

## MAPEAMENTO DE ÁREAS DE CAATINGA ATRAVÉS DO *RANDOM FOREST*: ESTUDO DE CASO NA BACIA DO RIO TAPEROÁ

### MAPPING OF AREAS OF CAATINGA FOREST THROUGH *RANDOM*: CASE STUDY IN THE TAPEROÁ RIVER BASIN

Phillipe Valente Cardoso<sup>1</sup>

Vinicius da Silva Seabra<sup>2</sup>

Rafael Albuquerque Xavier<sup>3</sup>

Erimáigna de Morais Rodrigues<sup>4</sup>

Aureliana Santos Gomes<sup>5</sup>

#### RESUMO

A caatinga é constituída por um grande mosaico de arbusto e florestas suscetíveis às alterações sazonais características do bioma. Esse fato, constitui uma das dificuldades no mapeamento do bioma em médias e grandes escalas. Com avanço metodológico nas técnicas de sensoriamento remoto, algoritmos têm surgido em auxílio ao mapeamento e monitoramento utilizando séries temporais. Dessa forma, o presente artigo tem o objetivo de fazer o mapeamento da cobertura de caatinga para a bacia hidrográfica do rio Taperoá através do algoritmo *random forest*. Como metodologia foi utilizada a coleção Landsat 8 e através da plataforma *Google Earth Engine* foi possível fazer a seleção de imagens para o ano de 2020, implementar a classificação e avaliá-la. Os resultados mostram uma eficácia na utilização do algoritmo, assim como 79% no acerto global do mapeamento. Quanto à vegetação destaca-se que 62% da área da bacia é composta por caatinga e 28% composta por áreas de caatinga degradada (antropizada). O artigo ainda visa contribuir metodologicamente no mapeamento da vegetação de caatinga.

**Palavras-chave:** *Random Forrest* ; *Google Earth Engine* ; Mapeamento da Caatinga .

#### ABSTRACT

<sup>1</sup>Professor Assistente, Departamento de Geografia /Faculdade de Formação de Professores/Universidade do Estado do Rio de Janeiro. [phillipevalente@gmail.com](mailto:phillipevalente@gmail.com) ORCID: <https://orcid.org/0000-0002-2850-6033>

<sup>2</sup>Professor Associado, Departamento de Geografia / Faculdade de Formação de Professores / Universidade do Estado do Rio de Janeiro. [vinegeobr@yahoo.com](mailto:vinegeobr@yahoo.com) ORCID: <https://orcid.org/0000-0003-1294-6015>

<sup>3</sup>Professor Associado, Departamento de Geografia/Centro de Educação/Universidade Estadual Paraíba. [xavier@ceduc.uepb.edu.br](mailto:xavier@ceduc.uepb.edu.br) ORCID: <http://orcid.org/0000-0002-1737-7547>

<sup>4</sup>Doutoranda, Programa de Pós-Graduação em Ecologia/Instituto de Biologia/Universidade Federal do Rio de Janeiro. [erimagnarodrigues@ufrj.br](mailto:erimagnarodrigues@ufrj.br) ORCID: <https://orcid.org/0000-0003-4281-3555>

<sup>5</sup>Graduanda, Unidade Acadêmica de Geografia / Centro de Humanidades/ Universidade Federal de Campina Grande. [aurelianagomes7@gmail.com](mailto:aurelianagomes7@gmail.com) ORCID: <https://orcid.org/0000-0002-9385-0842>

The caatinga consists of a large mosaic of shrubs and forests susceptible to seasonal changes characteristic of the biome. This fact constitutes one of the difficulties in mapping the biome at medium and large scales. With methodological advances in remote sensing techniques, algorithms have emerged to aid mapping and monitoring using time series. Thus, the present article aims to map the caatinga cover for the hydrographic basin of the Taperoá River through the random forrest algorithm. As a methodology, the Landsat 8 collection was used and through the Google Earth Engine platform it was possible to make the selection of images for the year 2020, implement the classification and evaluate it. The results show an efficiency in the use of the algorithm, as well as 79% in the global correctness of the mapping. As for vegetation, 62% of the basin's area is made up of caatinga and 28% is made up of degraded (anthropized) caatinga areas. The article also aims to contribute methodologically in the mapping of the caatinga vegetation.

