

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE CIÊNCIA FLORESTAL
CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL**

JOSÉ JORGE MONTEIRO JUNIOR

**USO DE *MACHINE LEARNING* E SENSORIAMENTO REMOTO PARA A
IDENTIFICAÇÃO DA FLORESTA TROPICAL SAZONALMENTE SECA NO
PARQUE NACIONAL DO CATIMBAU**

RECIFE-PE

2021

JOSÉ JORGE MONTEIRO JUNIOR

**USO DE *MACHINE LEARNING* E SENSORIAMENTO REMOTO PARA A
IDENTIFICAÇÃO DA FLORESTA TROPICAL SAZONALMENTE SECA NO
PARQUE NACIONAL DO CATIMBAU**

Trabalho de Conclusão de Curso apresentado
ao Departamento de Ciência Florestal -
DCFL, como parte dos requisitos necessários
à obtenção do título de Bacharel em
Engenharia Florestal.

Orientadora: Prof^ª. Dr^ª. Elisiane Alba
Coorientadora: Prof^ª. Dr^ª. Soraya G. El-Deir

RECIFE-PE

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

M775c Monteiro Junior, José Jorge
USO DE MACHINE LEARNING E SENSORIAMENTO REMOTO PARA A IDENTIFICAÇÃO
DA FLORESTA TROPICAL SAZONALMENTE SECA NO PARQUE NACIONAL DO CATIMBAU /
José Jorge Monteiro Junior. - 2021.
52 f. : il.

Orientadora: Elisiane Alba.
Coorientadora: Soraya Giovanetti El-Deir.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Engenharia Florestal, Recife, 2021.

1. Sensoriamento Remoto. 2. Caatinga. 3. Classificação de Uso e Cobertura da Terra. 4.
Inteligência Artificial . I. Alba, Elisiane, orient. II. El-Deir, Soraya Giovanetti, coorient. III. Título

CDD 634.9

JOSÉ JORGE MONTEIRO JUNIOR

**USO DE *MACHINE LEARNING* E SENSORIAMENTO REMOTO PARA A
IDENTIFICAÇÃO DA FLORESTA TROPICAL SAZONALMENTE SECA NO
PARQUE NACIONAL DO CATIMBAU**

Trabalho aprovado. RECIFE, 20 de janeiro 2021.

BANCA EXAMINADORA:

Elisiane Alba, Prof^a.Dra. (UFRPE/UAST)
(Presidenta/Orientadora)

Rodrigo Eiji Hakamada, Prof. Dr. (UFRPE/DCFL)

Hernande Pereira da Silva, Prof. Dr. (UFRPE/DEAGRI)

Rute Berger, Prof^a. Dra. (UFRPE/DCFL)
Suplente

RECIFE

2021

Dedico esta escrita à minha mãe, Glória Maria Azevedo Alves Monteiro.

AGRADECIMENTOS

A construção de um trabalho científico é a resultante de uma série de relações ao longo dos anos aos quais se foram adquiridos os saberes para compor a obra final. Nesta jornada muitos passaram e seria injusto da minha parte não mencionar estas pessoas. Portanto, agradeço a todos que passaram por minha vida e contribuíram com boas risadas, conselhos, choros, angústias, festas, almoços, caminhadas, estudos, horas e horas cantarolando, construindo, moldando, aprimorando e me proporcionando a capacidade de evoluir como ser humano e profissional que sou hoje.

Agradeço em particular:

A Deus, por me conceder companhia nos momentos (tantos momentos) onde estive só e ninguém conseguiria me entender. Me deu forças e sabedoria e me fez entender que somos parte de um projeto muito maior;

À Universidade Federal Rural de Pernambuco - UFRPE e ao Departamento de Ciência Florestal - DCFL, pelo ensino, pelo suporte, por ser um ambiente acolhedor e que me fez sentir como parte do processo criativo (de forma ativa);

À Professora Dr^a Elisiane Alba (Orientadora) pelo imenso suporte científico que envolveu a concepção, levantamento de dados, aprimoramento de técnicas, análise de dados, ensino, pesquisa, até a amizade que foi construída neste processo, estendo minha eterna admiração e gratidão;

À Professora Dr^a Soraya Giovanetti El-deir (Co-orientadora) e ao Grupo de Gestão Ambiental em Pernambuco – GAMPE/UFRPE, a contribuição técnica, estilística, psicológica, estratégica e humana, garantiram a finalização deste trabalho ao qual você está prestes a ler. Se existe esta escrita, é graças a Professora Soraya;

Aos laços de amizade e irmandade tão inesperados mas criados e estabilizados com a prova do tempo: a melhor cientista social e incrível parceira de vida Raíssa Beatriz; as parceiras *papers mode on*, Msc. Géssyca Sena e Dr^a Gabriela Salami; Beatriz Rodrigues que cansa minha beleza, mas está sempre ao meu lado me fazendo esquecer de todos os problemas com sua risada, Leonardo Modesto e Sandra Porfírio que me ajudaram tanto e nem sabem;

Aos membros do Laboratório de Sensoriamento Remoto – LASER/DCFL/UFRPE, em especial ao Professor Dr. Emanuel Silva, por investir tempo e confiança, além do ensino de técnicas de sensoriamento remoto ao longo de toda a minha graduação;

Aos meus amigos e colegas de turma que estão ao meu lado desde o início do curso, como a Vitória que foi o meu primeiro abraço dado na UFRPE e minha primeira árvore plantada, nossa filha, a Alfreda (um belíssimo Ipê que pode ser visitado, em frente ao prédio de pesca na UFRPE); à Yasmim, que estendo minha gratidão por todos os nossos momentos juntos, cada aluguel pago, cada surto, cada vitória, você é parte disso; aos terrores da OMS, Maria Eduarda Fernandes (Mariah) e Bruno Rocha, amo muito vocês; Aos grandes parceiros que construíram momentos únicos em minha memória: Evelyn Victória, Lucca S. Mousinho Móssio, Elcio Ferreira, Poliana Gomes e tantos outros.

Às minhas influências musicais que ficaram incansavelmente nos meus ouvidos enquanto eu produzia este trabalho de conclusão de curso, agradeço imensamente a Beyoncé, Lady Gaga, Pablllo Vittar, Anitta, Ariana Grande, P!nk, Adele, Little Mix, Paloma Faith, Vivaldi, Shubert, Emicida, Sunday Service Choir, Whitney Houston, Demi Lovato, Dua Lipa, Madonna, Gloria Groove, Ben Platt, Chaka Khan, Jessie Ware, Elton John e tantos outros incontáveis artistas presentes nas mais diversas e ecléticas *playlists* que fiz uso durante este processo.

Aos meus pais, Glória Maria e Jorge Monteiro, por todo investimento em minha educação ao longo dos anos, pelo suporte, carinho, desafios e compreensão por todos os meus estresses, ausências, silêncios. Muitíssimo obrigado.

Por fim, agradeço a todos mais uma vez e desejo uma boa leitura.

If you feel insignificant, you better think again Better
wake up because you're part of something way bigger
[...]

Not just a speck in the universe
Not just some words in a bible verse
You are the living word
You're part of something way bigger
Bigger than you, bigger than we
Bigger than the picture they framed us to see
But now we see it and it ain't no secret, no.

Understand the truth about that question in your soul
Look up, don't look down, then watch the answers
unfold
Life is your birthright; they hid that in the fine print
Uh, take the pen and rewrite it
Step out your estimate
Step in your essence and know that you're excel-lent
Rise, the spirit is teaching
Oh, I'm not just preaching, I'm taking my own
advice.

BEYONCÉ. **BIGGER.** Houston: Parkwood
Entertainment, 2019. Streaming (3:46).

MONTEIRO JUNIOR, JOSÉ JORGE. Uso de *machine learning* e sensoriamento remoto para a identificação da Floresta Tropical Sazonalmente Seca no Parque Nacional do Catimbau, 2021. Orientadora: Elisiane Alba, Co-orientadora: Soraya Giovanetti El-Deir.

RESUMO

A classificação de florestas tropicais sazonalmente secas é um dos maiores desafios das análises ambientais por sensoriamento remoto, tendo em vista as características fitofisionômicas da floresta que se assemelham remotamente das características do solo exposto, gerando erros amostrais em estudos de monitoramento florestal. O objetivo deste trabalho foi utilizar-se do aprendizado de máquinas para entender a dinâmica de uso e cobertura da terra no Parque Nacional do Catimbau nos períodos de maior precipitação (úmido) e menor precipitação (seco) a partir do imageamento LANDSAT. O trato metodológico ocorreu a partir da obtenção de dados landsat no ano de 2019 para o período úmido e período seco, os dados brutos foram pré processados em sistemas de informação geográficos a fim de (i) selecionar bandas; (ii) delimitar área de estudo; (iii) executar a correção atmosférica; e (iv) fazer a junção das bandas de satélite (*bandset*). Foi criado um *shapefile* para treinamento dos algoritmos de aprendizado de máquina contendo amostras das classes encontradas na área de estudo, estas sendo, as fitofisionomias Arbórea-Arbustiva e Arbustiva-Herbácea, áreas antropizadas, solo exposto, e outras áreas (i.e. nuvens, componentes hídricos, rodovias). No *software R* foram utilizados os algoritmos tanto para a classificação supervisionada (com base na validação cruzada, método *k-fold* e teste de Friedman e Nemenyi) quanto para a espacialização dos dados utilizando os algoritmos citados. Com os métodos descritos foi possível observar que os valores de NDVI fomentaram a ideia de que a fitofisionomia arbustiva-herbácea apresenta reflectância similar ao solo exposto em algumas áreas no período seco. No período úmido, o algoritmo kNN apresentou melhor performance na diferenciação das classes e identificação da vegetação (Kappa = 0,9887). Já no período seco, os algoritmos kNN, SVM e ANN não apresentaram diferenças estatísticas significativas quanto a sua performance, sendo estes considerados bons classificadores para o período (Kappa = 0,9965;0,9973;0,9962, respectivamente). Portanto, o presente estudo trouxe inovação no uso de técnicas de Inteligência Artificial para a solução de problemas no monitoramento, manejo e administração de florestas tropicais sazonalmente secas com dados remotos. Sendo um método alternativo para identificar, de forma rápida e econômica, as mudanças na estrutura florestal.

Palavras-chave: inteligência artificial, caatinga, semiárido, linguagem R.

MONTEIRO JUNIOR, JOSÉ JORGE. Using machine learning and remote sensing to identify the Seasonally Dry Tropical Forest in the Catimbau National Park, 2021. Advisor: Elisiane Alba, Co-Advisor: Soraya Giovanetti El-Deir.

ABSTRACT

The classification of seasonally dry tropical forests is one of the biggest challenges of environmental analysis by remote sensing, considering the forest physiognomic characteristics that are remotely similar to the characteristics of the exposed soil, generating sample errors in forest monitoring studies. The objective of this work was to use machine learning to understand the dynamics of land use and land cover in Catimbau National Park during periods of greater precipitation (wet) and less precipitation (dry) from LANDSAT imagery. The methodological treatment took place from the obtaining of LANDSAT data in 2019 for the wet period and dry period, the raw data were pre-processed in geographic information systems to (1st) select bands; (2nd) delimit the study area; (3rd) perform the atmospheric correction, and (4th) join the satellite bands (band set). A shapefile was created to train the machine learning algorithms containing samples of the classes found in the study area, these being the tree-shrub and shrub-herbaceous phytophysionomies, anthropized areas, exposed soil, and other areas (i.e. clouds, water bodies, highways). In the R application, algorithms were used both for supervised classification (based on cross-validation, k-fold method, and Friedman and Nemenyi test) and for data spatialization using the aforementioned algorithms. With the described methods, it was possible to observe that the NDVI values promoted the idea that the shrub-herbaceous phytophysionomy shows reflectance similar to the exposed soil in some areas in the dry period. In the wet period, the kNN algorithm performed better in-class differentiation and vegetation identification (Kappa = 0.9887). In the dry period, the kNN, SVM, and ANN algorithms did not show statistically significant differences regarding their performance, which are considered good classifiers for the period (Kappa = 0.9965; 0.9973; 0.9962, respectively). Therefore, the present study brought innovation in the use of Artificial Intelligence techniques to solve problems in the monitoring, management, and administration of seasonally dry tropical forests with remote data. Being an alternative method to identify, quickly and economically, changes in the forest structure

Keywords: artificial intelligence, caatinga, semiarid, R language.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema de uma camada de alimentação de uma rede neural.	21
Figura 2 – Localização do Parque Nacional do Catimbau/Vale do Catimbau, Pernambuco - Brasil.	22
Figura 3 – Diagrama metodológico para obtenção do NDVI, composição RGB, classificação supervisionada por meio dos algoritmos de aprendizado de máquina e espacialização dos usos.....	23
Figura 4 – Composição RGB Natural (bandas 7, 5, 3) das Imagens Landsat 8 para o vale do Catimbau no período chuvoso (Junho/2019) e seco (Dezembro/2019).	30
Figura 5 – Comparação das feições de uso e cobertura da terra no Parna Catimbau em diferentes visualizações (<i>Landsat 8 vs Google Earth vs in situ</i>)no período seco.	31
Figura 6 – Reflectância das classes Arbórea-arbustiva, Arbustiva-herbácea e Solo exposto nas Bandas B4 e B5 no período seco.	32
Figura 7 – Reflectância das classes Arbórea-arbustiva, Arbustiva-herbácea e Solo exposto nas Bandas B4 e B5 no período úmido.	33
Figura 8 – Normalized Difference Vegetation Index - NDVI para o Vale do Catimbau nos períodos chuvoso (junho/2019) e seco (Dezembro/2019)	34
Figura 9 – Relação da banda B4, B5 do Landsat 8 e do <i>Normalized Difference Vegetation Index</i> – NDVI para a diferenciação entre as classes de vegetação e o solo exposto. A – período úmido. B –Período seco.....	35
Figura 10 – Teste <i>post hoc</i> de Friedman e Nemenyi para os valores ranqueados das médias de Kappa no período chuvoso	37
Figura 11 – Gráfico plot do teste <i>post hoc</i> de Friedman e Nemenyi para os valores ranqueados das médias de Kappa no período seco.....	39
Figura 12 - Espacialização do uso e cobertura da terra a partir dos algoritmos kNN, RF, ANN e SVM para o Vale do Catimbau no período chuvoso.....	41
Figura 13 – Espacialização dos algoritmos de <i>Machine Learning</i> kNN, RF, ANN e SVM para o Vale do Catimbau no período seco (dezembro/2019).....	43

LISTA DE QUADROS

Quadro 1 – Características da resolução espectral e espacial do Landsat 8 sensor OLI das bandas utilizadas no estudo	19
Quadro 2 – Características de diferentes métodos de aprendizado.	20
Quadro 3 – Algoritmos de <i>aprendizado de máquina</i> e parâmetros utilizados para o processamento das imagens do período úmido e seco da área de estudo, em 2019	27
Quadro 4 – Qualidade da classificação segundo intervalos do coeficiente <i>Kappa</i>	29
Quadro 5 – Valores de acurácia global (\hat{G}) e índice de Kappa ($K+$) para diferentes algoritmos classificadores do uso e cobertura da terra.....	40

LISTA DE TABELAS

Tabela 1 - Dados bloqueados não replicados dos resultados de Kappa para os algoritmos RF, SVM, ANN e kNN e respectivos <i>rankings</i> de médias para o período chuvoso.....	36
Tabela 2 – Dados bloqueados não replicados dos resultados de Kappa para os algoritmos de <i>Machine Learning</i> RF, SVM, RNA/ANN e kNN (Método K-fold) e seus respectivos Rankings de médias para o período seco	38
Tabela 3 – Quantificação das classes temáticas em hectares (ha) e sua representação percentual obtida por meio da classificação supervisionada dos algoritmos de aprendizado de máquina no Parna Catimbau com dados Landsat 8 Sensor OLI para o período úmido..	42
Tabela 4 – Quantificação das classes temáticas em hectares (ha) e sua representação percentual obtida por meio da classificação supervisionada dos algoritmos de aprendizado de máquina no Parna Catimbau com dados Landsat 8 Sensor OLI para o período Seco (Dezembro/2019).....	44

