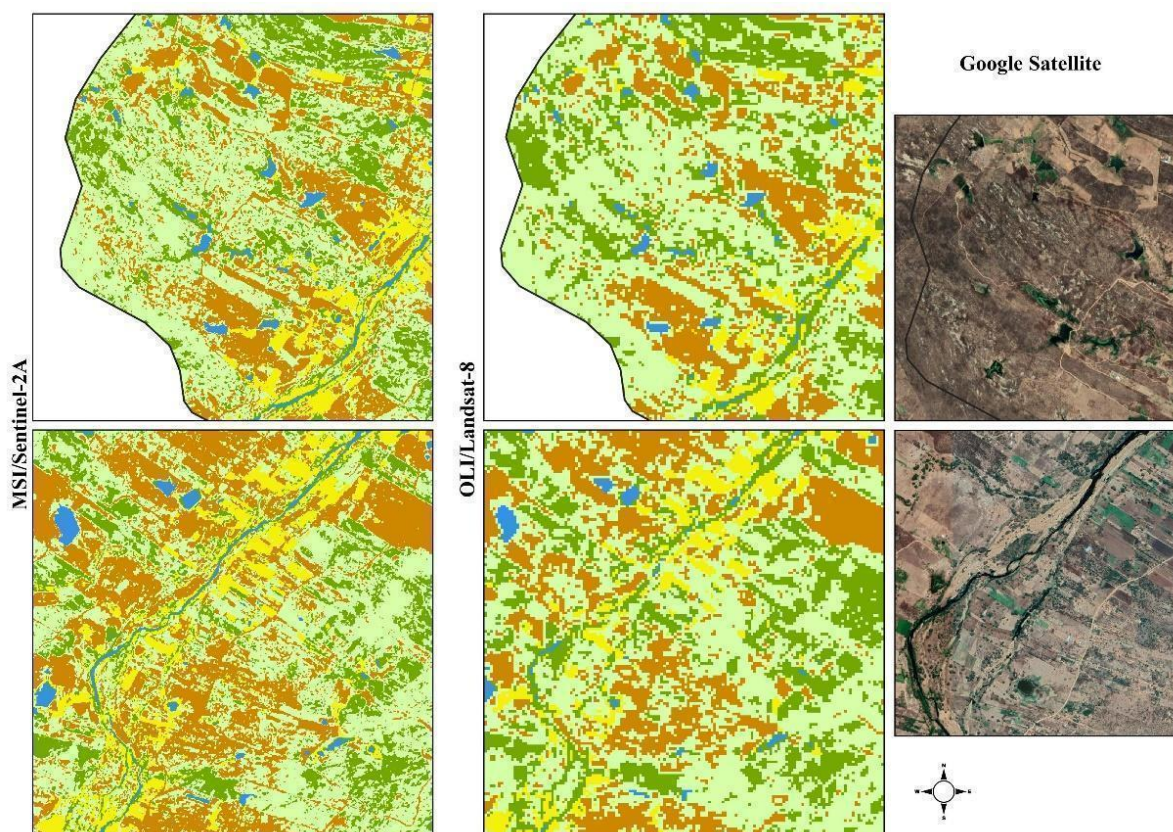


Figura 6. Diferença das classes Área Descoberta e Agricultura entre os sensores

A vegetação densa apresentou maior discrepância na parte indicada na Figura 7. O sensor MSI demonstrou maior cobertura dessa classe nessa região, ao passo que o sensor OLI foi mais homogêneo, aumentando a proporção da classe de vegetação aberta e solo exposto, sendo esta área caracterizada por uma vegetação de densidade variável, que inclui áreas desmatadas com vegetação esparsa.

Nas imagens do sensor OLI, as áreas agrícolas se apresentaram de forma discreta. Apesar disso, a classe “Agricultura” foi espacializada de maneira semelhante à classe de vegetação. Esse fator pode estar atrelado as configurações radiométricas, no qual o sensor MSI apresenta resolução radiométrica de 12 bits, enquanto o OLI possui 16 bits, isso traz maior capacidade de diferenciar os níveis de cinza,

aumentando a exatidão na detecção dos alvos. Refere-se, portanto, a capacidade do sensor de discriminação das intensidades da energia refletida pelo objeto (Florenzano, 2008).