**Keywords:** *Random Forrest 1; Google Earth Engine 2; Mapping of the caatinga 3.*

## INTRODUÇÃO

As áreas semiáridas do nordeste brasileiro tornaram-se protagonistas de discussões que vêm se renovando nas últimas décadas. O presente fato é atestado por uma reestruturação de ideias em torno da importância do bioma Caatinga e na busca de estratégias dentro de um Plano Nacional para a manutenção da região no que diz respeito à sociedade e natureza (MACIEL; PONTES, 2015).

A região do semiárido nordestino brasileiro está inserida em vários estados brasileiros, tendo como um dos destaques o estado da Paraíba, com aproximadamente 86,5% do seu território inserido nessa realidade. A fragilidade natural dessa região, ressaltada por Alves *et al.* (2009), está associada a práticas agrícolas inadequadas, queimadas e desmatamento. Esses são os principais fatores que têm contribuído para o avançado processo de erosão e do assoreamento de rios e açudes.

Compondo a região, o bioma caatinga constitui um grande mosaico de arbustos espinhosos e florestas sazonalmente secas englobando os estados do Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe e o norte de Minas Gerais (LEAL, *et al.*, 2005). Abrangendo cerca de 735.000 km<sup>2</sup>, ou aproximadamente 11% do território nacional, é dividido em duas regiões, o Agreste e o Sertão.

Tendo características únicas, o bioma Caatinga tem em sua vegetação uma forte influência do regime de chuvas na região. Desta forma, basta apenas poucos milímetros cúbicos de chuva para que a vegetação passe de um estado seco, com ausência total de folhas, para um estado no qual o

verde ganha um grande destaque na paisagem. Cabe também ressaltar a importância das áreas de fundo de vale e mata ciliar dos rios e açudes, que atuam como retentores de umidade e, por isso, também apresentam uma dinâmica da vegetação diferenciada na região.

Ainda sobre a vegetação, Loefdren (1910), em suas observações divide as espécies xerófilas nordestinas em permanentes e periódicas. As primeiras são adaptadas às condições de umidade, ou seja, tropófitas. Essas, segundo o autor, são xerófitas nas épocas secas e mesófitos nas épocas chuvosas, são caracterizadas por árvores e arbustos de porte considerável. Já as segundas só se desenvolvem na presença de água, sendo exclusivamente mesófitos. Em geral são herbáceas, apresentando também substâncias sublenhosas. Uma das características dessas espécies é a velocidade de rebrotação referente a um regime de chuvas. Com isso, a dinâmica desse bioma lança desafios para o mapeamento de suas paisagens e até mesmo diferenciação de suas espécies vegetais em escalas de maior abrangência.

Em decorrência disso, busca-se a incorporação crescente do uso de geotecnologias nos estudos do bioma, a capacidade de desenvolver análises mais consistentes se dá pela necessidade de trabalhar todo esse conjunto de dados, de naturezas variadas, sobre uma perspectiva integradora.

O sensoriamento remoto por sua vez, como parte dessas geotecnologias, pode ser entendido como a ciência, técnica e processo que permite a aquisição de determinado tipo de informação através de sensores de forma remota. Seu avanço tecnológico possibilitou o desenvolvimento de sensores com maiores resoluções espaciais, espectrais, radiométricas e temporais, além do surgimento de diferentes softwares de processamento de imagem e métodos de classificação. Somado a isso, veio a popularização das imagens de satélites e como consequência a possível aquisição de diversos desses produtos de forma gratuita para todo território nacional.

Segundo Stow (1993), a maioria dos estudos geoecológicos faz uso das imagens de sensoriamento remoto para identificar e representar a distribuição de suas estruturas e funções em diferentes escalas temporais e espaciais. Nesse contexto, surge o aprimoramento da interpretação das imagens orbitais, assim como a análise de séries temporais que com o aumento do processamento computacional agregam mais a qualidade das análises. (LACERDA e VICEZ, 2021).

Face o exposto, o presente artigo tem como objetivo fazer o mapeamento da vegetação de caatinga utilizando o algoritmo de classificação *Random Forrest* através do *Google Earth Engine* e tendo como área de estudo a bacia do rio Taperoá localizada no estado da Paraíba.