LISTA DE SIGLAS E ABREVIATURAS

AA	Áreas Antropizadas
ANN	<i>Artificial Neural Networks</i>
BS	Solo Exposto
DOS	<i>Dark Object Subtraction</i>
FAA	Fitofisionomia Arbórea-Arbustiva
FAH	Fitofisionomia Arbustiva-Herbácea
FTS	Florestas Tropicais Secas
FTSD	Floresta Tropical Seca Decídua
FTSS	Florestas Tropicais Sazonalmente Secas
IA	Inteligência Artificial
IAF/LAI	Índice de área foliar/ <i>Leaf area Index</i>
ICMBio	Instituto Chico Mendes de Conservação da Biodiversidade
IPA	Índice Pluviométrico Annual
IVM/SWIR	Infravermelho médio/ <i>Short-Wave Infrared</i>
IVP/NIR	Infravermelho Próximo/ <i>Near Infrared</i>
kNN	<i>k-Nearest Neighbor</i>
MAC	Metabolismo Ácido das Crassuláceas
ML	<i>Machine Learning</i>
NASA	<i>National Aeronautics and Space Administration</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NDVI _{máx}	Valores máximos de NDVI
OLI	<i>Operational Land Imager</i>
Parna	Parque Nacional
RF	<i>Random Forest</i>
RGB	<i>Red, Green, Blue</i>
SAVI	<i>Soil-Adjusted Vegetation Index</i>
SCP	<i>Semi-automatic Classification plugin</i>
SR	Sensoriamento Remoto
SVM	<i>Support Vector Machine</i>
TIRS	<i>Thermal Infrared Sensor</i>
UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>
USGS	<i>United States Geological Survey</i>

SUMÁRIO

1 INTRODUÇÃO.....	16
2 REVISÃO DE LITERATURA	18
2.1 ECOFISIOLOGIA DE PLANTAS EM ÁREAS DE FLORESTAS TROPICAIS SAZONALMENTE SECAS.....	18
2.2 SENSORIAMENTO REMOTO NO MONITORAMENTO DE FTS	18
2.2.1 Produtos Landsat para a classificação digital de imagens.....	19
2.3 APRENDIZADO DE MÁQUINAS PARA SERVIÇOS ECOSISTÊMICOS	19
3 MATERIAL E MÉTODOS	21
3.1. CARACTERIZAÇÃO DA ÁREA DE ESTUDO	21
3.2 CLASSIFICAÇÃO SUPERVISIONADA E ESPACIALIZAÇÃO DO USO E COBERTURA DA TERRA	23
3.3 FORMAÇÃO DE BANCO DE DADOS.....	24
3.4 PROCESSAMENTO DAS IMAGENS.....	25
3.4.1 Pré-processamento de imagens de satélite.....	25
3.5 CLASSIFICAÇÃO SUPERVISIONADA E ESPACIALIZAÇÃO DO USO E COBERTURA DA TERRA UTILIZANDO APRENDIZADO DE MÁQUINA	27
3.6 VALIDAÇÃO DA CLASSIFICAÇÃO	28
4 RESULTADOS E DISCUSSÃO.....	29
4.1 RESPOSTA ESPECTRAL DA VEGETAÇÃO E SOLO EXPOSTO NAS BANDAS B4 E B5	29
4.2 AVALIAÇÃO DO NDVI COMO UMA POTENCIAL VARIÁVEL NOS MODELOS DE CLASSIFICAÇÃO	33
4.2 VALIDAÇÃO DA CLASSIFICAÇÃO DOS ALGORITMOS.....	35
4.3 DESEMPENHO DOS ALGORITMOS NA CLASSIFICAÇÃO SUPERVISIONADA	39
5 CONCLUSÃO.....	45
REFERÊNCIAS.....	46

1 INTRODUÇÃO

Florestas Tropicais Secas (FTS) são encontradas em regiões onde possuem um grande período de seca severa durante o ano, com chuvas momentâneas, caracterizando um período úmido (SINGH; CHATURVEDI, 2018). Muitos termos vêm sendo implementados, ao longo dos anos e das regiões do mundo, para determinar o tipo de vegetação dessas florestas, como Floresta Tropical Sazonalmente Seca (FTSS); Floresta Tropical Seca Decídua (FTSD); Floresta de monções (para regiões da Índia); *Cuabal* (América Central) e Caatinga, sendo esta a Floresta Tropical Sazonalmente Seca (FTSS) do Brasil.

Caatinga é uma região ecológica que está inserida no semiárido do nordeste brasileiro, possuindo 13 diferentes fisionomias florestais que se expandem em uma larga quantidade de espécies florestais arbóreas e arbustivas (SILVA; LEAL; TABARELLI, 2018). O nome Caatinga (do tupi: *ka'a* [mata] + *tinga* [branca] = mata branca) é referido assim, possivelmente pela vista de uma grande quantidade de árvores em desfolia durante o período seco. Além de ser uma região ecológica com mais de 912.529 km², a Caatinga também é considerada uma região social possuindo mais de 28,6 milhões de habitantes, população esta que possui tradições culturais resilientes e fortes sobrevivendo da agricultura de subsistência. Sendo uma região distante de apoio político, industrial ou de grande potencial para a agricultura extensionista (bem desenvolvida em outras regiões do país), a caatinga é a região brasileira de menor índice de desenvolvimento humano (BUAINAIN; GARCIA, 2013).

Da área total da Caatinga, somente 0,28% são áreas destinadas à proteção da biodiversidade do bioma (SÁNCHEZ et al., 2018). Em um cenário onde este é um dos biomas mais ameaçados do Brasil, grande parte da sua mata nativa é desmatada como consequência dos avanços da pecuária (SILVA et al., 2020). Os problemas da Caatinga também estão associados e relacionados com processos de degradação, desertificação e a instabilidade climática (SILVA et al., 2019). Um dos exemplos para conservação das áreas de extrema importância ecossociológica do país foi a criação do Parque Nacional - Parna do Catimbau, este, originado em concordância com a Lei Federal Nº 9.985/2000 (Lei SNUC) e o Decreto Lei Nº 4.340 de 22 de Agosto de 2002 (BRASIL, 2000, BRASIL, 2002) com a finalidade de ser uma área referência para a manutenção da biodiversidade do bioma Caatinga.

Nesse sentido, a ausência de um monitoramento periódico na área desencadeia diversos impactos negativos para esse ecossistema, como a fragmentação e redução dos remanescentes florestais. Assim, o mapeamento do uso e cobertura da terra permite

identificar a distribuição da cobertura vegetal presente na paisagem, bem como as atividades antrópicas que ocorrem na superfície terrestre. Desse modo, as técnicas do sensoriamento remoto se apresentam como ferramentas potenciais nesse processo, possibilitando o monitoramento periódico.

No entanto, um dos maiores desafios no estudo da identificação e classificação do uso e cobertura da terra, contendo todas as alterações que ocorrem na FTSS, consiste em diferenciar áreas florestais de áreas não florestais (YANG et al., 2016; CUNHA et al., 2020). Nessas áreas de FTSS o verde da vegetação é fortemente relacionado com a precipitação anual média onde as características de uma mesma classe muda ao longo do ano. Este efeito espaço-temporal-climático-fito-ecológico dificulta e frequentemente mascara as ações humanas como desmatamento, agricultura, irrigação, construção, criação de novas vias, entre outras; quando em um período de seca extensa (HEIN et al., 2011; CUNHA et al., 2020). No período de seca, a vegetação comporta um baixíssimo nível de material fotossintético o que é usado como indicador de mudanças na cobertura classificada das florestas (SAHAGÚN et al., 2019).

Diante do exposto, é notória a necessidade de estudos que estabeleçam métodos mais eficientes de identificação das mudanças ocorridas nas FTSS. Nesse sentido, destacam-se os algoritmos de aprendizado de máquina, os quais têm demonstrados grandes contribuições no monitoramento da cobertura florestal (WOLFER et al., 2020), especialmente quando ocorre a mistura da resposta espectral dos alvos.

O *machine learning* ou aprendizado de máquinas se torna uma alternativa viável, oriunda da Inteligência artificial para a classificação de áreas. Tendo em vista a vasta capacidade de aprendizado de algoritmos e múltiplas aplicações dentro da administração de áreas florestais. Zhao et al. (2019), assinalam que algoritmos de aprendizado de máquina como *Support Vector Machine - SVM*, *Random Forest - RF*, *k- Nearest Neighbor - kNN*, e *Artificial Neural Networks – ANN*, por exemplo, são formas acuradas e rápidas de avaliação quantitativa florestal quando baseados em imagens de satélite.

Sob este olhar, o objetivo deste trabalho foi utilizar-se do aprendizado de máquinas para entender a dinâmica de uso e cobertura da terra no Parque Nacional do Catimbau nos períodos de maior precipitação (úmido) e menor precipitação (seco) a partir do imageamento LANDSAT. De forma complementar pretende-se (i) classificar imagens OLI/Landsat 8 no ano de 2019, em períodos diferentes, fazendo uso dos algoritmos SVM, RF, kNN e ANN e analisar o comportamento das bandas utilizando índice de vegetação por diferença normalizada - NDVI. (ii) Avaliar o desempenho dos algoritmos testados a partir da análise

do Índice Kappa e Acurácia Global para o período de Seca e de Chuva. E (iii) Realizar a espacialização do uso e cobertura da terra no Parna Catimbau para os períodos estudados.

2 REVISÃO DE LITERATURA

2.1 Ecofisiologia de plantas em áreas de Florestas Tropicais Sazonalmente Secas

O fator abiótico de maior importância a nível de influência na vida de plantas na FTSS, é a disponibilidade hídrica, além obviamente da fotoinfluência e a carga de micro e macronutrientes (ARAÚJO, CASTRO, ALBUQUERQUE, 2007). De acordo com estes autores, até 90% do que representa o Índice Pluviométrico Anual (IPA), ocorre durante os períodos chuvosos, essa dinâmica afeta germinação, crescimento e produção de flores e frutos. Essa dinâmica de temporalidade espacial quanto ao IPA acontece, de maneira similar, em outras FTSS a nível mundial (ANDRADE et al., 2020), o que influencia, também, na performance funcional do ecossistema assim como do indivíduo vegetal singular.

A sazonalidade pluviométrica induz as plantas da FTSS a um nível de stress hídrico elevado por um certo período. Prova disso são as diversas formas de respostas ecofisiológicas, como o aumento na densidade de tricomas e a mudança na respiração foliar (prevalência das plantas C4 e plantas com o Metabolismo Ácido das Crassuláceas - MAC)(OLIVEIRA et al., 2020). Assim, as FTSS possuem vegetação com fenologia adaptada e altamente dependente da precipitação pluviométrica (ALTHOFF et al., 2018; SANTOS et al., 2020).

2.2 Sensoriamento remoto no monitoramento de FTS

O monitoramento de florestas vem sendo realizado através de técnicas de obtenção de dados a distância (Sensoriamento Remoto - SR). Estudos de mudanças do uso e cobertura da terra em paisagens como a caatinga são raros ou insípidos, tanto que, de acordo com a base científica *ScienceDirect* nos últimos 5 anos até janeiro de 2021 foram publicados somente 100 artigos científicos com o foco em sensoriamento remoto e caatinga, destes sendo 13 com foco em *machine learning* e nenhum com uso de dados Landsat 8 voltados para o monitoramento das mudanças de uso e cobertura da terra (SILVA, et al., 2020). Esta falta de dados científicos que corroborem e embasem comparativamente a pesquisa dificulta o entendimento sobre o efeito da temporalidade e da variabilidade climática espacial que frequentemente pode mascarar ou dificultar possíveis ações humanas de desmatamento (que é comum na região) e comportamento da vegetação (principalmente a mais herbácea) já que seu material fotossintético é confundido com o solo exposto no período seco (ECKERT et

al., 2015; CUNHA et al., 2018).

2.2.1 Produtos Landsat para a classificação digital de imagens

O programa Landsat é formado por uma série de desenvolvimentos do Serviço Geológico dos Estados Unidos – USGS lotado na Administração Nacional da Aeronáutica e Espaço – NASA, desde 1972, com foco principal no fornecimento de dados topográficos, geológicos, estatísticos, geográficos, biológicos, e de informação geomática, por exemplo (SILVERO et al., 2021). O satélite Landsat 8 começou a disponibilizar imagens e dados ao público a partir de 2013 através de seus sistemas *Operational Land Imager* – OLI e o *Thermal Infrared Sensor* (TIRS), este satélite tem a capacidade de maior detecção de mudanças na superfície terrestre já que possui resolução radiométrica de 16 bits com alto nível de cinza (0 a 65.535) esta característica possibilita a facilidade na caracterização da imagem, consequentemente auxilia na diminuição do efeito causado pelas sombras das nuvens (ALBA, 2020) A análise da vegetação, de acordo com Otgonbayar et al. (2019), dá-se a partir das bandas B2 à B7 (Quadro 1) .

Quadro 1 - Características da resolução espectral e espacial do Landsat 8 sensor OLI das bandas utilizadas no estudo

Banda	Faixa espectral (μm)	Resposta espectral	Resolução Espacial
B2	0,45 – 0,51	Azul	30
B3	0,53 – 0,59	Verde	
B4	0,64 – 0,67	Vermelho	
B5	0,85 – 0,88	IVP (Near Infrared/NIR)	
B6	1,57 – 1,65	IVM 1 (SWIR 1)	
B7	2,11 – 2,29	IVM 2 (SWIR 2)	

Legenda: IVP – infravermelho próximo; IVM – infravermelho médio; NIR – *near infrared*; SWIR – *short-wave infrared*

Fonte: Adaptado de USGS (2020)

2.3 Aprendizado de máquinas para serviços ecossistêmicos

A inteligência artificial é uma ciência que possui influência de várias áreas como a Lógica matemática, Ciência da computação, Linguística, Engenharias, Filosofia, Psicologia, Biologia, por exemplo. Como também possui vários ramos definidos a partir dela, a exemplo da robótica, planejamento de *big data*, sistemas nebulosos, e aprendizado de máquina

(CARNEIRO NETO et al., 2018).

O aprendizado de máquinas (ML) é um processo utilizado no ajuste de modelos para um conjunto de dados, através de treinamento ou aprendizagem. Este processo é dado através do treinamento em sequência de um algoritmo. Onde, no fim do processo, uma amostra de dados independentes é utilizada para determinar a acurácia do treinamento.

Os algoritmos de *machine learning* possuem características de diferentes métodos de lidar com o aprendizado como habilidades de aprender com dados mistos, compreender estimar dados não apresentados reconhecer valores discrepantes, interpretação de dados e poder de previsão de dados não inseridos, como assinalam Hastie, Tibshirani e Friedman (2009). Nesta escrita serão abordados os algoritmos de *machine learning* mais utilizados nos estudos de classificação de uso e cobertura da terra voltado para ecologia, principalmente a área florestal, estes são *Support Vector Machine – SVM*; *Random Forest – RF*; *k-Nearest Neighbor – kNN* e *Artificial Neural Networks – ANN* (Quadro 2).

Quadro 2 – Características de diferentes métodos de aprendizado.

Característica	ANN	SVM	RF	k-NN
Habilidade natural de lidar com dados "misturados"	▼	▼	▲	▼
Habilidade com valores não apresentados	▼	▼	▲	▲
Robustez em reconhecer dados de valor discrepante do padrão de entrada	▼	▼	▲	▲
Mudanças em entradas de informações	▼	▼	▲	▼
Escala Computacional (alto N amostral)	▼	▼	▲	▼
Habilidade de lidar com dados irrelevantes	▼	▼	▲	▼
Habilidade de extrair combinações lineares de feições	▲	▲	▼	◆
Capacidade de Interpretação	▼	▼	◆	▼
Poder de Previsão	▲	▲	▼	▲

Legenda: ▲ = Bom; ◆ = Satisfatório; ▼ = fraco.

