



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21d/2024/03.21.14.14-TDI

**INTEGRAÇÃO DE ÍNDICES ESPECTRAIS E  
PRODUTOS METEOROLÓGICOS PARA CLASSIFICAR  
USO E COBERTURA DA TERRA USANDO SÉRIES  
TEMPORAIS DE IMAGENS DE SATÉLITE E MACHINE  
LEARNING**

Pedro Vinícius da Silva Brito

Dissertação de Mestrado do Curso  
de Pós-Graduação em Computação  
Aplicada, orientada pelos Drs.  
Karine Reis Ferreira Gomes, e  
Rafael Duarte Coelho dos Santos,  
aprovada em 13 de março de 2024.

URL do documento original:

<http://urlib.net/8JMKD3MGP3W34T/4AUKGUP>

INPE  
São José dos Campos  
2024

**PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE  
Coordenação de Ensino, Pesquisa e Extensão (COEPE)  
Divisão de Biblioteca (DIBIB)  
CEP 12.227-010  
São José dos Campos - SP - Brasil  
Tel.:(012) 3208-6923/7348  
E-mail: pubtc@inpe.br

**CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):**

**Presidente:**

Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra (CGCT)

**Membros:**

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)  
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e Ciência Espaciais (CGCE)  
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e Pesquisas Aplicadas (CGIP)  
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

**BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon  
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

**REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)  
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

**EDITORAÇÃO ELETRÔNICA:**

Ivone Martins - Divisão de Biblioteca (DIBIB)  
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21d/2024/03.21.14.14-TDI

**INTEGRAÇÃO DE ÍNDICES ESPECTRAIS E  
PRODUTOS METEOROLÓGICOS PARA CLASSIFICAR  
USO E COBERTURA DA TERRA USANDO SÉRIES  
TEMPORAIS DE IMAGENS DE SATÉLITE E MACHINE  
LEARNING**

Pedro Vinícius da Silva Brito

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Karine Reis Ferreira Gomes, e Rafael Duarte Coelho dos Santos, aprovada em 13 de março de 2024.

URL do documento original:

<http://urlib.net/8JMKD3MGP3W34T/4AUKGUP>

INPE  
São José dos Campos  
2024

Dados Internacionais de Catalogação na Publicação (CIP)

---

Brito, Pedro Vinícius da Silva .

B777i      Integração de índices espectrais e produtos meteorológicos para classificar uso e cobertura da terra usando séries temporais de imagens de satélite e machine learning / Pedro Vinícius da Silva Brito. – São José dos Campos : INPE, 2024.  
xxii + 115 p. ; (sid.inpe.br/mtc-m21d/2024/03.21.14.14-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2024.

Orientadores : Drs. Karine Reis Ferreira Gomes, e Rafael Duarte Coelho dos Santos.

1. Séries temporais. 2. Cubo de dados. 3. Aprendizado de máquina. 4. Uso e cobertura da terra. 5. Sensoriamento Remoto.  
I.Título.

CDU 681.3.019:528.8

---



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



## INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

### DEFESA FINAL DE DISSERTAÇÃO PEDRO VINÍCIUS DA SILVA BRITO BANCA Nº 028/2024, REG. 428492/2022

No dia 13 de março de 2024, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Computação Aplicada, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

**Novo Título: "Integração de índices espectrais e produtos meteorológicos para classificar uso e cobertura da terra usando séries temporais de imagens de satélite e machine learning"**

#### Membros da Banca:

Dr. Pedro Ribeiro de Andrade Neto – Presidente – INPE

Dra. Karine Reis Ferreira Gomes – Orientadora – INPE

Dr. Rafael Duarte Coelho dos Santos – Orientador – INPE

Dr. Michel Eustáquio Dantas Chaves – Membro Externo – UNESP/Campus Tupã

Dra. Magna Soelma Beserra de Moura – Membro Externo – Embrapa



Documento assinado eletronicamente por **Pedro Ribeiro de Andrade Neto, Tecnologista**, em 15/03/2024, às 08:05 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael Duarte Coelho dos Santos, Pesquisadora**, em 15/03/2024, às 08:35 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Karine Reis Ferreira Gomes, Chefe da Divisão de Observação da Terra e Geoinformática**, em 15/03/2024, às 09:09 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Michel Eustáquio Dantas Chaves (E), Usuário Externo**, em 18/03/2024, às 08:54 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Magna soelma beserra de moura (E)**, **Usuário Externo**, em 18/03/2024, às 10:15 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site <https://sei.mcti.gov.br/verifica.html>, informando o código verificador **11776677** e o código CRC **642928B0**.