### **BACIA DO RIO TAPEROÁ**

A bacia está integralmente inserida dentro do Bioma da Caatinga e engloba 29 municípios, corresponde a uma área de aproximadamente 5.882 km<sup>2</sup>, onde 4 municípios estão totalmente e 21 parcialmente inseridos em seu limite (Figura 1). Os municípios de São João do Cariri e Taperoá, apesar de estarem parcialmente inseridos devido a sua extensão territorial, ocupam 19% da área da bacia. Já outros municípios, localizados nos divisores de água, ocupam menos de 1% do território da bacia, são eles: Areia de Baraúnas, Barra de Santa Rosa, Boqueirão, Campina Grande, Cubati, Salgadinho, São Domingos do Cariri, São Vicente do Seridó, Sumaré e Teixeira.

O rio principal é o que dá nome a bacia Taperoá, de regime intermitente, com nascente na Serra do Teixeira, desaguando no rio Paraíba. Sendo considerado de grande importância estratégica para a gestão de recursos hídricos no Estado da Paraíba. Além disso, abastece o açude Epitácio Pessoa que teve sua obra concluída em 1956 e abastece 8 municípios da mesorregião da Borborema e 6 do Agreste Paraibano, incluindo a segunda maior cidade do estado Campina Grande, através das adutoras do Cariri e do Boqueirão. (ALMEIDA, 2012). Tendo capacidade inicial prevista para 534.680.00 m<sup>3</sup> de água, esse reservatório vem sofrendo um grave processo de assoreamento em virtude das erosões do solo causadas pelos desmatamentos na região. Britto e Vianna (2006) alertam para a diminuição de sua capacidade máxima que passa a ser de 411.686.287 m<sup>3</sup> em 2004. É importante destacar também que os rios da BHRT abastecem 11 reservatórios distribuídos ao longo de toda a bacia.

Almeida (2012) destaca que apesar do grau de importância hídrica dessa bacia, ela vem sendo protagonista de uma série de problemas ambientais como desmatamento, poluição atmosférica, hídrica e do solo, erosão, escassez de recursos naturais na região, levando a diversos problemas sociais, e ausência de políticas públicas que visam a mitigação desses problemas. A autora em seu trabalho destaca três pontos chaves de conflito na BHRT: 1) a conservação dos



consideração os fatores relacionados à vegetação e que também atendesse uma escala representativa para a bacia de forma a escolher legenda compatível.

Para isso foi escolhida as imagens do Landsat 8 disponibilizadas através da plataforma *Google Earth Engine* (GEE), que é uma plataforma computacional que permite através da infraestrutura da Google análises geoespaciais, além da disponibilidade de pentabytes de dados para utilização. A plataforma possui um ambiente de desenvolvimento integrado (IDE) baseado na web que possibilita o desenvolvimento e execução de *scripts*. Sendo assim, a coleção de imagens escolhida para o mapeamento foi a *Landsat 8 Collection 1 Tier 1 calibrated top-of-atmosphere (TOA) reflectance* por conta da correção radiometria já nativa.

A partir das classes de mapeamento foram determinadas:

Caatinga verde: caatinga arbórea e arbustiva com grande densidade, com presença de folhas, podendo apresentar atividades agropastoris não identificáveis na escala de mapeamento.

Caatinga seca: caatinga arbórea e arbustiva com grande densidade, em sua maioria seca, podendo apresentar atividades agropastoris não identificáveis na escala de mapeamento.

Caatinga degradada: caatinga com grande presença de exposição solo, em geral apresentando características de atividades agropastoris e/ou mineração passadas ou recentes podendo ser perceptíveis em poucos casos em virtude da escala de mapeamento.