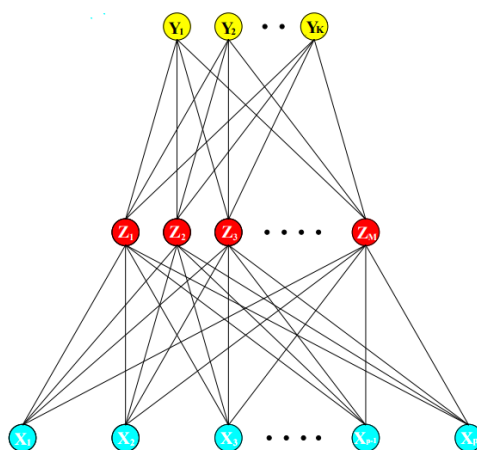
Fonte: Adaptado de Hastie, Tibshirani e Friedman (2009)

As máquinas de vetores de suporte ou SVM são algoritmos que agrupam seus dados em hiperplanos e separam suas classes linearmente, produzindo barreiras não lineares para construir uma barreira linear numa escala maior (QIAN et al., 2015). O algoritmo k-vizinhos

mais próximos ou kNN é um algoritmo é um método de criação de *clusters* através das médias dos valores espaciais próximos, então se o valor está perto de um valor próximo é formado um conjunto classificado. Já o algoritmo *Random Forest* é uma técnica de ML que agrupa uma sequência de árvores treinadas (através do conjunto de dados de treinamento) e as organiza numa variedade de pacotes, esta ação vem do inglês *bagging* (tradução livre = empacotar). Algoritmos de RF capturam interações complexas entre modelos e fazem o *bagging*, a fim de reduzir a variância entre os dados (HASTIE; TIBSHIRANE; FRIEDMAN 2009).

O termo redes neurais representa uma regressão de duas fases ou um modelo de classificação, no caso deste estudo, um modelo de classificação. A ideia deste modelo de classificação é gerar para cada unidade classificatória a possibilidade de interação com todas unidades classificatórias subsequentes (Figura 1).

Figura 1 – Esquema de uma camada de alimentação de uma rede neural.



Fonte: Hastie, Tibshirane e Friedman

3 MATERIAL E MÉTODOS

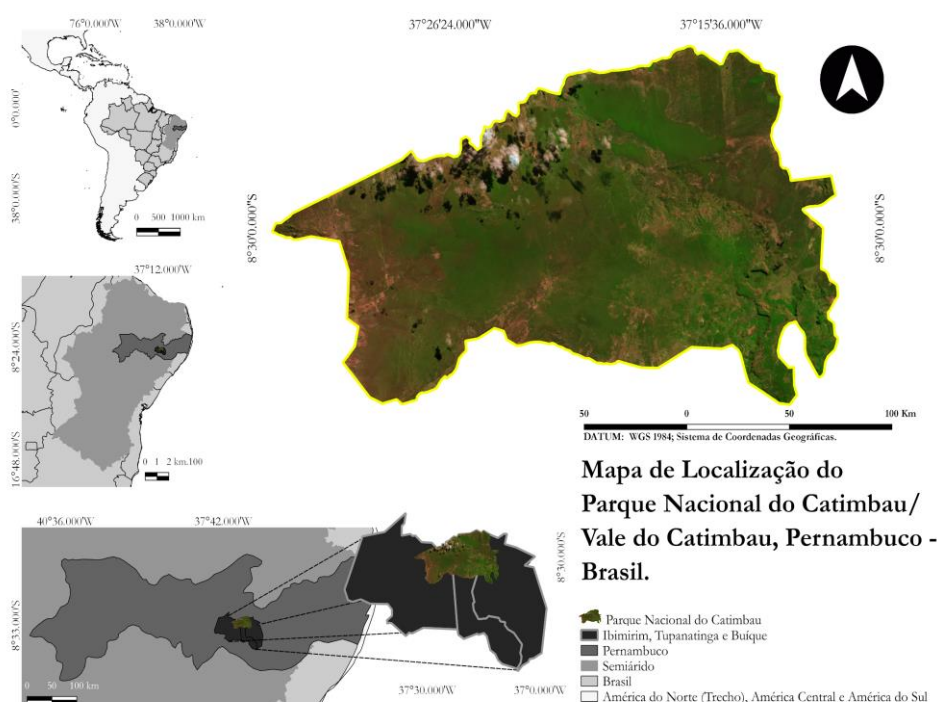
O presente estudo foi desenvolvido no bioma da Caatinga, tendo passos metodológicos relativos ao levantamento de dados primários por meio de imagens de satélite, com análises multifatoriais, assim como dados secundários, os quais corresponderam ao estudo da legislação, documentos e artigos científicos.

3.1. Caracterização da Área de Estudo

O Parque Nacional do Catimbau (Parna Catimbau) ou Vale do Catimbau (Figura 2), inscrito no diploma legal de criação Decreto s/nº (BRASIL, 2002), está situado nos

municípios de Ibimirim, Tupanatinga e Buíque no estado de Pernambuco, na Região Nordeste do Brasil (ICMBIO, 2020). O parque é considerado área núcleo da Reserva da Biosfera da Caatinga pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO, 2001). O centróide deste Parna está posicionado nas coordenadas geográficas 08° 37' S e 37° 09' W, dista 286,2 km da cidade de Recife - PE. O Parna Catimbau está inserido na mesorregião do Agreste Pernambucano, nas microrregiões do Vale do Ipanema (Buíque e Tupanatinga) e Moxotó (Ibimirim), possuindo uma área de 62.278,56 hectares, cerca de 623 km² (Autor, ano).

Figura 2 – Localização do Parque Nacional do Catimbau/Vale do Catimbau, Pernambuco - Brasil.



Fonte: O autor.

O Agreste é caracterizado por ser uma zona de transição, sendo o encerramento da Mata Atlântica e começo da Caatinga. O clima predominante é o tipo BShs', segundo a classificação de Köppen, com transições características de FTSS para o tropical chuvoso do tipo As', no período chuvoso. As características pluviométricas da área de estudo apontam uma alta variação entre períodos de estiagem e chuva variando de 480mm à 1100mm por ano (RITO et al., 2017) e temperatura média anual de 23°C (RITO et al., 2017; PEREIRA et al., 2020) A maior quantidade de chuvas ocorre no período do verão-outono (março-maio) podendo se estender até junho a julho. Estas chuvas podem corresponder até 75% das chuvas

Anuais. O período de setembro a janeiro é considerado o período de menor precipitação e outubro assume a característica de mês com menor índice pluviométrico (SCHRÖDER; RODRÍGUEZ; GÜNTER, 2021).

As classes de solos mais presentes no Parna Catimbau são Neossolos Quartzarênicos (platôs e depressão) estando em 70% do domínio do parque; Latossolos (platôs); Neossolos Litólicos (Complexos rupestres e formações areníticas) (RITO et al., 2017). Também são encontradas registros de Argissolos, Cambissolos, Planossolos e Luvisolos (RITO et al., 2017; PEREIRA et al., 2020).

A vegetação do Parque Nacional do Catimbau é predominante caatinga hiperxerófila nos extratos herbáceos, arbustivos e arbóreos, compondo mosaicos cobrindo grandes áreas. Nos períodos de maior pluviosidade, estas áreas são chamadas de “tapetes verdejantes”, de acordo com Fonseca, Freire (2018) e Freire et al. (2020). As fitofisionomias de caatinga presentes são espécies estão divididas em espécies de Caatinga, Floresta estacional semidecidual, Campo rupestre e espécies de cerrado edáfico (SANO et al., 2019).

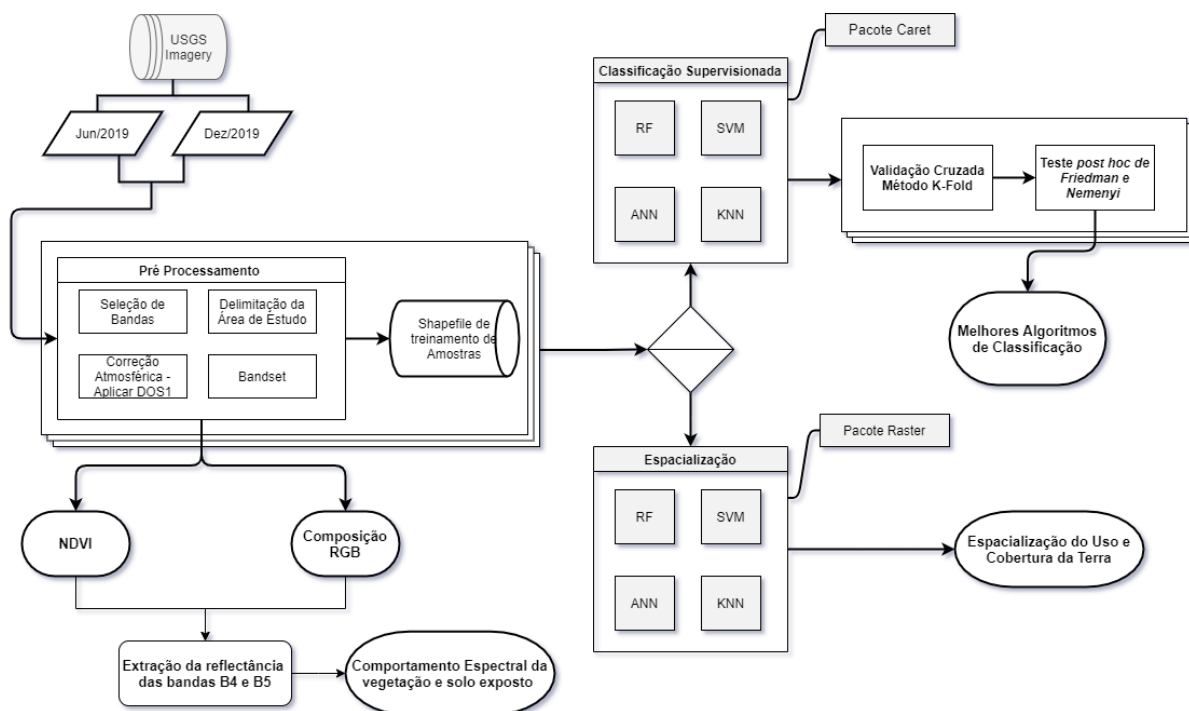
As Famílias vegetacionais predominantes nestas áreas são Asteraceae; Bromeliaceae; Bignoniaceae; Cactaceae; Convolvulaceae; Fabaceae; Malvaceae; Malpighiaceae; Myrtaceae; e Sapindaceae (RODAL et al., 1998, FIGUEIREDO et al., 2000, ANDRADE et al., 2004, GOMES et al., 2006, COSTA FILHO et al., 2012). É relevante assinalar que há espécies endêmicas e recém descobertas na região, como apontam os estudos de Souza-Silva et al. (2010), a exemplo das espécies de Bromeliaceae *Tillandsia Catimbauensis* (LEME; TILL; SIQUEIRA); Apocynaceae *Mandevilla Catimbauensis* (SOUZA-SILVA; RAPINI; MORALES); Compositae: Eupatorieae: Ageratinae *Acritopappus buiquensis* (HIND; BAUTISTA), como registrado por alguns autores (HIND; BAUTISTA, 2009, SOUZA-SILVA; RAPINI; MORALES, 2010, SILVA et al., 2018).

3.2 Classificação supervisionada e espacialização do uso e cobertura da terra

A classificação supervisionada e espacialização dos uso por meio dos algoritmos de aprendizado de máquina na área de estudo foi dada por uma sequência metodológica (Figura 3), iniciando-se em: (i) *download* dos dados Landsat 8 (bandas de imagens) na plataforma do serviço geológico do governo dos Estados Unidos da América para o período úmido (junho/2019) e seco (dezembro/2019); (ii) pré-processamento dos dados e compilação das informações para o mesmo *datum* e sistema de coordenadas geográficas, seleção de bandas, delimitação da área de estudo, correção atmosférica e criação de um *bandset* de bandas; (iii) composição das Bandas coloridas RGB e observação do comportamento espectral das

bandas do sensor OLI/Landsat 8; (iv) obtenção do *Normalized Difference Vegetation Index* – NDVI; (v) classificação utilizando algoritmos de aprendizado de máquinas; (vi) espacialização em do uso e cobertura da terra para o Parna e a utilização de pacotes no *software R*.

Figura 3 – Diagrama metodológico para obtenção do NDVI, composição RGB, classificação supervisionada por meio dos algoritmos de aprendizado de máquina e espacialização dos usos.



Fonte: O autor.

3.3 Formação de Banco de dados

Para a composição e formação do banco de dados espectrais, foram utilizados dados do satélite Landsat 8 sensor OLI/TIRS (*Operational Terra Imager/Thermal Infrared Sensor*), disponibilizados pelo Instituto de Pesquisa Geológica dos Estados Unidos da América - *United States Geological Survey* (USGS, 2020). Observando os fatores climáticos da área de estudo para o período úmido, foi selecionada a imagem de 02 de junho de 2019, a qual é disponibilizada com um pré-processamento de nível 1, contendo desse modo, o georreferenciamento da cena. Seguindo as observações anteriores, escolheu-se a cena do dia 27 de dezembro de 2019 para o período seco/menor pluviosidade, com o mesmo nível de processamento. As imagens estão localizadas na órbita/ponto 215/66. Por serem dados Landsat 8, as imagens possuem resolução espacial de 30 m. Todas as bandas imagens selecionadas obtinham menos de 20% de cobertura de nuvens e este critério foi essencial

para a escolha destes dados. Isto, deu-se principalmente pela não execução de caminhamento em campo, sendo assim, prezou-se pela melhor qualidade na obtenção das imagens Landsat.

Para a obtenção dos limites do estado, municípios e unidade de conservação, foram coletados dados do Sistema de Informações Geoambientais de Pernambuco - SIG CABURÉ, gerido pela a Agência Estadual de Meio Ambiente - CPRH (CPRH, 2020). Para a padronização dos dados, o sistema de coordenadas utilizado foi *Universal Transversa de Mercator* – UTM e *Datum* WGS 1984 (TRUE, 2004, KOSHALE; SINGH, 2020).

3.4 Processamento das Imagens

3.4.1 Pré-processamento de imagens de satélite

Seguindo as diretrizes de Otgonbayar et al. (2019) e Alba (2020), foram utilizadas as bandas de 2 à 7 (quadro 1) para o processamento de dados, pois estas apresentam uma maior resposta fitocromática (OTGONBAYAR et al., 2019). Os dados foram pré-processados no *software* QGis versão 2.18.21, que por meio do *plug-in* SCP (*Semi-automatic Classification plugin*) foi realizado a correção atmosférica (OTGONBAYAR et al., 2019, QGIS, 2020).

No *plugin* SCP, inicialmente realizou-se a delimitação da área de estudo e, em seguida, foi aplicada a correção atmosférica, através do método *Dark Object Subtraction* (DOS) (SANTOS et al., 2015), seguido da criação de um *bandset* (todas as bandas dispostas em um só arquivo). Para Chavez (1996) e Congedo e Munafo (2012), a correção atmosférica se faz necessária, pois alguns *pixels* estão sombreados pela radiância que o satélite recebe e, também, pelo fenômeno de espalhamento atmosférico, combinados com o fato de que alguns alvos na superfície terrestre são pretos, a exemplo da água.

Chavez (1996) assumiu que a mínima reflectância corrigida é melhor que nenhuma reflectância corrigida (CHAVEZ, 1996). Então, definiu-se a existência de objetos pretos e aplica-se a correção atmosférica DOS (CONGEDO; MUNAFO, 2012, FRANCH et al., 2013). Antonio Sobrino, Jiménez-Muñoz e Paolini (2004) destacam a fórmula da correção atmosférica ou a radiância orbital (L_p) (Equação 1), como:

Equação 1 – Correção atmosférica/radiância orbital DOS1.

$$L_p = L_{\min} - L_{DO1\%} \quad (1)$$

Em que,

L_{\min} = valor de contagem digital, soma de todos os pixels com valores mínimos ou iguais a este valor que seja igual a 0,01% (ANTONIO SOBRINO; JIMENÉZ-MUÑOZ; PAOLINI, 2004)

$L_{DO1\%}$ = radiância do objeto preto, reflectância de 0,01 (CONGEDO; MUNAFO, 2012)

A correção atmosférica das bandas selecionadas permitiu diversas aplicações do Sensoriamento Remoto a partir do uso de imagens de satélite, sobretudo do Landsat 8. Na Engenharia Florestal, índices espectrais de vegetação como o *Normalized Difference Vegetation Index* (NDVI) (Equação 2) permitem assegurar quais áreas estão tendo atividade fotossintética, e qual o seu vigor vegetativo (ROUSE JUNIOR et al., 1974).