---

Referência: Processo nº 01340.001988/2024-11

SEI nº 11776677

*“Paciência e perseverança tem o efeito mágico de fazer as dificuldades desaparecerem e os obstáculos sumirem”.*

*JOHN QUINCY ADAMS*





*A meus pais **Cicero e Maria**, à minhas irmãs **Alaide e**  
**Andressa***



## AGRADECIMENTOS

Esses anos de mestrado foram marcados por muitos desafios, construções, conquistas e amadurecimento. Aprendi que nem tudo acontece como planejado, mas, tudo tem um motivo de acontecer. O caminho até aqui não foi fácil, porém, com perseverança e uma grande pitada de fé, nunca pareceu impossível.

Gostaria de começar agradecendo a Deus por me dar força e ânimo para seguir na busca dos meus objetivos, tranquilidade nos momentos de aflição e colocar pessoas incríveis no meu caminho, que me deram oportunidades, contribuíram para o meu conhecimento e dispuseram as suas amizades.

Agradeço aos meus pais, Maria Valdejane e Cicero Figueiredo, os maiores mestres da minha vida, que nunca deixaram de acreditar em mim, e que mesmo diante das dificuldades, mantiveram a fé na minha formação e mostraram a importância do estudo, mesmo não tendo eles as mesmas oportunidades no passado.

A minha tia, Maria das Graças, que sempre esteve presente como uma segunda mãe.

A todos os meus familiares, em especial as minhas irmãs Alaide Cristina e Andressa Karla, que foram as minhas primeiras inspirações no estudo, por todo apoio e acreditar em minha jornada acadêmica e meu irmão João Vitor, que sempre me deu apoio e vibra em todas as minhas conquistas.

À minha namorada, Herica Carvalho, que me motivou nos momentos difíceis dessa caminhada, foi companheira e compreensível nos momentos que estive ausente.

Aos orientadores Dra. Karine Reis Ferreira e Dr. Rafael Duarte Coelho dos Santos pelas oportunidades, confiança, ensinamentos e paciência.

Ao professor, Gilberto Ribeiro de Queiroz, pelos conselhos, direcionamentos e confiança.

Ao Felipe Souza e Michel Chaves, pelo tempo disponibilizado, trocas de conhecimentos, paciência e apoio nos momentos de dificuldades.

Ao Baggio Silva e Juliana Balera, por compartilhar seus conhecimentos e experiências, os quais foram fundamentais para os meus primeiros passos.

A Fernanda Paiva, Jéssica Barbosa, Jonas Oliveira, Mateus Miranda e Rafael Marinho, que sempre se mostraram prestativos e presentes nos meus momentos de

dificuldade e celebraram os de felicidade.

A todas as pessoas que, ao longo da minha trajetória, se tornaram amigos.

Ao Instituto Nacional de Pesquisas Espaciais e, em especial, a Pós-Graduação do INPE, seus professores, estudantes e técnicos, pelo ambiente acolhedor e conhecimentos compartilhados.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro da bolsa concedida durante os dois anos do curso.