Algaroba e caatinga: Áreas contendo a espécie exótica algaroba e vegetação nativa de caatinga

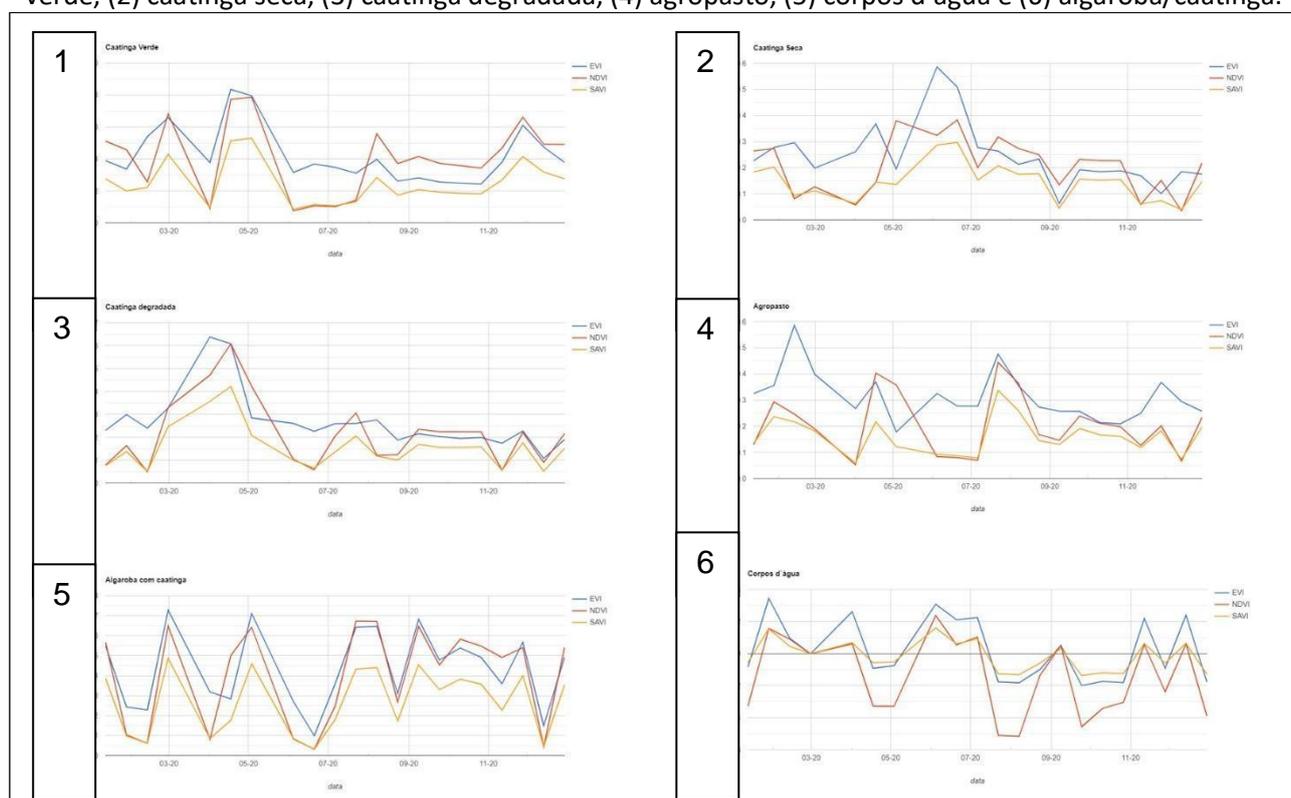
Agropasto: atividades agropastoris identificáveis nas imagens Landsat 8.

Corpos d'água: açudes naturais e artificiais presentes na bacia.

Em seguida, a coleção de imagens passou por uma filtragem e foram consideradas apenas as imagens para o ano de 2020, para em seguida aplicar uma filtragem para minimizar o efeito de nuvens na região. Como próxima etapa foram calculados os índices de vegetação: *Normalized Difference Vegetation Index (NDVI)*, *Soil-Adjusted Vegetation Index (SAVI)* e *Enhanced Vegetation Index (EVI)* e logo após foi gerada uma imagem com a média de todos os pixels da coleção. A figura 2 apresenta o comportamento das classes escolhidas ao longo do ano de 2020, os índices apontam os maiores valores no período 05/2020 a 07/2020 para as classes caatinga verde e caatinga seca.

Na classe caatinga degradada podemos observar valores mais baixos e poucas flutuações ao longo do ano. A classe algaroba/caatinga mostra valores altos com algumas quedas ao longo do ano, mostrando um comportamento bem diferente das classes de caatinga verde, seca e degradada. Quanto aos corpos d'água, a classe apresenta valores baixos ou negativos ficando próximos a 0, como já se era esperado. Por fim, a classe agropasto também parece acompanhar o período úmido e seco da bacia, seguindo o regime de chuvas.