Apesar do percentual igual dos corpos hídricos, os mapas classificados (Figura 8) mostram que o sensor OLI foi capaz de detectar água somente nas áreas mais largas do leito do rio principal Piranhas-Açu, muitas vezes classificando como vegetação densa. Nos demais corpos hídricos, ambos os sensores foram eficientes. No entanto, a água em alguns locais foi confundida com a resposta espectral da área de agricultura, provavelmente pela presença de partículas orgânicas, inorgânicas, fitoplâncton ou macrófitas. A água turva pode afetar a resposta do sensor, pois a transparência da água resulta em alta absorção e baixa reflectância (Novo, 2008; Moreira, 2012).

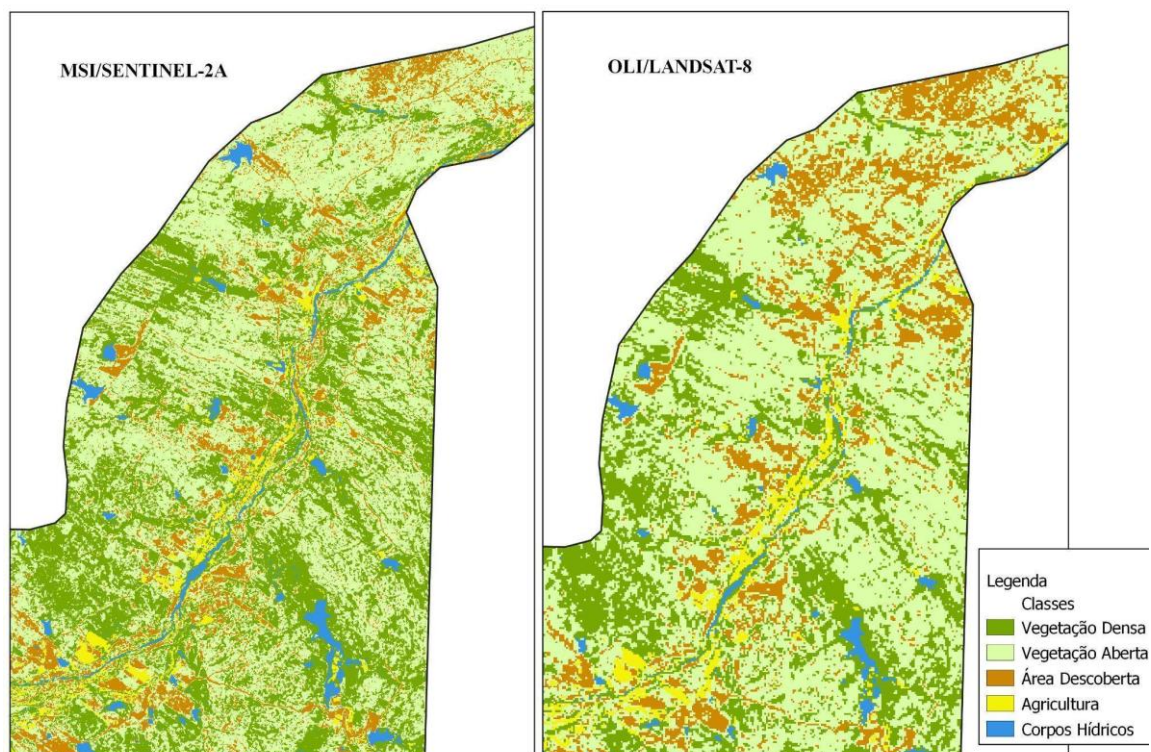


Figura 7. Distinções nas classes Vegetação Densa e Área Descoberta entre os sensores.