Equação 2 – Cálculo do *Normalized Difference Vegetation Index* – NDVI.

$$NDVI = \frac{IVP - Vermelho}{IVP + Vermelho} \quad (2)$$

Em que,

IVP = Infravermelho próximo; e Vermelho

Foi realizada a composição RGB, na falsa-cor, fazendo uso das bandas 7, 5 e 3, respectivamente. Essa técnica facilita a visualização dos alvos presentes na área imageada e assim, melhor auxiliar na amostragem das classes.

Desse modo, foi possível criar o arquivo vetorial de amostras de uso e cobertura da terra a partir da seleção de polígonos para as classes temáticas, a fim de estabelecer o padrão espectral de cada tipo de uso, sendo essa técnica similar a utilizada quando se busca o processo de classificação por Máxima Verossimilhança (SILVA et al., 2020.b, ALMEIDA et al., 2020, LOUZADA; BERGIER; ASSINE, 2020). Assim, organizou-se um arquivo vetorial contendo a resposta espectral dos usos para cada banda da imagem OLI/Landsat 8 englobada no estudo. As classes utilizadas e que serviram de base para a criação do arquivo vetorial de amostras de uso da terra, foram adaptadas do relatório de pesquisa de Freire et al. (2015), sendo estas: Fitofisionomia Arbórea-Arbustiva, Fitofisionomia Arbustiva-Herbácea, Áreas Antropizadas (Áreas de Construção, como casas, prédios administrativos do parque e áreas de agricultura), Solo Exposto e Outras Classes (nuvens, sombra de nuvens, açudes, riachos e reservatórios de água).

A fim de mensurar o padrão espectral da vegetação da Caatinga, visando aprimorar o entendimento dos fatores causadores da mistura espectral das classes vegetacionais com o solo exposto em imagens orbitais, foi elaborado gráficos com os valores de reflectância das bandas B4 e B5 para a vegetação Arbórea-Arbustiva e Arbustiva-Herbácea, bem como para o solo exposto. Os gráficos foram gerados no *software* R com o pacote *ggplot2* (WICKHAM, 2016). Todos os comportamentos foram executados para os períodos úmido e seco.

3.5 Classificação supervisionada e espacialização do uso e cobertura da terra utilizando aprendizado de máquina

Foram utilizados algoritmos de aprendizado de máquina (*Machine Learning*) que são apresentados em estudos de classificação supervisionada de espécies florestais em diversas regiões climáticas (SEGAL, 2004, GLEASON; IM, 2012, XI et al., 2020, LI et al., 2020, POURGHASEMI et al., 2020). Os Algoritmos testados foram *Random Forest* (RF), *Artificial Neural Networks* (ANN), *K-Nearest Neighbor* (kNN) e *Support Vector Machine* (SVM) (Quadro 3).

Quadro 3 – Algoritmos de *aprendizado de máquina* e parâmetros utilizados para o processamento das imagens do período úmido e seco da área de estudo, em 2019

Algoritmo	Sigla	Tipo de Parâmetro	Parâmetro
<i>Random Forest</i>	RF	<i>mtry</i> : número de variáveis amostradas aleatoriamente	Sqrt(p)
		<i>ntree</i> : número de árvores	500
<i>Artificial Neural Networks</i>	ANN	<i>maxit</i> : número máximo de interações	100
		<i>weights</i> : peso para cada exemplo	1
<i>K-Nearest Neighbor</i>	kNN	<i>K</i> : comparação entre vetores de características e conjunto de instâncias próximos	1
<i>Support Vector Machine</i>	SVM	<i>Kernel</i> : função utilizada para predição	Radial
		<i>C</i> : ajuste de sensibilidade na margem (em caso de vetores de suporte classificados de forma errada)	1

Fonte: Adaptado de Souza et al. (2016) e Alba (2020)

O arquivo vetorial contendo o padrão espectral dos usos foi inserido no processo como sendo as amostras de treinamento dos algoritmos, procedendo com os ajustes nos parâmetros dos modelos, tanto para a imagem do período úmido quanto para o período seco da área de estudo. A precisão de cada parâmetro de treinamento para os algoritmos foi dada a partir de testes que apontaram maior precisão e ajuste dos modelos (SOUZA et al., 2016, ALBA, 2020). O processo de classificação supervisionada das imagens do satélite OLI/Landsat 8 foi operada e desenvolvida na linguagem de programação estatística e gráfica do programa R - versão 3.6.2., utilizando-se de regressões não paramétricas e consonância (Hong et al. 2020).

Foram utilizados para a obtenção dos dados os pacotes (i) *raster*, o qual possui

implementos básicos e funções de alto nível para funções com dados raster (HIJMANS, 2020), como no caso de dados Landsat; (ii) *caret*, é uma sigla para treinamento de regressões e classificações; que converte funções e classes treinadas em matrizes de confusão, além de treinar e criar gráficos de classificação e modelos de regressão (KUHNN, 2020); (iii) *rgdal*, uma biblioteca onde são armazenados os dados geoespaciais (BIVAND, 2020); (iv) *RandonForest*, baseado na floresta de inserções aleatórias de árvores de Breiman (2001) (LIAW, 2018); (v) *nnet*, pacote para redes neurais artificiais avançadas com uma única camada para modelos logísticos lineares multinomiais (RIPLEY; VENABLES, 2020). Estes pacotes (*caret* e *raster*) auxiliaram, também, na espacialização do uso e cobertura da terra, a partir dos modelos e equações estabelecidas para os algoritmos RF, SVM, ANN e kNN.

3.6 Validação da Classificação

A escolha do melhor algoritmo de *Machine Learning* na classificação, principalmente da vegetação da caatinga presente na área de estudo, foi baseada em critérios de alta performance como a acurácia global (Equação 3) já que, estatisticamente, é responsável por determinar o quão correto e isento de erros é o valor apresentado (PONTIUS JUNIOR; GILMORE; MILLONES, 2011), bem como o Índice de Kappa (k) (Equação 4), que avalia a qualidade das classificações (LANDIS; KOCH, 1977).

Equação 3 – Obtenção da acurácia global dos dados.

$$\hat{G} = \sum_{i=1}^H \hat{p}_{ii}$$

Em que:

\hat{G} = acurácia global, igual ao somatório da diagonal principal da matriz de proporções \hat{P}_{ij} (PONTIUS JÚNIOR; GILMORE; MILLONES, 2011)

Equação 4 – Obtenção do Índice de Kappa

$$K = \frac{N \sum_{i=1}^r X_{ij} - \sum_{i=1}^r X_{i+} X_{+i}}{N^2 - \sum_{i=1}^r X_{i+} X_{+i}}$$

Em que:

r = número de classes

X_{ij} = número de elementos classificados corretamente

X_{i+} = total de elementos classificados para uma categoria i

X_{+i} = total de elementos de referência amostrados para uma categoria i

N = número total de amostras

Os valores do Índice de Kappa (K) variam entre -1 e 1, o qual o valor 1 reflete uma total concordância entre a imagem classificada e a imagem real, já o valor -1 ocorre quando

há uma total discordância. Esses valores são comparados e confrontados com os conceitos definidos por Landis e Koch (1977), que servem para avaliar a exatidão do mapa temático obtido da classificação da imagem (Quadro 4).

Quadro 4 – Qualidade da classificação segundo intervalos do coeficiente *Kappa*

Valor do <i>Kappa</i>	Qualidade da Classificação
<0,00	Péssima
0,0-0,20	Ruim
0,20-0,40	Razoável
0,40-0,60	Boa
0,60-0,80	Muito Boa
0,80-1,00	Excelente

Fonte: Landis e Koch (1977)

A avaliação do desempenho dos algoritmos de aprendizado de máquina na classificação dos usos presentes no Parna Catimbau foi baseada na validação cruzada (método *k-fold*) com 30 repetições, observando para cada algoritmo o valor da acurácia global e do índice de kappa (KOHAVI, 1995). É atribuído, de acordo com Pahikkala et al. (2008), o peso “1” para o algoritmo mais eficiente, criando um *ranking* de valores sucessivos (ALBA, 2020).

A fim de estabelecer um nível de comparação estatística entre os resultados obtidos de acurácia e índice Kappa dos algoritmos, verificou-se a existência de diferenças significativas entre estes. Foi aplicado o teste estatístico *post hoc* pareado de comparações múltiplas das somas das médias de classificação dos dados bloqueados não replicados, o teste de Friedman e Nemenyi (FRIEDMAN, 1940; NEMENYI, 1962). A hipótese de nulidade do teste de Friedman é que todos os algoritmos são equivalentes. Já o teste *post hoc* de Nemenyi é aplicado quando a hipótese de nulidade é rejeitada, automaticamente apontando as diferenças significativas entre os algoritmos (TAVARES JÚNIOR et al., 2020). Já para a obtenção deste teste, foi utilizado o pacote “*tsutils*” do *software* R (KOURENTZES, 2020).

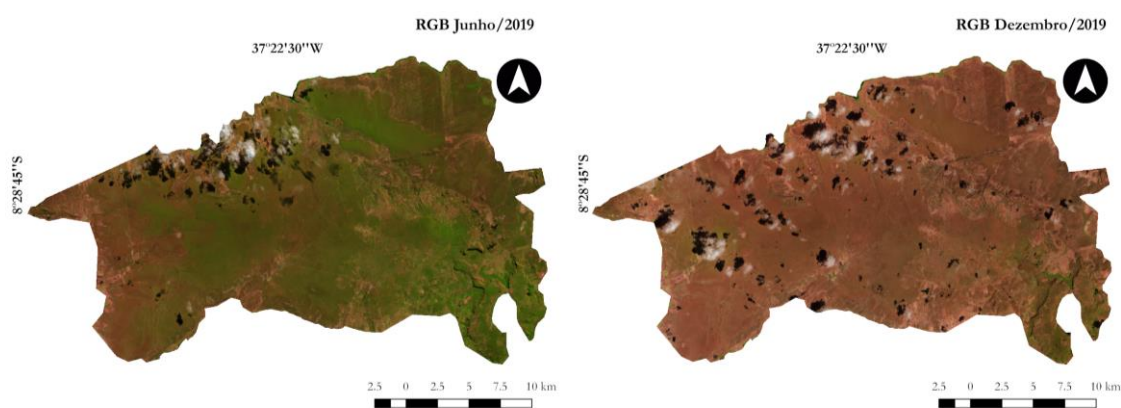
4 RESULTADOS E DISCUSSÃO

4.1 Resposta espectral da vegetação e solo exposto nas bandas B4 e B5

É possível notar a influência da ausência de chuvas na vegetação da caatinga (Figura 4), onde a folhagem é perdida, gerando diminuição da tonalidade da vegetação nas imagens OLI/Landsat 8 (RGB Dezembro/2019), porém no pós-chuvas, é notório a recuperação da coloração verde e consequentemente da vegetação (RGB Junho/2019). Esse comportamento corrobora com a literatura, uma vez que a vegetação sadia, com folhas, apresenta um pico de reflectância na região do visível, no canal do verde, devido a presença de pigmentos foliares como a clorofila, proporcionando desse modo, relacionar a cor verde à vegetação.

Do mesmo modo, esta sensação visual foi identificada em estudos de Souza Sobrinho et al. (2016), onde florestas tropicais sazonalmente secas e florestas tropicais secas possuem um padrão de regeneração alto, principalmente em áreas preservadas. As áreas preservadas possuem árvores com um ciclo de vida natural e completo, assim como maior diversidade de espécies, o que auxilia na manutenção da recuperação, sucessão, polinização, diversificação e estoque de carbono, por exemplo.

Figura 4 – Composição RGB Natural (bandas 7, 5, 3) das Imagens Landsat 8 para o vale do Catimbau no período chuvoso (Junho/2019) e seco (Dezembro/2019).



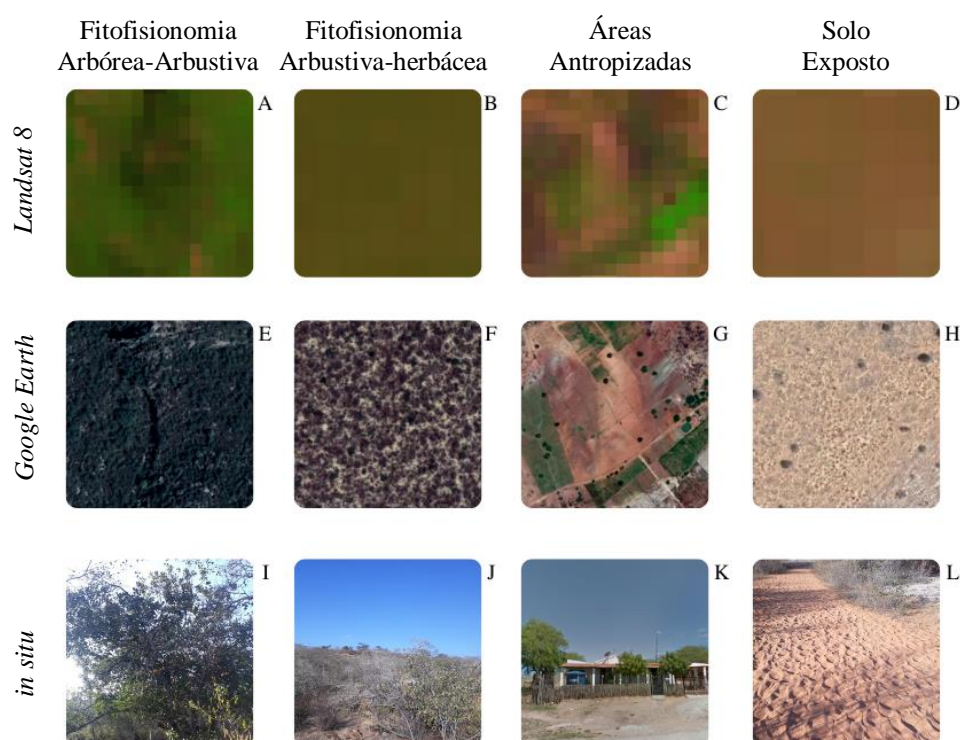
Fonte: O autor.

Algumas das classes presentes neste estudo apresentam feições próximas visualmente, quando abordado o período seco. É o caso da fitofisionomia Arbustiva-Herbácea que estende por todo o parque, ocupando mais de 80%. Esta se confunde com o solo exposto em algumas regiões, o que é típico da Caatinga e florestas tropicais secas. O solo exposto da região tem características de erosão da rocha sedimentar arenito, o Neossolo Quartzarênico presente em 70% do parque (RITO et al., 2017), apresentando uma resposta espectral muito próxima à vegetação denominada Caatinga Hiperxerófila.

É possível identificar proximidades visuais (Figura 5) nos pixels da imagem OLI/Landsat 8, imagens do *Google Earth* e Imagens *in situ* entre as Fitofisionomias

Arbustiva-Herbácea e o Solo Exposto, o que evidencia a necessidade de utilização de novas abordagens para a classificação de uso e cobertura da terra para áreas de floresta tropical seca, tendo em vista a dificuldade visual de classificação automática ou sem treinamento dessas áreas. Na figura 5 as figuras 5A e 5E, 5B e 5F, 5C e 5G, 5D e 5H, representam a mesma área geográfica, já as figuras 5I a 5L são representações destas feições/classes dentro da área de estudo, mas não necessariamente na mesma localização pontual geográfica.

Figura 5 – Comparação das feições de uso e cobertura da terra no Parna Catimbau em diferentes visualizações (*Landsat 8* vs *Google Earth* vs *in situ*) no período seco.