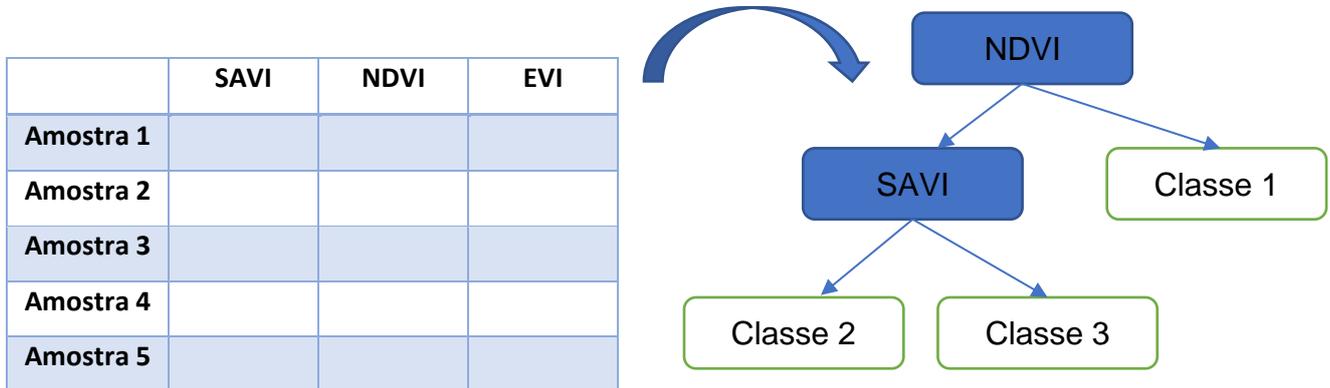
Figura 2 - Índices de vegetação NDVI, EVI e SAVI das classes mapeadas na bacia do rio Taperoá. (1) caatinga verde, (2) caatinga seca, (3) caatinga degradada, (4) agropasto, (5) corpos d'água e (6) algaroba/caatinga.



Fonte: Elaborado pelos autores.

Por fim foi utilizado o algoritmo *Random Forest* implementado no próprio GEE. Ele estabelece regras para a tomada de decisão na classificação, estabelecendo “nós” que, após ter a condição verificada, segue o fluxo através de um ramo direito ou esquerdo, levando sempre a outro nó até a finalização da árvore (Figura 3).

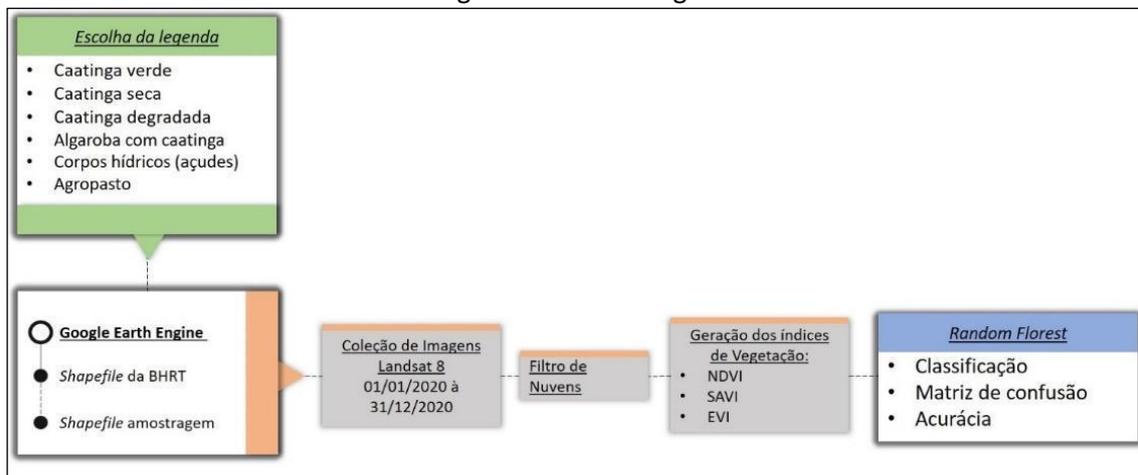
Figura 3 - Árvore de Decisão.