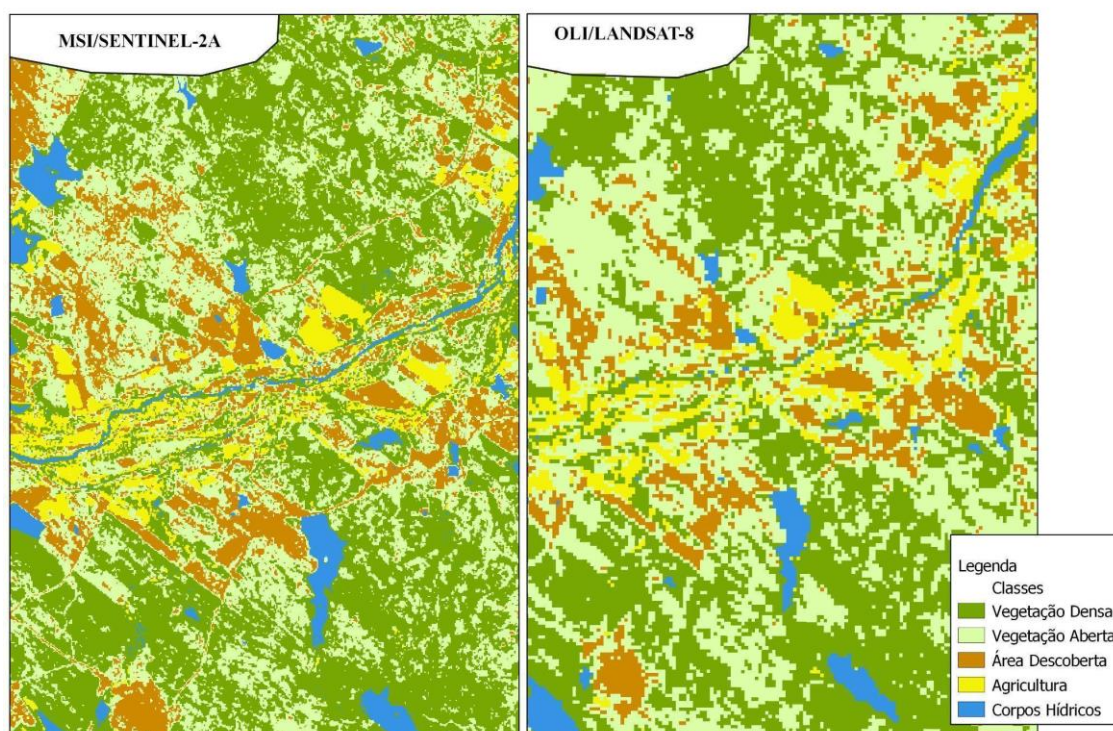


Figura 8. Diferença da classe Corpos Hídricos entre os sensores.

Diversos estudos exploram as respostas dos sensores OLI/Landsat-8 e MSI/Sentinel para análise ambiental no semiárido. A exemplo da pesquisa feita por Freires *et al.* (2019), que comparou ambos os sensores para mapeamento de uso e cobertura do solo no Maciço de Uruburetama, Ceará, concluindo que o MSI permite maior detalhamento dos alvos e maior acurácia na classificação. Machado (2020) realizou o mapeamento de uso e cobertura da terra da Ilha do Maranhão e diferenças espaciais nas classes nos produtos, apesar da acurácia ser satisfatória, concluindo que o MSI/Sentinel tem maior nível de detalhe.

4 Conclusões

A plataforma Google Earth Engine demonstrou agilidade nas etapas de processamento de imagens de satélite. A possibilidade de criar rotinas, sem necessitar do uso de computador de alta performance, torna-se um potencial para ser utilizado em grandes áreas e em outros modelos de machine learning. O algoritmo Random Forest foi eficiente na identificação de padrões ambientais, incluindo áreas de Caatinga, atividades agrícolas e solos expostos.

Ao comparar a aplicação dos sensores OLI/Landsat-8 e MSI/Sentinel-2A nos produtos do índice de vegetação NDVI, verificou-se uma alta correlação ($r = 0.89$) e valores estatísticos aproximados, no entanto, o gráfico histograma e a

separação das classes conseguiu revelar nuances espaciais que diferem os dois sensores, estando ligados, principalmente, pela diferença da resolução espacial dos equipamentos.

As classificações de uso e cobertura vegetal da terra gerados para o município de Jardim de Piranhas/RN, a partir de bandas espectrais semelhantes e os NDVIs, obtidos das imagens do OLI e do MSI, apresentaram resultados satisfatórios: a exatidão global foi de 89% para o OLI e 92% para o MSI e índice Kappa foi de 0.86/OLI e 0.90/MSI. As matrizes de confusão obtiveram desempenho variado, com baixa acurácia do usuário para "Vegetação Densa" no OLI (70%) e maior precisão no MSI (83%), podendo estar relacionado com as dificuldades em distinguir vegetação em áreas densas e abertas, assim afetando a acurácia na resolução de 30m.