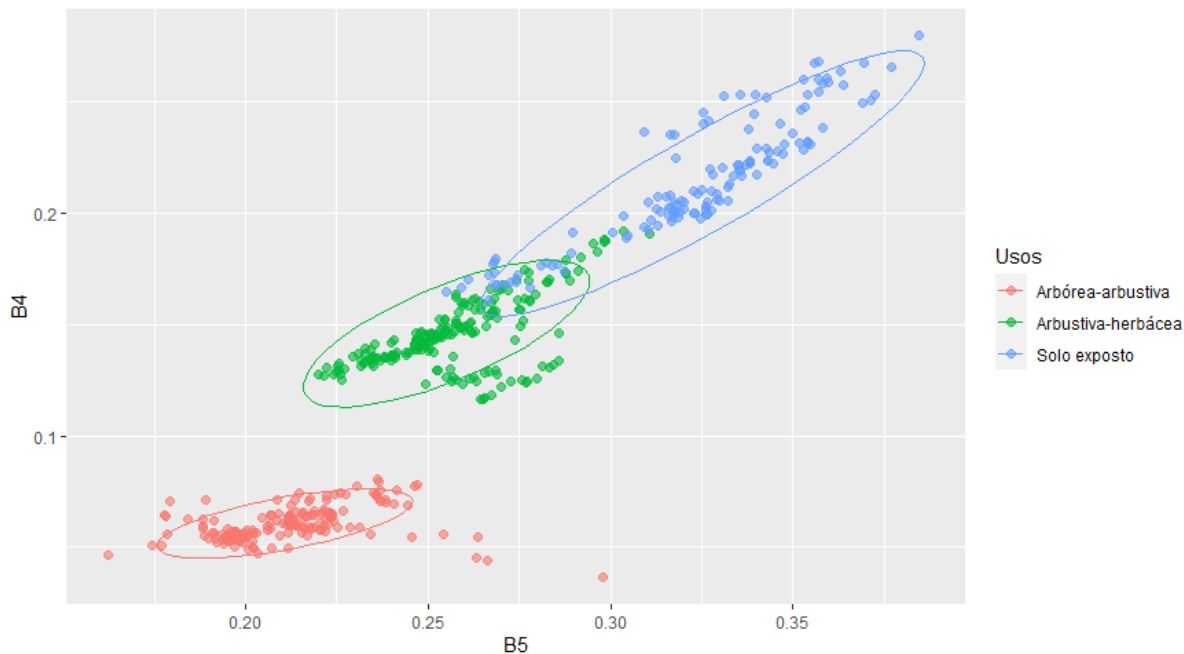


Fonte: *Landsat 8* (USGS, 2020) (A-D), *Google Earth* (2020) (E-H), Luciano Cavalcanti (s/a) (I-J; L), *Google Street View* (2018) (K).

A Figura 6 ilustra o comportamento espectral das diferentes fitofisionomias nas bandas do vermelho e do infravermelho próximo. Observou-se que a vegetação Herbácea-arbustiva apresenta, por vezes, resposta espectral similar àquela apresentada pelo solo exposto, gerando erros significativos no produto obtido pelo classificador. A resposta espectral das classes de vegetação e solo exposto considerando as B4 e B5 permitiu a diferenciação das classes e a possibilidade de distinção entre si, mesmo havendo uma matriz de confusão causada pelas relações intrínsecas da vegetação e o índice pluviométrico. A banda B4 consegue demonstrar melhor individualização da vegetação arbórea-arbustiva, por

possuir média e desvio padrão altos, sendo imprescindível a inclusão dessa banda na identificação de espécies vegetacionais, de acordo com Alba (2020).

Figura 6 – Reflectância das classes Arbórea-arbustiva, Arbustiva-herbácea e Solo exposto nas Bandas B4 e B5 no período seco.



Fonte: O autor

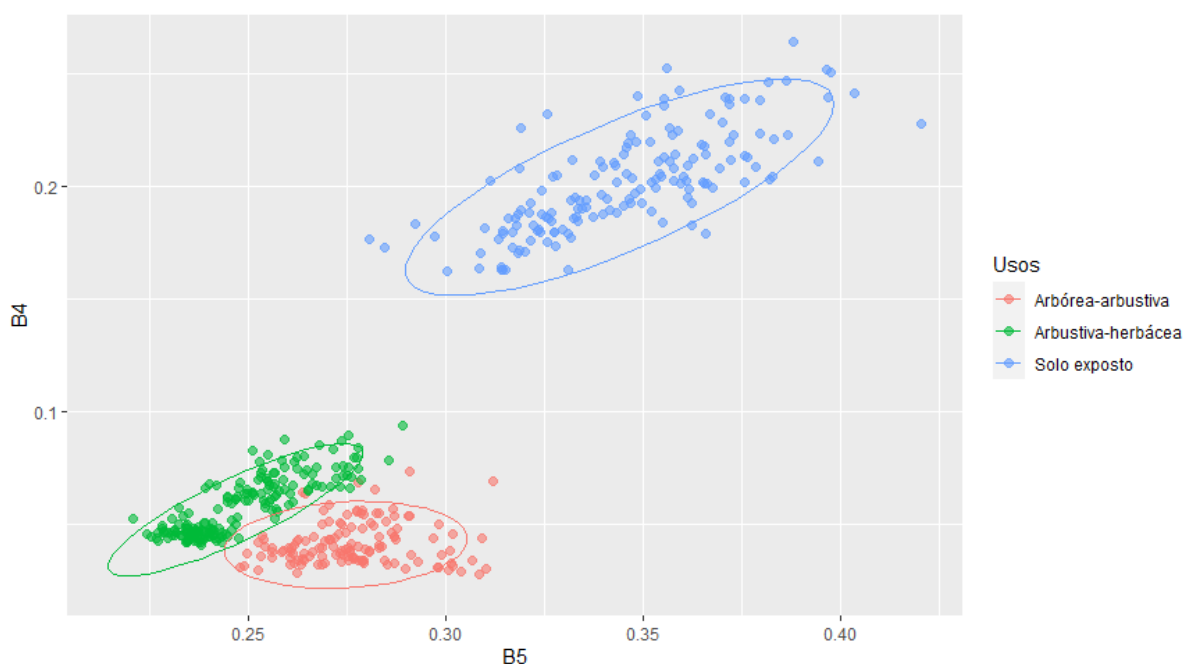
Em contrapartida, no período úmido (Figura 7) esse comportamento se normaliza e as classes de vegetação estão mais próximas do esperado e distantes das fisionomias de solo exposto. É observada alta correlação entre as classes Arbórea-Arbustiva e Arbustiva-herbácea, sendo que isso se explica uma vez que, tendo maior índice pluviométrico, há desaceleração no processo de senescência e aceleração metabólica, com produção de folhas, resultando, conseqüentemente, na produção de clorofila e aumento da reflectância na banda B5.

Ressaltando que quanto maior a densidade de folhas, menor é a reflectância no vermelho, pois existe uma maior quantidade de pigmentos foliares ativos fazendo a absorção de Radiação eletromagnética para os processos metabólicos da planta, como assinalam Miura, Huete e Yoshioka (2006). Este fenômeno dá-se a partir da observação dos valores da B4 no período úmido, que foi próximo a 0,04, enquanto no período seco (sem atividade dos metabólitos foliares), os valores subiram ($\pm 0,08$).

Do mesmo modo, observou-se que a B4 apresenta maior contribuição para a distinção das classes de vegetação para o período úmido. No entanto, destaca-se que as amostras de treinamento do solo exposto apresentaram valores mais elevados na B5 quando

comparados com o período seco, contribuindo significativamente para a individualização/diferenciação dessa cobertura. Assim, o uso combinado da resposta espectral desses comprimentos de onda em modelos de classificação, juntamente com técnicas aprimoradas de leitura e diferenciação desses padrões espectrais, pode contribuir para a minimização dos problemas encontrados em estudos de mapeamento do uso e cobertura em regiões da Caatinga.

Figura 7 – Reflectância das classes Arbórea-arbustiva, Arbustiva-herbácea e Solo exposto nas Bandas B4 e B5 no período úmido.



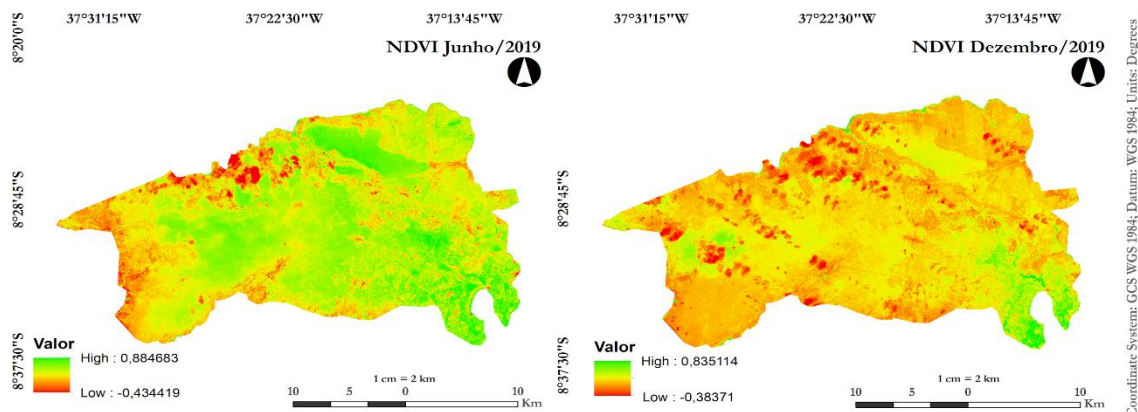
Fonte: O autor.

4.2 Avaliação do NDVI como uma potencial variável para modelos de classificação

Sabendo que o NDVI é usado como um indicador de atividade vegetacional, é factível afirmar que o vigor de áreas da caatinga está vulnerável em períodos de seca. A mutação de características fitofisionômicas do período úmido (Figura 8) (Esquerda) para o Período seco (Figura 8) (Direita) é de extrema relevância para a compreensão do motivo que se faz necessário um sistema seguro de separação e identificação de classes, com menor propensão a erros, como a Inteligência Artificial e Aprendizado de máquinas. A classificação de uso e cobertura da terra para áreas de floresta tropical sazonalmente seca possui particularidades de mistura espectral causadas justamente pela razão Chuva/NDVI na escala pluviométrica anual, além da relação de séries estatísticas que correlacionam esta razão

(BARBOSA et al., 2019).

Figura 8 – Normalized Difference Vegetation Index - NDVI para o Vale do Catimbau nos períodos



chuvoso (junho/2019) e seco (Dezembro/2019)

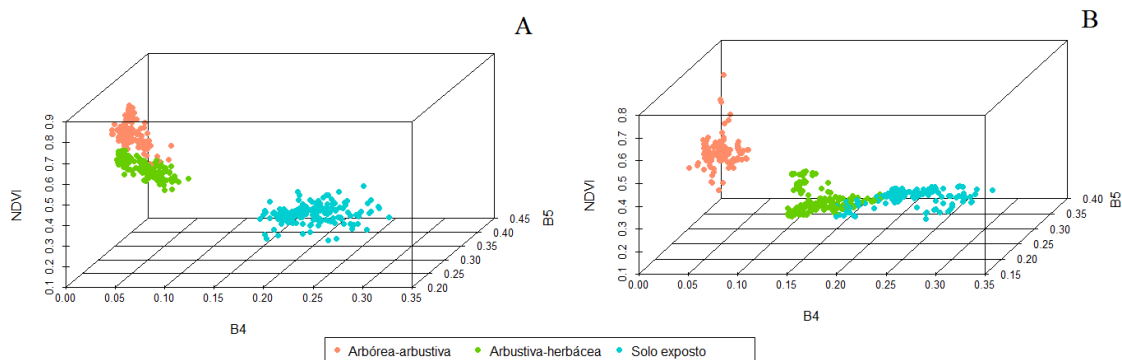
Fonte: O autor.

Observando os valores de Máxima do NDVI ($NDVI_{m\acute{a}x}$) nas duas imagens (NDVI Junho/2019 e NDVI Dezembro/2019) (Figura 8), é possível definir que em períodos úmidos identificou-se valores mais elevados desse índice, expressando uma vegetação com maior vigor vegetativo ($NDVI_{m\acute{a}x}$: 0,8864), enquanto que em períodos de seca, foram detectados valores menos expressivos ($NDVI_{m\acute{a}x}$: 0,8351), como assinala Silva et al. (2020.b).

A interação da banda B4, B5 e o NDVI (Figura 9) proporcionou uma melhor compreensão sobre o comportamento das classes vegetacionais, principalmente a de menor porte (Fitofisionomia Arbustivo-Herbácea) e as transições em um processo de seca e alto índice pluviométrico. Observou-se uma redução significativa dos valores de NDVI expressos pela vegetação Arbustivo-Herbácea no período seco, assumindo um comportamento similar ao solo exposto, enquanto que no período úmido ficou evidenciado a recuperação de suas funções fotossintéticas e densidade de copa, com valores mais elevados de NDVI, próximos a 0,70.

Entretanto, o NDVI apresentou pouca contribuição para a redução do processo de mistura espectral entre as classes temáticas (Figura 9). O NDVI tem grande importância no entendimento dos processos fisiológicos da vegetação diante da variação pluviométrica anual, porém não demonstrou ter influência na distinção de padrões espectrais de alvos com respostas similares, objeto de estudo do presente trabalho.

Figura 9 – Relação da banda B4, B5 do Landsat 8 e do *Normalized Difference Vegetation Index* – NDVI para a diferenciação entre as classes de vegetação e o solo exposto. A – período úmido. B – Período seco.



Fonte: O autor.

Resultados dessa magnitude demonstram o potencial do uso do NDVI como um índice eficiente na rápida resposta quando há interesse na estrutura vegetal, tendo-se demonstrado tendências naturais quando submetido a diferentes variações na composição da FTSS. Outros índices de vegetação, como Albedo de superfície, *Soil-Adjusted Vegetation Index*- SAVI e *Leaf Area Index* – IAF, podem apresentar inúmeras aplicações no manejo de áreas reais, com soluções factíveis e necessárias, como assinala Jin et al. (2018).

4.2 VALIDAÇÃO DA CLASSIFICAÇÃO DOS ALGORITMOS

A classificação por aprendizado de máquina depende essencialmente da validação das estimativas apresentadas pelos resultados. Os dados bloqueados não replicados utilizados para a validação cruzada dos resultados de K+, pelo método *k-fold* (Tabela 1), baseia-se em 30 repetições para cada algoritmo. Os dados referentes ao período de maior pluviosidade (Junho/2019), geraram um *ranking* de médias, ilustrando a posição dos algoritmos de acordo com o coeficiente de concordância de Kappa (Tabela 1).