Fonte: Elaborado pelos autores.

No *Random Forest* a variável no primeiro nó acontece de forma aleatória, escolhendo de uma ou mais variáveis. Após realizar as análises com base na amostragem é definida a variável que irá compor o primeiro nó, esse processo irá se repetir nos nós subsequentes excluindo as variáveis já selecionadas nos nós anteriores. A quantidade de árvores utilizadas no processo de classificação é definida no projeto, para esse foram utilizadas 400 árvores. Assim, após alguns testes, essas árvores são analisadas e agrupadas, selecionando as melhores variáveis na separação das classes do mapeamento. Após a classificação, as matrizes de erro são geradas e o acerto global da classificação também é baseado na amostragem. A figura 4 apresenta o fluxograma de atividades.

Figura 4 - Metodologia.



Fonte: Elaborado pelos autores.

## RESULTADOS E DISCUSSÕES

Os resultados da matriz de erro e da acurácia da classificação são apresentados no console do GEE (Figura 5). A listagem de elementos da matriz refere-se ao id das classes. Ela foi transcrita para o quadro 1, já considerando as respectivas classes. A matriz apresenta a quantidade de acertos e erros referentes aos pontos de validação, além de mostrar quais foram as classes que apresentaram maior erro.

Figura 5 - Resultados apresentados no console do GEE.

```

Inspector Console Tasks
Use print(...) to write to this console.

▶ List (378 elements) JSON

Matriz de Erro - RF1 JSON
▼ List (7 elements) JSON
  ▶ 0: [0,0,0,0,0,0,0]
  ▶ 1: [0,30,6,4,0,0,0]
  ▶ 2: [0,3,6,0,0,1,1]
  ▶ 3: [0,1,0,13,0,1,0]
  ▶ 4: [0,1,1,0,17,0,0]
  ▶ 5: [0,0,4,0,0,15,1]
  ▶ 6: [0,0,0,0,0,2,17]

Acuracia - RF1 JSON
0.7903225806451613
    
```

Fonte: Elaborado pelos autores.

Quadro 1 - Matriz de erro da classificação

	Caatinga Verde	Caatinga Seca	Caatinga Degradada	Algaroba/Caatinga	Agropasto	Corpos D'água	Total	% de acerto da classe
Caatinga Verde	30	6		4			40	75%
Caatinga Seca	3	6	1		1		11	55%
Caatinga Degradada		4	15		1		20	75%
Algaroba/Caatinga	1		1	13			15	87%
Agropasto			2		17		19	89%
Corpos D'água	1	1				17	19	89%

Fonte: Elaborado pelos autores.

Ao fazer uma leitura do respectivo quadro, observa-se que a classe caatinga verde utilizou 40 pontos de validação, atingindo um total de 75% de acerto (30 pontos de amostragem), apresentando também 6 pontos de validação para a classe caatinga seca e 4 pontos para a classe algaroba/caatinga. Já a classe caatinga seca tem um total de 11 pontos selecionados, atingindo 55% de acerto na classe (6 pontos), o restante é representado nas classes caatinga verde e agropasto, respectivamente com 3 e 1 pontos. Esse erro apresentado entre a caatinga verde e caatinga seca não tem grande significância para esse mapeamento, já que somatório dessas classes representam o total de área de caatinga na bacia, vide o quadro 2.