O MSI apresentou resultados mais precisos para vegetação esparsa e densa. A precisão da classificação na categoria "Área Descoberta" foi notavelmente elevada, sendo atribuída ao comportamento espectral do solo desmatado. No entanto, houve dificuldades na distinção da classe "Vegetação Aberta" devido à considerável heterogeneidade em determinadas áreas desprovidas de cobertura vegetal e com exposição de rochas. É relevante destacar que essa classe prevalece em ambos os sensores, correspondendo a ambientes propícios para atividades agropecuárias.

Quantificar a estrutura da vegetação por meio de métricas foi fundamental para comparar os

mapeamentos nos dois sensores, podendo-se concluir que o nível de detalhe das imagens do sensor MSI foi melhor para delinear e classificar os alvos em escala maior. Embora ambos os sensores possuam diferenças na resolução espectral e radiométrica, principalmente, o sensor OLI demonstrou significativo potencial na identificação de áreas homogêneas.

A metodologia aplicada e os resultados obtidos podem contribuir para o desenvolvimento de políticas públicas de planejamento ambiental do município de Jardim de Piranhas e a mesma pode ser reproduzida para outras áreas do semiárido brasileiro. O estudo de dados espaciais através de geotecnologias, torna-se fundamental para o ordenamento territorial, visando o equilíbrio da proteção dos ecossistemas, a segurança hídrica e o desenvolvimento urbano, na perspectiva de um mundo de incertezas socioambientais a partir das mudanças climáticas globais em curso.

Agradecimentos

As autoras agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo a pesquisa e concessão de bolsa de Pós-Graduação, ao Laboratório de Hidrografia, Climatologia e Cartografia e ao Laboratório Didático de Geociências do Centro de Ensino Superior do Seridó - Universidade Federal do Rio Grande do Norte, pelo apoio e disponibilidade de espaço e equipamentos para o desenvolvimento da pesquisa.

Referências

- Amani, M. *et al.*, 2020 Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE journal of selected topics in applied earth observations and remote sensing*, v. 13, p. 5326–5350.
- Breiman, L., 2001 Random forests. *Machine Learning*, Boston, 45, p. 5-32
- Congalton, R. G. 1991A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46.
- Congalton, R. G.; Green, K, 1999. *Assessing the Accuracy of Remotely Sensed Data: principles and practices*. New York, Lewis Publishers. 137.
- Chang, N.-B.; Bai, K, 2018. *Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing*. [s.l.] CRC Press.
- Chueh, A. M, 2005 *Análise do potencial de degradação ambiental na bacia hidrográfica do Rio Pequeno em São José dos Pinhais/PR, por meio do DFC - Diagnóstico Físico Conservacionista*. Dissertação de Mestrado-UFPR, Curitiba, 2005 (p. 61-71).
- Drusch, M. *et al.*, 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, v. 120, p. 25–36.
- ESA. AGÊNCIA ESPACIAL EUROPEIA 2023. *Introduzindo o Sentinel-2*. Disponível: <https://earth.esa.int/eogateway/missions>. Acesso em: 05 mai. 2023.
- Florenzano, T. G, 2002. *Imagens de satellite para estudos ambientais*. Oficina de textos.
- Freires, E. V. *et al.*, 2019. Comparison of OLI/Landsat 8 and MSI/Sentinel-2 Images in Cover and Land Use Mapping in the Uruburetama Massif, Ceará. *Anuário do Instituto de Geociências - UFRJ*, 42, 427–442, 12 dez.
- Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. 2016. *Google Earth Engine: Planetary-scale geospatial analysis for everyone*. *Remote Sensing of Environment*, <http://dx.doi.org/10.1016/j.rse.2017.06.031>
- IPCC 2021- Painel Intergovernamental sobre Mudanças Climáticas. *Mudança do Clima: A base científica*. Tradução: OLIVEIRA, Mariane Arantes Rocha de. IPCC/WMO/UNEP, 2021.
- Landis, J.R.; Koch, G.G., 1977The measurement of observer agreement for categorical data. *Biometrics*, s.l, v. 33, 159-174.
- Lorenzetti, J. A, 2015. *Princípios físicos de sensoriamento remoto*. São Paulo: Blucher, 2015.
- Machado, A. M. B, 2020. Comparação de Imagens OLI/LANDSAT-8 e MSI/SENTINEL-2 no Mapeamento de Uso e Cobertura da TERRA na Ilha do Maranhão. *Revista Equador*, v. 9, 209–232, 2 jul.
- Moreira, M. A, 2012. *Fundamentos do Sensoriamento Remoto e Metodologias de Aplicação*. 4 ed. Viçosa: UFV.
- Mukaka, M.M. 2012. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3): 69-71.
- Mutanga, O.; Kumar, L, 2019 *Google Earth Engine applications*. *Remote sensing*, 11, 591.
- Naushad, R.; Kaur, T.; Ghaderpour, E. 2021. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. *Sensors*, 21, 8083, 3.
- Novo, E. M. L. M. 2008. *Sensoriamento remoto: princípios e aplicações*. 3 ed. São Paulo: Edgard Blucher. 2008