Tabela 1 - Dados bloqueados não replicados dos resultados de Kappa para os algoritmos RF, SVM, ANN e kNN e respectivos *rankings* de médias para o período chuvoso

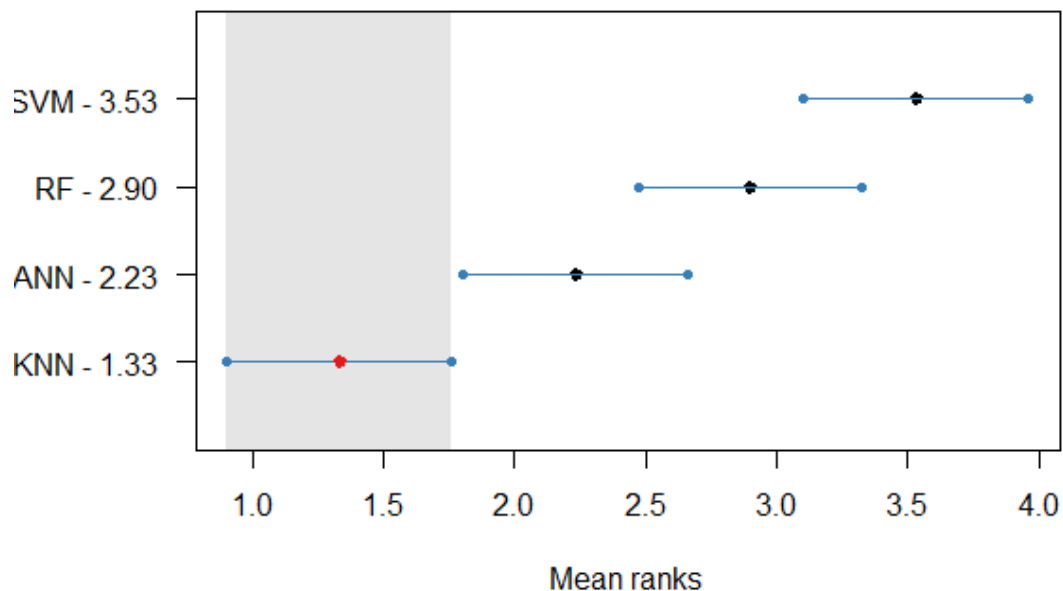
Dados bloqueados não replicados de Kappa (método k-fold)					Ranking de médias			
K(+)	RF	SVM	ANN	kNN	RF	SVM	ANN	kNN
úmido								
Seed 1	0,981282	0,969714	0,984754	0,988786	3	4	2	1
Seed 2	0,981155	0,981024	0,985317	0,988591	3	4	2	1
Seed 3	0,977396	0,981005	0,977582	0,984721	4	2	3	1
Seed 4	0,981244	0,985035	0,981600	0,992411	4	2	3	1
Seed 5	0,980992	0,965731	0,988853	0,988411	3	4	1	2
Seed 6	0,981692	0,977291	0,988630	0,988477	3	4	1	2
Seed 7	0,981122	0,950921	0,973716	0,992474	2	4	3	1
Seed 8	0,981094	0,984646	0,985061	0,988761	4	3	2	1
Seed 9	0,981135	0,927957	0,989031	0,992401	3	4	2	1
Seed 10	0,977077	0,980986	0,969705	0,988305	3	2	4	1
Seed 11	0,980909	0,981005	0,985385	0,992302	4	3	2	1
Seed 12	0,980802	0,977122	0,988938	0,988621	3	4	1	2
Seed 13	0,980926	0,980706	0,981623	0,988699	3	4	2	1
Seed 14	0,984726	0,980710	0,977099	0,988646	2	3	4	1
Seed 15	0,984866	0,980778	0,985359	0,984621	2	4	1	3
Seed 16	0,981141	0,985014	0,985004	0,988684	4	2	3	1
Seed 17	0,981115	0,962517	0,977323	0,988305	2	4	3	1
Seed 18	0,980932	0,977449	0,988730	0,988998	3	4	2	1
Seed 19	0,981385	0,980898	0,977605	0,992401	2	3	4	1
Seed 20	0,985072	0,981508	0,989112	0,988227	3	4	1	2
Seed 21	0,981113	0,980931	0,988953	0,984607	3	4	1	2
Seed 22	0,984699	0,961754	0,973622	0,988807	2	4	3	1
Seed 23	0,981300	0,973423	0,988894	0,984854	3	4	1	2
Seed 24	0,985051	0,980864	0,985173	0,984625	2	4	1	3
Seed 25	0,981261	0,980530	0,981458	0,988603	3	4	2	1
Seed 26	0,981266	0,980776	0,981142	0,992488	2	4	3	1
Seed 27	0,981335	0,942963	0,980916	0,992239	2	4	3	1
Seed 28	0,977167	0,981026	0,981248	0,988635	4	3	2	1
Seed 29	0,984798	0,935648	0,977847	0,985014	2	4	3	1
Seed 30	0,977031	0,981167	0,981243	0,988531	4	3	2	1
MÉDIA	0,981369	0,973037	0,982697	0,988708	2,90	3,53	2,23	1,33
	RF	SVM	ANN	kNN	RF	SVM	ANN	kNN

Fonte: O autor.

O cálculo do teste *post hoc* de Friedman e Nemenyi para o período de maior pluviosidade (Figura 10) com os valores ranqueados das médias de K+ aponta que a distância crítica (*Critical Distance* – CD) foi de $\pm 0,856$. Assim, o algoritmo kNN apresentou maior performance. Pode-se afirmar, estatisticamente, que kNN é superior ao SVM, ANN e RF no processo de diferenciação de Florestas Tropicais Sazonalmente Secas no período úmido.

Este registro é inovador, pois devido a sua organização simples, normalmente, não é incluído em estudos de modelagem, porém Deng et al. (2016) destaca a sua robustez e a importância do algoritmo.

Figura 10 – Teste *post hoc* de Friedman e Nemenyi para os valores ranqueados das médias de Kappa no período chuvoso



Fonte: O autor.

O período seco, por sua vez, obteve boa classificação para os algoritmos kNN, ANN e SVM na etapa de treinamento dos dados, apresentando valores no *ranking* de médias de 2,07, 2,27 e 1,95, respectivamente (Tabela 2).

Tabela 2 – Dados bloqueados não replicados dos resultados de Kappa para os algoritmos de *Machine Learning* RF, SVM, RNA/ANN e kNN (Método K-fold) e seus respectivos Rankings de médias para o período seco

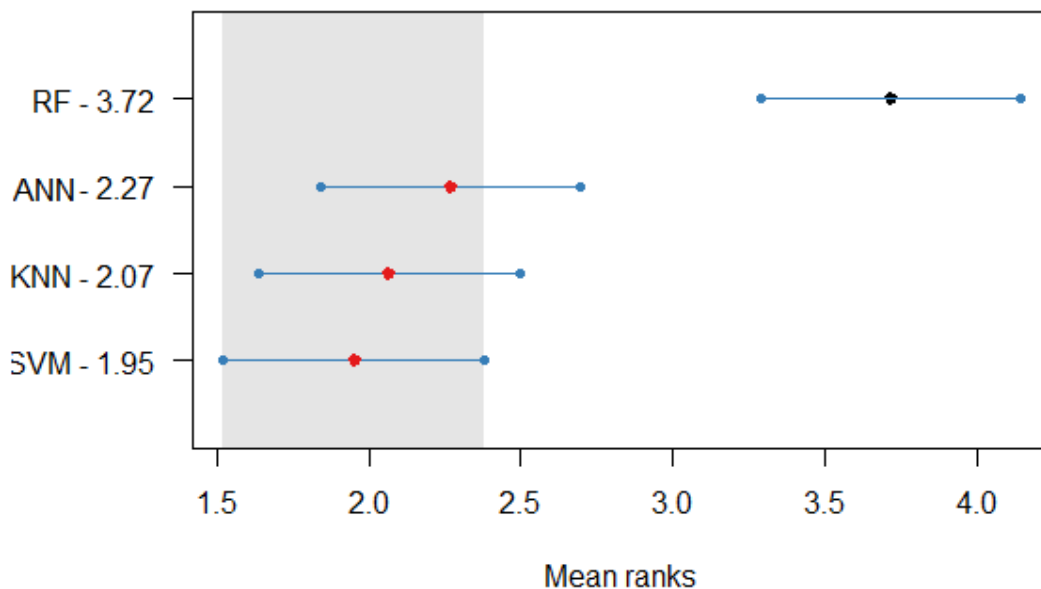
Dados bloqueados não replicados de Kappa (método k-fold)					Ranking de médias			
K(+)	RF	SVM	ANN	kNN	RF	SVM	ANN	kNN
Seed 1	0,9956825	0,9955882	0,9955882	1	2	3,5	3,5	1
Seed 2	0,9907565	0,9957766	0,986385	0,9953301	3	1	4	2
Seed 3	0,9910781	0,995509	0,99523	0,995509	4	1,5	3	1,5
Seed 4	0,9914125	1	0,9954186	0,9954186	4	1	2,5	2,5
Seed 5	0,9911169	1	0,9957766	1	4	1,5	3	1,5
Seed 6	0,9909669	0,9868479	0,9953301	0,9955157	3	4	2	1
Seed 7	0,9953301	1	1	0,9955882	4	1,5	1,5	3
Seed 8	0,9907704	1	0,99523	0,9953301	4	1	3	2
Seed 9	0,9906936	1	0,9954114	0,9956825	4	1	3	2
Seed 10	0,9907637	1	1	0,99523	4	1,5	1,5	3
Seed 11	0,9820616	0,991003	0,9954114	0,995614	4	3	2	1
Seed 12	0,9905763	1	0,9954474	1	4	1,5	3	1,5
Seed 13	0,9908805	1	0,9954474	0,9953301	4	1	2	3
Seed 14	0,9907058	1	0,9954474	1	4	1,5	3	1,5
Seed 15	0,9911169	0,99523	0,9954114	0,995509	4	3	2	1
Seed 16	0,9909471	1	1	0,9955423	4	1,5	1,5	3
Seed 17	0,9910806	1	1	1	4	2	2	2
Seed 18	0,9953301	0,99523	0,9955882	0,9953301	2,5	4	1	2,5
Seed 19	0,9907657	0,9909559	0,995614	0,9953301	4	3	1	2
Seed 20	0,9911414	1	1	0,9957064	4	1,5	1,5	3
Seed 21	0,9905751	1	1	0,9908504	4	1,5	1,5	3
Seed 22	0,9908541	1	1	1	4	2	2	2
Seed 23	0,9911489	0,9954114	0,9955882	0,9957766	4	3	2	1
Seed 24	0,9827098	1	0,9955157	1	4	1,5	3	1,5
Seed 25	0,9908658	1	0,9954114	0,9954114	4	1	2,5	2,5
Seed 26	0,9908877	1	1	0,9955423	4	1,5	1,5	3
Seed 27	0,9914305	1	0,9955157	0,9954114	4	1	2	3
Seed 28	0,99523	0,9824884	0,990552	0,9956825	2	4	3	1
Seed 29	0,995614	1	1	0,9955423	3	1,5	1,5	4
Seed 30	0,9906648	0,9956461	0,9910065	1	4	2	3	1
MÉDIA	0,991105	0,997323	0,996211	0,996539	3,72	1,95	2,27	2,07
	RF	SVM	ANN	kNN	RF	SVM	ANN	kNN

Fonte: O autor.

Ao avaliar estatisticamente o desempenho dos algoritmos para o período seco, observou-se que os algoritmos SVM, kNN e ANN não se diferem. Desse modo, ambos podem ser indicados para o monitoramento dos usos em regiões do semiárido no período de

menor disponibilidade hídrica. (Figura 11).

Figura 11 – Gráfico plot do teste *post hoc* de Friedman e Nemenyi para os valores ranqueados das médias de Kappa no período seco



Fonte: O autor.

Entretanto, o algoritmo RF apresentou desempenho inferior, diferindo significativamente dos demais algoritmos testados. O baixo desempenho expressado pelo RF pode ser atribuído a amostragem, uma vez que Rana e Suryanarayana (2020) afirmam que RF apresenta valores inferiores nos parâmetros de avaliação da qualidade em classificações de uso e cobertura da terra quando se tem abordagens com pequenas quantidades de amostra por classe, como é o caso deste estudo.

4.3 DESEMPENHO DOS ALGORITMOS NA CLASSIFICAÇÃO SUPERVISIONADA

Todos os algoritmos apresentados nesta pesquisa foram considerados excelentes de acordo com a média dos valores para o índice de *Kappa* ($K+$), assim como para a acurácia global (\hat{G}) em ambos os períodos de análise (Quadro 5). Para o período úmido, o algoritmo kNN apresentou os maiores valores, com \hat{G} e $K+$ de 0,9918 e 0,9887, respectivamente.. Já no período seco, o algoritmo SVM apresentou maiores valores de \hat{G} e $K+$, sendo esse no valor de 0,9973.

Quadro 5 – Valores de acurácia global (\hat{G}) e índice de Kappa (K^+) para diferentes algoritmos classificadores do uso e cobertura da terra.

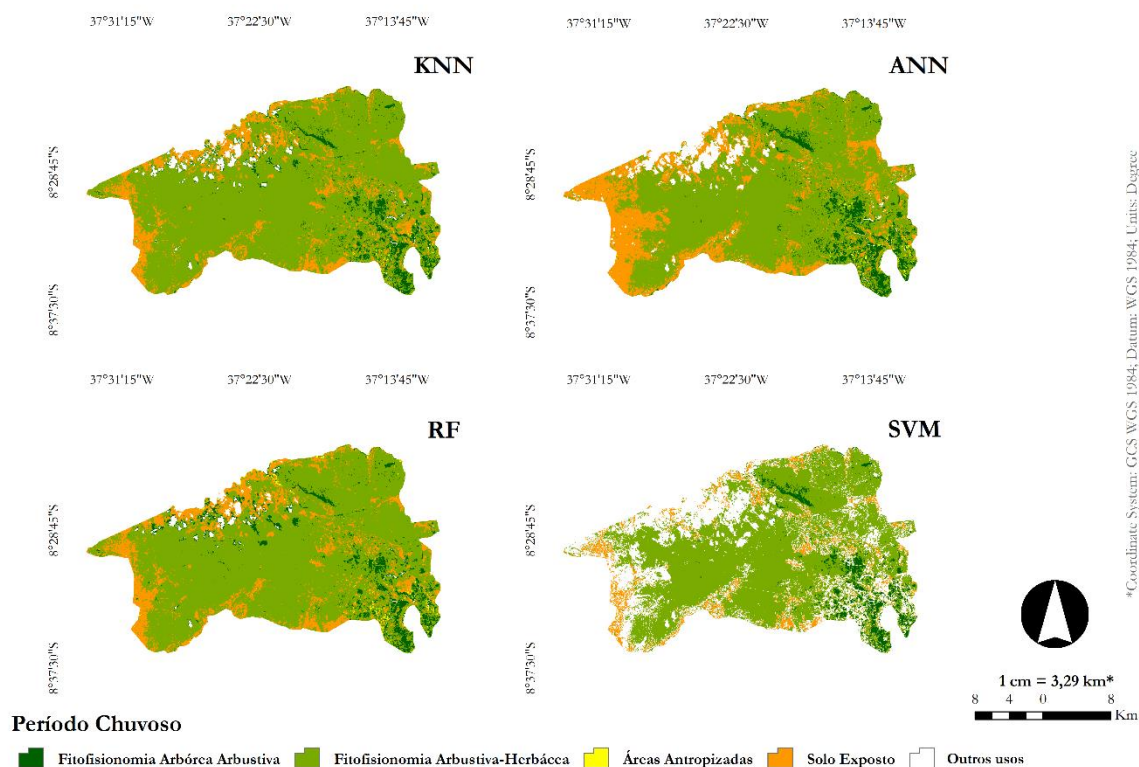
Período	Algoritmos	Acurácia Global (\hat{G})	Kappa (K^+)
Úmido	RF	0,9855	0,9814
	SVM	0,9815	0,9730
	ANN	0,9864	0,9827
	kNN	0,9918	0,9887
Seco	RF	0,9911	0,9911
	SVM	0,9973	0,9973
	ANN	0,9962	0,9962
	kNN	0,9965	0,9965

Fonte: O autor.

Comparar diferentes classificadores pode promover informações importantes que corroboram com a distribuição das classes de dados (SHAO; LUNETTA, 2012). O comportamento de SVM em relação à kNN, sendo o primeiro inferior no período úmido, indica que se kNN fornece bons resultados, as classes são bastante separáveis (DENG et al., 2016). O resultado do kNN no período úmido pode ser explicado pela robustez do método, que permitiu identificar as classes separadas utilizando o cálculo da distância euclidiana entre os seus vizinhos.

A classificação temática para o Parna Catimbau, obtida a partir da modelagem e espacialização dos usos para o período úmido baseando-se nos algoritmos estudados está demonstrada na Figura 12. O algoritmo kNN foi o único que apresentou valores gráficos de classificação próximos à realidade, corroborando as estatísticas realizadas acima, bem como com as conclusões de Deng et al. (2016), onde os resultados experimentais mostraram que a classificação kNN teve um bom desempenho em termos de precisão e eficiência, sendo adequada para lidar com *big data*.

Figura 12 - Espacialização do uso e cobertura da terra a partir dos algoritmos kNN, RF, ANN e SVM para o Vale do Catimbau no período chuvoso



Fonte: O autor.

Para esse período é possível notar que ANN não se mostrou eficiente na identificação da classe Solo exposto, a qual englobou 21,82% da área de estudo (Tabela 3). Este resultado difere do encontrado para o algoritmo kNN (14,24%), que se apresentou como o melhor classificador para o período úmido em regiões de FTS. Outra diferença refere-se ao algoritmo SVM, o qual quantificou apenas 7,51% como sendo solo exposto.