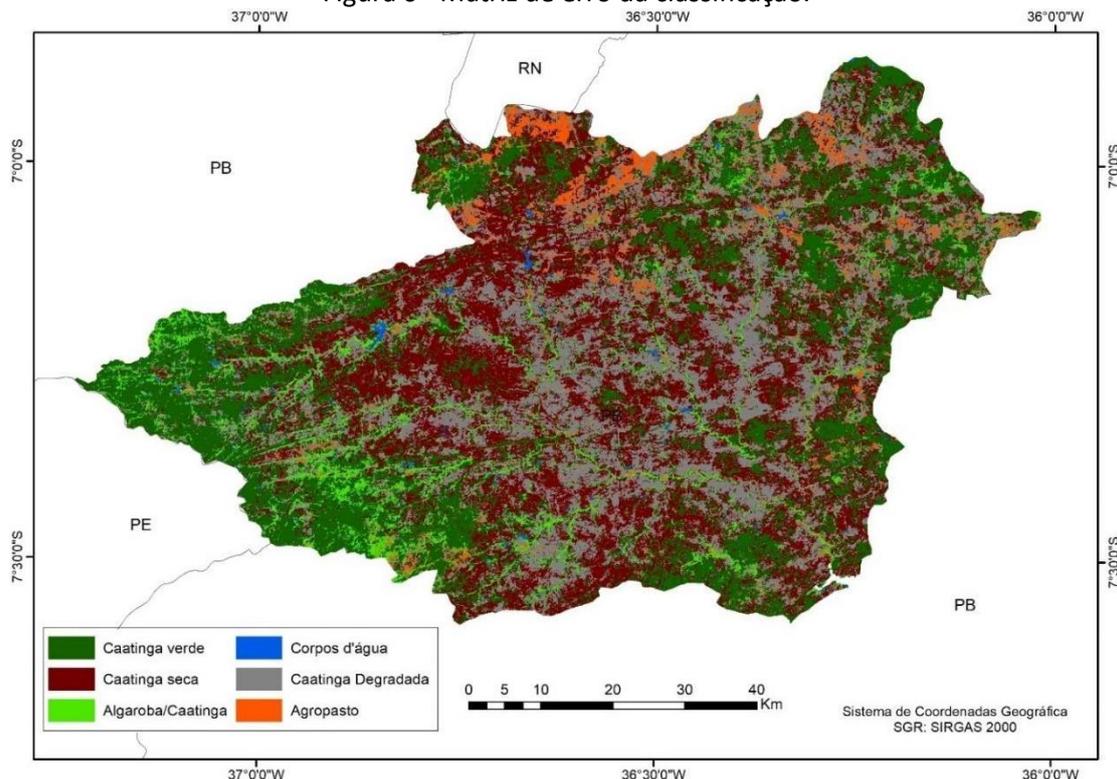
Nas áreas de algaroba/caatinga foram encontrados 87% de acerto (13 pontos), em um total de 15 pontos avaliados. Já a caatinga degradada apresentou 75% de acerto (15 pontos), sendo o maior desencontro na validação relacionado a classe caatinga seca. As classes agropasto e água apresentaram ambos 89% de acerto (17 pontos) de um total de 19 utilizados.

Cabe lembrar que a imagem utilizada no mapeamento é uma média de bandas da coleção do satélite Landsat 8 para todo período de 2020. As amostragens, apesar de terem seguidos critérios como dados de campo e interpretação visual, ainda podem sofrer com a dinâmica da região. Isso porque a média resultante, após a seleção de datas e filtragem, pode ter mascarado valores de pixel

extremos. Por exemplo, um açude pode estar seco na maioria do ano e sendo selecionado na amostragem, podendo gerar conflitos na modelagem e, conseqüentemente, na validação.

Quanto ao produto do mapeamento (Figura 6), a classe caatinga verde e caatinga seca apresentam respectivamente 30% e 32% do total da área de bacia, ou seja, 62% de área de caatinga. Quanto a sua distribuição, elas ficam concentradas em áreas pontuais, com uma concentração menor e uma grande fragmentação a leste e norte da bacia e com uma concentração maior e baixa fragmentação a oeste da bacia.

Figura 6 - Matriz de erro da classificação.



Fonte: Elaborado pelos autores.

Quadro 2 - Matriz de erro da classificação

		Área em Km <sup>2</sup>	%
Caatinga	Caatinga Verde	1734,69	30%
	Caatinga Seca	1893,87	32%
	Caatinga Degradada	1625,02	28%
	Algaroba/Caatinga	384,85	7%
	Agropasto	186,98	3%
	Corpos D'água	52,59	1%
	<b>Total</b>	<b>5877,99</b>	<b>100%</b>

Fonte: Elaborado pelos autores.