- Pereira, Paulo Emanuel Batista 2021. Diagnóstico de impacto ambiental no trecho do Rio Piancó-Piranhas- Açú em Jardim de Piranhas -RN. Orientador: Walker Gomes de Albuquerque. 2021. 101 f. Monografia (Bacharelado em Engenharia Ambiental) - Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia Agroalimentar, [S. l.], 2021
- RefatI, D. C. *et al.*, 2023. Influence of drought and anthropogenic pressures on land use and land cover change in the brazilian semiarid region. *Journal of South American Earth Sciences*, v. 126, 104362, 1.
- Rouse, J. W.; Haas, R. H.; Schell, J. A.; Deering, D. W.; Harlan, J. C, 1974. Monitoring the vernal advancement of retrogradation (greenwave effect) of natural vegetation. *Greenbelt: NASA*, 371 p., 1974.
- Santos, S. de L. *et al.*, 2019. Prejuízos e impactos ambientais do Rio Piranhas em Jardim de Piranhas-RN: Preservação e conservação deste afluente. Em: *Semiárido Brasileiro – Volume 3*. Belo Horizonte/MG: Editora Poisson. p. 122–125.
- Santos, A. B. A., 2023. Impactos ambientais da cadeia produtiva da indústria têxtil em Jardim de Piranhas/RN. Orientador: Edna Lucia da Rocha Linhares. 2021. 51 f. Monografia (Graduação) - Curso de Ciência e Tecnologia, Universidade Federal Rural do Semi-árido, 2021. Disponível: https://repositorio.ufersa.edu.br/bitstream/prefix/9445/1/AndressaBAS_MONO.pdf. acesso em: 5 jul.
- Shwartzshalev, S.; Bendavid, S, 2014. *Understanding Machine Learning: From theory to algorithms*. NY, USA: Cambridge University Press.
- Silva, J. M. C., Barbosa, L. C. F., 2017. Impact of Human Activities on the Caatinga. In: Silva, J. M. C.; Leal, I. R.; Tabarelli, M. *Caatinga: the Largest Tropical Dry Forest Region in South America*. Cham, Switzerland: Springer Publishing International, 359-368. Disponível: 10.1007/978-3-319-68339-3_13. Acesso: 02 mai. 2023.
- Sun, A. Y.; Scanlon, B. R. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters*, v. 14073001, 3 jul. 2019
- USGS, 2023. United States Geological Survey. Disponível: <https://www.usgs.gov/landsat-missions>. Acesso: 05 mai. 2023.
- Vale, T. M. C. *et al.*, 2020. Subsistence Agriculture Productivity and Climate Extreme Events. *Atmosphere*, v. 11, n. 12, p. 1287, 2020. DOI: <https://doi.org/10.3390/atmos11121287>.
- Yang, J.; Gong, P.; FU, R.; Zhang, M.; Chen, J.; Liang, S.; Xu, B.; SHI, J.; Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. *Nat. Clim. Chang.* 3, 875–883. Doi: <https://doi.org/10.1038/nclimate1908>
- Zhang, F. Yang, X. 2020. Improving land cover classification in an urbanized coastal area by Random forests: The role of variable selection. *Remote Sensing of Environment*. s.l, v. 251 n. 1, p. 112105. 2020. Doi: <https://doi.org/10.1016/j.rse.2020.112105>.