A classe temática Fitofisionomia Arbustiva-Herbácea na classificação pelo kNN englobou uma cobertura de 77,51%, corroborando com a análise visual em campo e das imagens de satélite de alta resolução espacial, uma vez que corresponde ao uso predominante do Parque. Por outro lado, SVM classificou o mesmo uso com um desconto na cobertura com cerca de -17% em relação ao kNN, um total de 60,22%. Não houve alterações significativas, numericamente, nas classes Fitofisionomia Arbóreo-Arbustiva e Áreas Antrópicas.

As discrepâncias nos valores numéricos de cada classe, podem estar associados a forte influência das nuvens (incluídas na classe: áreas não classificadas) (Figura 12). Há uma

predominância de nuvens nas imagens orbitais para a região do Nordeste brasileiro, fator esse que limita a quantidade de imagens disponíveis para estudos dos elementos da superfície.

Tabela 3 – Quantificação das classes temáticas em hectares (ha) e sua representação percentual obtida por meio da classificação supervisionada dos algoritmos de aprendizado de máquina no Parna Catimbau com dados Landsat 8 Sensor OLI para o período úmido.

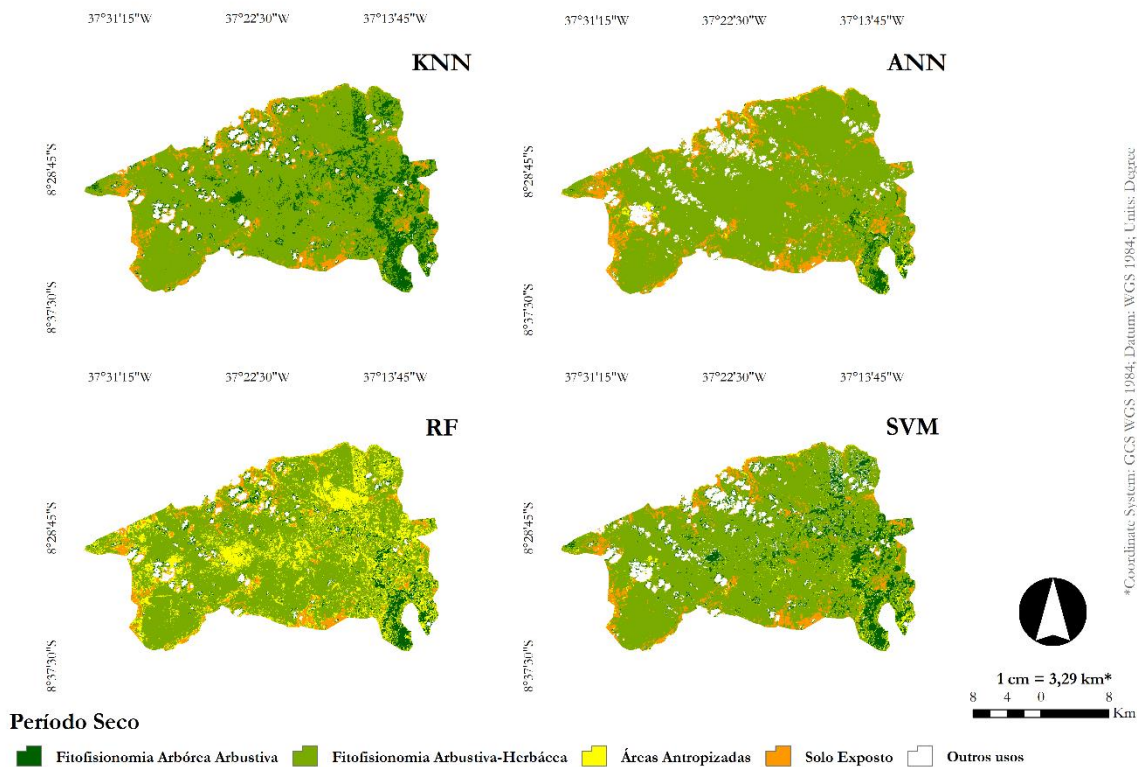
Junho/2019	RF		SVM		ANN		KNN	
Classes temáticas	ha	%	ha	%	ha	%	ha	%
FAA	3599,82	5,78	2779,65	4,46	3670,11	5,89	3366,45	5,41
FAH	47314,35	75,97	37504,89	60,22	41736,96	67,02	48273,66	77,51
AA	464,31	0,75	93,6	0,15	276,57	0,44	206,28	0,33
BS	9889,2	15,88	4677,84	7,51	13590,09	21,82	8865,36	14,24
Outros usos	1010,88	1,62	17222,58	27,65	3004,83	4,82	1566,81	2,52
Total	62278,56	100	62278,56	100	62278,56	100	62278,56	100

Legenda: FAA, Fitofisionomia Arbórea-Arbustiva; FAH, Fitofisionomia Arbustiva-Herbácea; AA, Áreas Antropizadas; BS, Solo Exposto; Outros usos: Áreas de Nuvens, Sombras, Pequenos corpos hídricos.

Fonte: O autor.

Na espacialização dos usos e cobertura da terra para o período seco (Figura 13), observou-se o melhor ajuste dos algoritmos SVM, kNN e ANN na identificação das classes temáticas existentes na área de estudo, os quais corroboram as análises estatísticas demonstradas anteriormente. Destaca-se que a vegetação Arbórea-Arbustiva, correspondente a cobertura de maior densidade de copa, estando situada em sua maioria, na porção leste do Parna.

Figura 13 – Espacialização dos algoritmos de *Machine Learning* kNN, RF, ANN e SVM para o Vale do Catimbau no período seco (dezembro/2019)



Fonte: O autor.

A Fitofisionomia Arbustiva-Herbácea continuou representando a maior cobertura da terra no Parna Catimbau (Tabela 4), demonstrando a eficácia dos algoritmos testados na identificação dessa cobertura no período seco, sendo esse o que engloba maior confusão na identificação dessa vegetação por sensores remotos. Ao comparar os resultados obtidos na quantificação da vegetação Arbustiva-Herbácea nos dois períodos de estudo, observou-se que o kNN (algoritmo de melhor desempenho para o período úmido) englobou uma área total de 77,51% para essa classe, enquanto o SVM (algoritmo de melhor desempenho para o período seco) expressou um total de 69,98%, evidenciando similaridades na identificação dessa cobertura em ambos os períodos pelos algoritmos selecionados.

A diferença na comparação dessas quantificações está associada, em grande parte, ao maior predomínio de nuvens no Parna, interferindo em maior intensidade, na classe de maior abrangência, que correspondeu a vegetação Arbustiva-Herbácea. Do mesmo modo que no período úmido, a imagem contou com a presença de nuvens, reforçando a dificuldade de obtenção de dados para o monitoramento da região sem essa influência, de modo que a escolha das imagens se dá pela menor concentração de nuvens.

Tabela 4 – Quantificação das classes temáticas em hectares (ha) e sua representação percentual obtida por meio da classificação supervisionada dos algoritmos de aprendizado de máquina no Parna Catimbau com dados Landsat 8 Sensor OLI para o período Seco (Dezembro/2019).

Dezembro/2019	RF		SVM		ANN		kNN	
Classes temáticas	ha	%	ha	%	ha	%	ha	%
FAA	3906	6,27	6835,59	10,98	1340,28	2,15	7971,57	12,80
FAH	37903,05	60,86	43580,34	69,98	49104	78,85	46117,62	74,05
AA	14327,82	23,01	2278,8	3,66	867,42	1,39	334,17	0,54
BS	3863,79	6,20	4509,36	7,24	6343,47	10,19	4984,29	8,00
Outros usos	2277,9	3,66	5074,47	8,15	4623,39	7,42	2870,91	4,61
Total	62278,56	100	62278,56	100	62278,56	100	62278,56	100

Legenda: FAA, Fitofisionomia Arbórea-Arbustiva; FAH, Fitofisionomia Arbustiva-Herbácea; AA, Áreas Antropizadas; BS, Solo Exposto; Outros usos: Áreas de Nuvens, Sombras, Pequenos corpos hídricos.

Fonte: O autor.

O Solo Exposto apresentou confusão no processo de classificação com as áreas antropizadas no período seco, fator que ocasionou a quantificação de parte do solo exposto como áreas antropizadas. Assim, o algoritmo RF no período seco acarretou maiores erros na separação das classes, influenciando na espacialização erroneamente, especialmente, da classe Áreas Antropizadas, a qual ocupou 23,01%. Estes valores foram conflitantes com a realidade e com os valores dos outros algoritmos para a mesma classe.

Manejar, administrar e monitorar áreas protegidas, como o caso do Parque Nacional do Catimbau, através de técnicas de sensoriamento remoto, a nível de presença ou não de vegetação, se faz importante para a conservação das condições de manutenção das formas de vida (faunística e florística) destes ambientes (NAGENDRA et al., 2013).

Nesse sentido, as técnicas de sensoriamento remoto aplicadas na solução de problemas no monitoramento de FTS demonstraram resultados promissores, em que mais estudos devem ser realizados a fim de investigar melhor os elementos que influenciam na interação da vegetação com a radiação eletromagnética, possibilitando uma melhor compreensão do comportamento espectral dessas fitofisionomias.

5 CONCLUSÃO

A classificação de imagens oriundas do sensor OLI/Landsat 8 em áreas de floresta tropical sazonalmente seca, como é o caso da vegetação encontrada no Parque Nacional do Catimbau, em Pernambuco, através da utilização de algoritmos de *machine learning* se mostrou como uma alternativa na detecção de mudanças ocorridas no uso e cobertura da terra. Acredita-se que esta será uma contribuição para facilitar a classificação da cobertura vegetal do Bioma Caatinga em todas as épocas do ano, mesmo quando a vegetação esteja em estágio caducifólia, momento em que pode haver dificuldades de diferenciação entre esta e o solo exposto. Desta forma, servirá para o aprimoramento das técnicas de manejo e conservação da vegetação da Caatinga.

No período úmido o algoritmo kNN evidenciou melhores resultados de separação e classificação dos usos, diferindo-se estatisticamente dos demais algoritmos testados. No período seco, este sendo o período motivador para a realização deste estudo, obteve os algoritmos SVM, kNN e ANN como eficientes na identificação da vegetação, especialmente a vegetação com característica arbustiva-herbácea, conseguindo diferenciar esta do solo exposto.

Assim, o kNN se destaca como um importante algoritmo para o mapeamento da cobertura vegetal em regiões do semiárido nos diferentes períodos de disponibilidade hídrica, podendo ser indicado em estudos de monitoramento da vegetação, pois descreve a variação no padrão espectral dos usos e cobertura da terra.

Portanto, o presente estudo trouxe inovação no uso de técnicas de Inteligência Artificial para a solução de problemas no monitoramento, manejo e administração de florestas tropicais sazonalmente secas com dados remotos, sendo um método alternativo para identificar, de forma rápida e econômica, as mudanças na estrutura florestal.

Para próximos estudos recomenda-se (i) a realização de maior amostragem dos usos, de forma sistemática, assim facilitando o treinamento dos algoritmos buscando a diminuição de erros; (ii) a criação de uma rede internacional de estudos de sensoriamento remoto em florestas secas para compartilhamento de dados e monitoramento. Todas essas abordagens devem ser interdisciplinares pois o mapeamento também depende das interações sociais presentes para que seja efetivo, usual e adaptável ao manejo de áreas protegidas assim com aos seus contextos geográficos (iii) realizar a verdade de campo, observando *in situ* o ajuste da classificação realizada pelos algoritmos testados.

REFERÊNCIAS

- ALBA, E. **Influência e análise da cobertura florestal na modificação do albedo com uso de inteligência artificial e sensoriamento remoto**. Tese de Doutorado. UFSM, Santa Maria - RS. 121p. 2020.
- ALMEIDA, R. A.; SANTOS, L. V.; PINTO, D. B. F.; FERRAZ, C. M. L. Temporal analysis of soil occupation in Mucuri watershed between 1989 and 2015. **Revista Engenharia na Agricultura-Reveng**, v. 28, p. 58-68, 2020.
- ALTHOFF, T. D.; MENEZES, R. S. C.; PINTO, A. S.; PAREYN, F. G. C.; CARVALHO, A. L.; MARTINS, J. C. R.; SAMPAIO, E. V. D. S. B. Adaptation of the century model to simulate C and N dynamics of Caatinga dry forest before and after deforestation. **Agriculture, ecosystems & environment**, v. 254, p. 26-34, 2018.
- ANDRADE, E. M.; GUERREIRO, M. J. S.; PALÁCIO, H. A. Q.; CAMPOS, D. A. Ecohydrology in a Brazilian tropical dry forest: thinned vegetation impact on hydrological functions and ecosystem services. **Journal of Hydrology: Regional Studies**, v. 27, p. 100649, 2020.
- ANTONIO SOBRINO, J.; JIMÉNEZ-MUÑOZ, J. C.; PAOLINI, L. Land surface temperature retrieval from LANDSAT TM 5. **Remote Sensing of environment**, v. 90, n. 4, p. 434-440, 2004.
- ARAÚJO, E. L.; CASTRO, C. C.; ALBUQUERQUE, U. P. Dynamics of Brazilian caatinga—A review concerning the plants, environment and people. **Functional Ecosystems and communities**, v. 1, n. 1, p. 15-28, 2007.
- BARBOSA, H. A.; KUMAR, T. L.; PAREDES, F.; ELLIOTT, S.; AYUGA, J. G. Assessment of Caatinga response to drought using meteosat-SEVIRI normalized difference vegetation index (2008–2016). **ISPRS journal of photogrammetry and remote sensing**, v. 148, p. 235-252, 2019.
- BIVAND, R. **RGDAL: Bindings for the Geospatial Data Abstraction Library**. R package version 1.2-6. 2020. Disponível em: '<https://CRAN.R-project.org/package=rgdal>'. Acesso em 16 de nov. 2020.
- BRASIL. Decreto Lei nº 4340, de 22 de agosto de 2002. Regulamenta artigos da Lei no 9.985, de 18 de julho de 2000, que dispõe sobre o Sistema Nacional de Unidades de Conservação da Natureza - SNUC, e dá outras providências.. **Decreto Nº 4.340 de 22 de Agosto de 2002**. Brasília, Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/2002/d4340.htm. Acesso em: 06 set. 2020.
- BRASIL. Lei nº 9985, de 18 de julho de 2000. **Sistema Nacional de Unidades de Conservação da Natureza – Snuc**. Brasília, Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19985.htm. Acesso em: 05 set. 2020
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- BUAINAIN, A. M.; GARCIA, J. R. Pobreza rural e desenvolvimento do semiárido nordestino: resistência, reprodução e transformação. **A nova cara da pobreza rural: desenvolvimento e a questão regional**. Brasília, DF: IICA, p. 217-235, 2013.
- CARNEIRO NETO, J. A.; DIAS, W. R. A.; SANTOS, J. R. V. D.; JESUS, W. B. D. Inteligência Artificial: um estudo guiado por patentes. **IVENPI**. Vol. 4, p. 679. 2018.