A classe caatinga degradada compõe 28% da área da bacia, ficando bem concentrada na região centro-sul. Esse alto percentual é preocupante, pois essas são áreas que sofrem ou que já tiveram uma forte antropização por conta das atividades e ciclos econômicos da região. Desta forma, a vegetação bastante antropizada em grandes áreas de solo exposto, podem ainda ser indicativos de um embrionário processo de desertificação na região, apesar disso alguns outros fatores devam ser considerados dentro desse fenômeno complexo.

As áreas referentes a classe agropasto representam 3% da bacia, cabendo lembrar que essa classe é composta apenas por áreas onde se é possível identificar a atividade. As outras classes, exceto a classe água, tem um uso misto podendo conter outras atividades, como foi descrito na metodologia. Quanto a concentração a classe fica localizada em sua maioria na região norte e nordeste da bacia.

Seguindo a trajetória dos rios da bacia, a classe algaroba/caatinga estende-se por toda a bacia, compondo 7% da área. E por fim, a classe corpos d'água, composta em sua maioria por açudes, representam 1% da bacia. Sua concentração fica a norte e a noroeste da bacia. Por fim, a acurácia do mapeamento ficou em 79%.

### CONSIDERAÇÕES FINAIS

Como uma das principais abordagens no presente mapeamento, cabe destaque a aprendizagem de máquinas através do algoritmo *Random Forest* implementado na plataforma

GEE. As árvores de decisão do algoritmo, executadas na plataforma, que além de um processamento em nuvem, proporciona um banco de dados de imagens em diferentes escalas, apresentando também um ganho prático e assertivo, apesar das pequenas questões já levantadas no estudo. Com isso, além do aumento na qualidade do mapeamento, há também o ganho de tempo e a possibilidade de trabalhar com séries temporais, importantes, principalmente, no mapeamento de uma vegetação tão dinâmica no espaço e no tempo como a caatinga.

Quanto ao mapeamento, entende-se que foi satisfatório dentro do que se propôs em relação às classes mapeadas. Desta forma, a dinâmica sazonal do semiárido nordestino e, conseqüentemente, da caatinga, deve ser considerada no que diz respeito ao mapeamento da vegetação. Assim, mapeamentos que consideram uma única data de imageamento, aumentam as possibilidades de confusão e/ou omissão entre as classes de vegetação mapeadas.

## REFERÊNCIAS

ALMEIDA, N. V. **Ordenamento territorial geoambiental Da bacia Hidrográfica do rio Taperoá/Semiárido Paraibano**. [s.l.] Universidade Federal Fluminense, 2012.

BRITO, F. B. de ; VIANNA, P. C. G. . Conflito pelo uso da água do açude Epitácio Pessoa-PB. In: Seminário Luso-Brasileiro-Caboverdiano - agricultura familiar em regiões de desertificação, João Pessoa. **Anais do Seminário Luso-Brasileiro-Caboverdiano** - agricultura familiar em regiões de desertificação. João Pessoa: Editora da UFPB, 2006. v. Único. p. 86-100. 2006.

LACERDA, E. R.; VICENS R., Detecção de áreas de florestas invariantes em séries temporais utilizando random forest. **GEOgraphia Niterói**, Universidade Federal Fluminense ISSN 15177793 (eletrônico) GEOgraphia, vol: 23, n. 50, 2021.

ALVES, J. J. A.; ARAUJO, M. A.; NASCIMENTO, S. S. **Degradação da caatinga: uma investigação ecogeográfica**. Caatinga, p. 126–135, 2009.

LEAL, I. R.; SILVA, J. M. C.; TABARELLI, M.; LACHER JR, T. E. **Mudando o curso da conservação da biodiversidade na Caatinga do nordeste do Brasil**. In: Conservação Internacional do Brasil (ed). Megadiversidade. Belo Horizonte, v.1.p. 139-146, 2005.

LOEFGREN, A. **Inpetoria de obras contra as secas**, Notas botânicas (Ceará). v. 2, p. 3–39, 1910.  
MACIEL, C. A. A; PONTES, E. T. M. **Seca e convivência com o semiárido**. Adaptação ao meio e patrimonialização no Nordeste brasileiro. Rio de Janeiro: Consequência Editora, 2015.

STOW, D. A. **The role of geographical information systems for landscape ecological studies.** In: Landscape ecology and GIS. Ed. Haines-Young et al., Taylor & Francis: 11-21. 1993.