- CHAVEZ, P. S. Image-based atmospheric corrections-revisited and improved. **Photogrammetric engineering and remote sensing**, v. 62, n. 9, p. 1025-1035, 1996.
- CONGEDO, L.; MUNAFÒ, M. Development of a methodology for land cover classification in Dar es Salaam using Landsat imagery. **Working Paper, Sapienza University, Rome**, 2012.
- CPRH. AGÊNCIA ESTADUAL DE MEIO AMBIENTE. Monte seu mapa. Recife: Sistema de Informações Geoambientais de Pernambuco (SIG CABURÉ), 2020. Disponível em: <http://sigcabure.cprh.pe.gov.br/>. Acesso em: 25 jun. 2019.
- CUNHA, J.; NÓBREGA, R. L.; RUFINO, I.; ERASMI, S.; GALVÃO, C.; VALENTE, F. Surface albedo as a proxy for land-cover clearing in seasonally dry forests: evidence from the Brazilian Caatinga. **Remote Sensing of Environment**, v. 238, p. 111250, 2020.
- DENG, Z.; ZHU, X.; CHENG, D.; ZONG, M.; ZHANG, S. Efficient kNN classification algorithm for big data. **Neurocomputing**, v. 195, p. 143-148, 2016.
- ECKERT, S.; HÜSLER, F.; LINIGER, H.; HODEL, E. Trend analysis of MODIS NDVI time series for detecting land degradation and regeneration in Mongolia. **Journal of Arid Environments**, v. 113, p. 16-28, 2015.
- FONSECA, R. C.; FREIRE, N. C. F. Evaluation of vegetation cover from NDVI vegetation indexes: possible effects of climate change at Catimbau National Park. **Journal of Hyperspectral Remote Sensing**, v. 7, n. 6, p. 367-377, 2018.
- FRANCH, B.; VERMOTE, E. F.; ANTONIO SOBRINO, J.; FÉDÈLE, E. Analysis of directional effects on atmospheric correction. **Remote Sensing of Environment**, v. 128, p. 276-288, 2013.
- FREIRE, N. C. F.; MOURA, D. C.; SILVA, J. B.; MOURA, A. S. S.; MELO, J. I. M.; NÓBREGA, R.; LUCENA, J.; ARAÚJO, A. P.; LOPES, A.; MENEZES, A.; LIMA, F.; FRANCIS, I.; DANTAS, R. **Mapeamento e Análise Espectro-Temporal das Unidades de Conservação de Proteção Integral da Administração Federal no Bioma Caatinga - Parque Nacional do Catimbau**. Fundação Joaquim Nabuco - Centro Integrado de Estudos Georreferenciados. Recife - PE. 53 p. 2015.
- FREIRE, N. C. F.; MOURA, D. C.; SILVA, J. B.; PACHECO, A. P. Spectro-temporal mapping and analysis of integral protection conservation units of federal administration in the caatinga biome. **Brazilian Journal of Development**, v. 6, n. 5, p. 24773-24781, 2020.
- FRIEDMAN, M. A comparison of alternative tests of significance for the problem of m rankings. **The Annals of Mathematical Statistics**, v. 11, n. 1, p. 86-92, 1940.
- WOLFER M. F. et al. Análise do Uso e Cobertura da Terra Utilizando Imagens Sentinel-2A e Inteligência Artificial. **Anuário do Instituto de Geociências**, v. 43, n. 4, p. 395-403, 2020.
- GLEASON, Colin J.; IM, Jungho. Forest biomass estimation from airborne LiDAR data using machine learning approaches. **Remote Sensing of Environment**, v. 125, p. 80-91, 2012.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. Springer Science & Business Media, 2009.
- HEIN, L.; RIDDER, N.; HIERNAUX, P.; LEEMANS, R.; WIT, A.; SCHAEPMAN, M. Desertification in the Sahel: Towards better accounting for ecosystem dynamics in the

interpretation of remote sensing images. **Journal of Arid Environments**, v. 75, n. 11, p. 1164-1172, 2011.

HIJMANS, R. J. **Raster: Geographic data analysis and modeling**. R package version 3.4-5. Disponível em: '<https://CRAN.R-project.org/package=raster>'. Acesso em 16 de nov. 2020.

HIND, D. J. N.; BAUTISTA, H. P. *Acritopappus buiquensis* (Compositae: Eupatorieae: Ageratinae), a new species from the Brejo forests of Pernambuco State, Brazil. **Key Bulletin**, v. 64, n. 4, p. 701-704, 2009.

HONG, D.; YOKOYA, N.; XIA, G. S.; CHANUSSOT, J.; ZHU, X. X. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 167, p. 12-23, 2020.

ICMBIO. INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE . **Parna Catimbau**. Disponível em: <https://www.icmbio.gov.br/portal/unidadesdeconservacao/biomas-brasileiros/caatinga/unidades-de-conservacao-caatinga/2135-parna-do-Catimbau>. Acesso em: 02 set. 2020.

JIN, Y.; LIU, X.; CHEN, Y.; LIANG, X. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. **International Journal of Remote Sensing**, v. 39, n. 23, p. 8703-8723, 2018.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Ijcai**. 1995. p. 1137-1145.

KOSHALE, J. P.; SINGH, C. **Multi-Temporal Land Use/Land Cover (LULC) Change Analysis Using Remote Sensing and GIS Techniques of Durg Block, Durg District, Chhattisgarh, India**. In Sustainable Development Practices Using Geoinformatics (eds S. Kanga, V.N. Mishra and S.K. Singh). Wiley Library. 2020.

KOURENTZES, N. **TSUTILS: Time Series Exploration, Modelling and Forecasting**. R package version 0.9.2. 2020. Disponível em: '<https://cran.r-project.org/web/packages/tsutils/tsutils.pdf>'. Acesso em 16 de nov. 2020.

KUHN, M. **CARET: Classification and Regression Training**. R package version 6.0-86. 2020. Disponível em: '<https://CRAN.R-project.org/package=caret>'. Acesso em 16 de nov. 2020.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **biometrics**, p. 159-174, 1977.

LI, Y.; LI, M.; LI, C.; LIU, Z. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. **Nature Scientific reports**, v. 10, n. 1, p. 1-12, 2020.

LIAW, A. **Breiman and Cutler's Random Forests for Classification and Regression**. R package version 4.6-14. 2018. Disponível em: '< <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>>'. Acesso em 16 de nov. 2020.

LOUZADA, R. O.; BERGIER, I.; ASSINE, M. L. Landscape changes in avulsive river systems: Case study of Taquari River on Brazilian Pantanal wetlands. **Science of The Total Environment**, v. 723, p. 138067, 2020.

MIURA, T.; HUETE, A.; YOSHIOKA, H. An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 Hyperion data. **Remote Sensing of Environment**, v. 100, n. 2, p. 223-236, 2006.

NAGENDRA, H.; LUCAS, R.; HONRADO, J. P.; JONGMAN, R. H.; TARANTINO, C.; ADAMO, M.; MAIROTA, P. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. **Ecological Indicators**, v. 33, p. 45-59, 2013.

NEMENYI, P. Distribution-free multiple comparisons. In: **Biometrics**. 1441 I ST, NW, SUITE 700, WASHINGTON, DC 20005-2210: **International Biometric Soc**, 1962. p. 263-&.

OLIVEIRA, A. C. P.; NUNES, A.; RODRIGUES, R. G.; BRANQUINHO, C. The response of plant functional traits to aridity in a tropical dry forest. **Science of The Total Environment**, v. 747, p. 141177, 2020.

OTGONBAYAR, M.; ATZBERGER, C.; CHAMBERS, J.; DAMDINSUREN, A. Mapping pasture biomass in Mongolia using Partial Least Squares, Random Forest regression and Landsat 8 imagery. **International Journal of Remote Sensing**, v. 40, n. 8, p. 3204-3226, 2019.

PAHIKKALA, T.; AIROLA, A.; BOBERG, J.; SALAKOSKI, T. Exact and efficient leave-pair-out cross-validation for ranking RLS. **Proceedings of AKRR**, v. 2008, p. 1-8, 2008.

PEREIRA, S., LEAL, I. R., TABARELLI, M., & SANTOS, M. G. Intense mycorrhizal root colonization in a human-modified landscape of the Caatinga dry forest. **Forest Ecology and Management**, v. 462, p. 117970, 2020.

PONTIUS JUNIOR., GILMORE, R.; MILLONES, M.. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, p. 4407-4429, 2011

POURGHASEMI, H. R.; GAYEN, A.; LASAPONARA, R.; TIEFENBACHER, J. P. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. **Environmental research**, p. 109321, 2020.

QGIS. QuantumGis Development Team. **QGIS Geographic Information System. Open source Geospatial Foundation Project**. Disponível em: <<http://www.qgis.org/>>. Acesso em: 10 nov. 2020.

QIAN, Y.; ZHOU, W.; YAN, J.; LI, W.; HAN, L. Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. **Remote Sensing**, v. 7, n. 1, p. 153 - 168, 2015.

RIPLEY, B.; VENABLES, W. **Feed-Forward Neural Networks and Multinomial Log-Linear Models**. R package version 1.5-18. 2020. Disponível em: ‘<https://cran.r-project.org/web/packages/nnet/nnet.pdf>’ . Acesso em 16 de nov. 2020.

RITO, K. F.; ARROYO-RODRÍGUEZ, V.; QUEIROZ, R. T.; LEAL, I. R.; TABARELLI, M. Precipitation mediates the effect of human disturbance on the Brazilian Caatinga vegetation. **Journal of Ecology**, v. 105, n. 3, p. 828-838, 2017.

RODAL, M. J. N.; ANDRADE, K. V. A.; SALES, M. F.; GOMES, A. P. S. Fitossociologia do componente lenhoso de um refúgio vegetacional no município de Buíque, Pernambuco.

Revista Brasileira de Biologia, v. 58, n. 3, p. 517-526, 1998.

ROUSE JUNIOR, J. W.; HAAS, R. H.; DEERING, D. W.; SCHELL, J. A.; HARLAN, J. C. **Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation**. [Great Plains Corridor]. 1974.

SAHAGÚN, E. C.; ANDRADE, J. L.; ZOTZ, G.; REYES-GARCÍA, C. Dew can prolong photosynthesis and water status during drought in some epiphytic bromeliads from a seasonally dry tropical forest. *Tropical Conservation Science*, v. 12, p. 1940082919870054, 2019.

SÁNCHEZ, A. S.; ALMEIDA, M. B.; TORRES, E. A.; KALID, R. D. A.; COHIM, E.; GASPARATOS, A. Alternative biodiesel feedstock systems in the Semi-arid region of Brazil: Implications for ecosystem services. **Renewable and Sustainable Energy Reviews**, v. 81, p. 2744-2758, 2018.

SANO, E. E.; RODRIGUES, A. A.; MARTINS, E. S.; BETTIOL, G. M.; BUSTAMANTE, M. M.; BEZERRA, A. S.; BOLFE, E. L. Cerrado ecoregions: A spatial framework to assess and prioritize Brazilian savanna environmental diversity for conservation. **Journal of environmental management**, v. 232, p. 818-828, 2019.

SANTOS, J. E. O.; NICOLETE, D. A. P.; FILGUEIRAS, R.; LEDA, V. C.; ZIMBACK, C. R. L. Imagens do Landsat-8 no mapeamento de superfícies em área irrigada. **Irriga**, v. 1, n. 2, p. 30-36, 2015.

SANTOS, J. M.; PESSOA, M. M. L.; FERREIRA, R. L. C.; SILVA, E. A. Land-use and coverage in the structure of the landscape in a Tropical Dry Forest in northeast Brazil. **Journal of Environmental Analysis and Progress**, v. 5, n. 1, p. 088-097, 2020.

SCHRÖDER, J. M.; RODRÍGUEZ, L. P. Á.; GÜNTER, S. Research trends: Tropical dry forests: The neglected research agenda?. **Forest Policy and Economics**, v. 122, 2021.

SEGAL, M. R. Machine learning benchmarks and random forest regression. **UCSF: Center For Bioinformatics and Molecular Biostatistics**. 2004.

SHAO, Y.; LUNETTA, R. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 70, p. 78-87, 2012.

SILVA, J. F.; CICERELLI, R. E.; ALMEIDA, T.; NEUMANN, M. R. B.; SOUZA, A. L. F. Land use/cover (LULC) mapping in Brazilian Cerrado using neural network with Sentinel-2 data. **Floresta**, v. 50, n. 3, p. 1430-1438, 2020.a

SILVA, J. L. S. E.; CRUZ-NETO, O.; PERES, C. A.; TABARELLI, M.; LOPES, A. V. Climate change will reduce suitable Caatinga dry forest habitat for endemic plants with disproportionate impacts on specialized reproductive strategies. **PloS one**, v. 14, n. 5, p. e0217028, 2019.

SILVA, J. M. C.; LEAL, I. R.; TABARELLI, M. (Ed.). **Caatinga: the largest tropical dry forest region in South America**. Springer, 2018.

SILVA, K. A.; EL-DEIR, S. G.; MONTEIRO JUNIOR, J. J.; SANTOS, J. P. O.; SILVA, E. A. Analysis of vegetation dynamics using the normalized difference vegetation index (NDVI) at the archipelago of Fernando de Noronha, Pernambuco, Brazil. **Interações (Campo Grande)**, p. 885-901, 2020.b

SILVA, L. F.; FREIRE, K. T.; ARAÚJO-MAGALHÃES, G. R.; AGAMEZ-MONTALVO, G. S.; SOUSA, M. A.; COSTA-SILVA, T. A.; SOUZA-MOTTA, C. M. *Penicillium* and *Talaromyces endophytes* from *Tillandsia Catimbauensis*, a bromeliad endemic in the Brazilian tropical dry forest, and their potential for L-asparaginase production. **World Journal of Microbiology and Biotechnology**, v. 34, n. 11, p. 162, 2018.

SILVA, L. P.; XAVIER, A. P. C.; SILVA, R. M.; SANTOS, C. A. G. Modeling land cover change based on an artificial neural network for a semiarid river basin in northeastern Brazil. **Global Ecology and Conservation**, v. 21, p. e00811, 2020.

SILVA, M. V.; PANDORFI, H.; LOPES, P. M. O.; SILVA, J. L. B.; ALMEIDA, G. L. P.; SILVA, D. A. O.; JARDIM, A. M. D. R. F. Pilot monitoring of Caatinga spatial-temporal dynamics through the action of agriculture and livestock in the Brazilian semiarid. **Remote Sensing Applications: Society and Environment**, v. 19, p. 100353, 2020.

SILVERO, N. E. Q.; DEMATTÊ, J. A. M.; AMORIM, M. T. A.; SANTOS, N. V.; RIZZO, R.; SAFANELLI, J. L.; BONFATTI, B. R. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. **Remote Sensing of Environment**, v. 252, p. 112117, 2021.

SINGH, J. S.; CHATURVEDI, R. K. **Tropical dry deciduous forest: Research trends and emerging features**. Springer, 2018.

SOUZA SOBRINHO, M. S.; TABARELLI, M.; MACHADO, I. C.; SFAIR, J. C.; BRUNA, E. M.; LOPES, A. V. Land use, fallow period and the recovery of a Caatinga forest. **Biotropica**, v. 48, n. 5, p. 586-597, 2016.

SOUZA, C. G.; CARVALHO, L.; AGUIAR, P.; ARANTES, T. B. Algoritmos de aprendizagem de máquina e variáveis de sensoriamento remoto para o mapeamento da cafeicultura. **Boletim de Ciências Geodésicas**, v. 22, n. 4, p. 751-773, 2016.

SOUZA-SILVA, R. F.; RAPINI, A.; MORALES, J. F. Mandevilla *Catimbauensis* (Apocynaceae), a new species from the semi-arid region, Pernambuco, Brazil. **Edinburgh Journal of Botany**, v. 67, n. 1, p. 1, 2010.

TAVARES JÚNIOR, I. S.; TORRES, C. M. M. E.; LEITE, H. G.; CASTRO, N. L. M.; SOARES, C. P. B.; CASTRO, R. V. O.; FARIAS, A. A. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. **Ecological Indicators**, v. 117, p. 106685, 2020.

TRUE, S. A. Planning the future of the World Geodetic System 1984. In: **PLANS 2004. Position Location and Navigation Symposium (IEEE Cat. No. 04CH37556)**. IEEE, 2004. p. 639-648.

USGS. UNITED STATES GEOLOGICAL SURVEY. **Earth Explorer**. 2020. Disponível em: <<https://earthexplorer.usgs.gov/>>. Acesso em: 15 mar. 2020.a

USGS. UNITED STATES GEOLOGICAL SURVEY. **What are the band designations for the Landsat Satellites?** 2020. Disponível em: <https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products>. Acesso em: 10 nov. 2020.b

WICKHAM, H. **ggplot2: elegant graphics for data analysis**. Springer, 2016.

XI, Z.; HOPKINSON, C.; ROOD, S. B.; PEDDLE, D. R. See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species

classification from terrestrial laser scanning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 168, p. 1-16, 2020.

YANG, Y.; WANG, Z.; LI, J.; GANG, C.; ZHANG, Y.; ZHANG, Y.; ODEH, I.; QI, J. Comparative assessment of grassland degradation dynamics in response to climate variation and human activities in China, Mongolia, Pakistan and Uzbekistan from 2000 to 2013. **Journal of Arid Environments**, v. 135, p. 164-172, 2016.

ZHAO, Q.; YU, S.; ZHAO, F.; TIAN, L.; ZHAO, Z. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. **Forest Ecology and Management**, v. 434, p. 224-234, 2019.