## RESUMO

As mudanças de uso e cobertura da Terra (em inglês, *Land Use and Land Cover*, LULC) impulsionam diversos impactos ambientais. A combinação de bandas e índices espectrais derivados de imagens de satélites de observação da Terra e de produtos meteorológicos podem melhorar a identificação de padrões de LULC. A organização desses dados em cubos de observação da Terra (em inglês, *Earth Observation*, EO) permite a extração de séries temporais de imagens de satélite (em inglês, *Satellite Image Time Series*, SITS). Essas séries temporais tem sido amplamente usadas com métodos de aprendizado de máquina (em inglês, *Machine Learning*, ML) para produzir mapas de LULC a partir de imagens de satélite. Portanto, este estudo tem como objetivo avaliar a combinação de diferentes bandas, índices espectrais e produtos meteorológicos para melhorar a precisão de classificação de mapas de LULC do município de Petrolina usando SITS e algoritmos de ML. Amostras de LULC para treinar os classificadores e validar os mapas classificados foram coletadas utilizando a plataforma *TerraCollect*. Dados de bandas, índices espectrais e produtos meteorológicos foram combinados usando a abordagem *early fusion*. Após essa etapa, as séries temporais das amostras de LULC foram extraídas e avaliadas utilizando *Self-Organizing Maps* (SOM). Os modelos *Random Forest* (RF) e *Temporal Convolutional Neural Network* (TempCNN) foram ajustados, treinados e as imagens classificadas. A precisão dos mapas classificados foi avaliada utilizando a técnica de área ponderada. De acordo com os resultados observou-se que, as maiores acurácias globais foram registradas no quarto experimento, no qual ambos os modelos RF e TempCNN atingiram 100% de precisão em ambos os anos. A maior diferença nos valores de acurácia global entre os modelos RF e TempCNN foi de 5% no primeiro experimento em 2018/2019 utilizando apenas bandas espectrais. Neste caso, o modelo RF obteve 99% e o TempCNN 94%. As classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, tanto com o modelo RF quanto com o TempCNN com exceção do primeiro experimento, e a classe de Vegetação arbórea/arbustiva nos experimentos 04, 06 e 07. Por outro lado, a classe de Área construída/Urbana obteve os menores valores percentuais de acurácia do produtor e usuário na maioria dos experimentos alcançando apenas 51% e 44% na acurácia do produtor no segundo experimento de 2018/2019 nos modelos RF e TempCNN, respectivamente. Conclui-se que as combinações de bandas, índices espectrais e produtos meteorológicos podem aumentar o desempenho das classificações de LULC no município de Petrolina. As classes de Agropecuária/Solo exposto e Área construída/Urbana, que são historicamente desafiadoras em termos de separabilidade na literatura, demonstraram acurácias iguais ou superiores a 83% e 74% (para a acurácia do produtor em 2019/2020 e 2017/2018, respectivamente) ao usar o modelo TempCNN, em todos os experimentos. Além disso, todas as classes analisadas apresentaram acurácias satisfatórias com a inclusão de dados meteorológicos em todos os experimentos.

Palavras-chave: Séries temporais. Cubo de dados. Aprendizado de máquina. Uso e cobertura da terra. Sensoriamento Remoto.



# INTEGRATION OF SPECTRAL INDICES AND METEOROLOGICAL PRODUCTS FOR LAND USE AND LAND COVER CLASSIFICATION USING TIME SERIES OF SATELLITE IMAGES AND MACHINE LEARNING

## ABSTRACT

Changes in land use and land cover (LULC) drive various environmental impacts. The combination of bands and spectral indices derived from Earth observation satellite images and meteorological products can improve the identification of LULC patterns. The organization of these data into Earth Observation (EO) cubes enables the extraction of Satellite Image Time Series (SITS). These time series have been widely used with machine learning (ML) methods to produce LULC maps from satellite images. Therefore, this study aims to evaluate the combination of different bands, spectral indices and meteorological products to improve the classification accuracy of LULC maps of the municipality of Petrolina using SITS and ML algorithms. LULC samples to train the classifiers and validate the classified maps were collected using the *TerraCollect* platform. Band data, spectral indices and meteorological products were combined using the *early fusion* approach. After this step, the time series of the LULC samples were extracted and evaluated using Self-Organizing Maps (SOM). The Random Forest (RF) and Temporal Convolutional Neural Network (TempCNN) models were adjusted, trained, and the images classified. The accuracy of the classified maps was assessed using the weighted area technique. According to the results, the highest overall accuracies were recorded in the fourth experiment, in which both the RF and TempCNN models achieved 100% accuracy in both years. The largest difference in global accuracy values between the RF and TempCNN models was 5% in the first experiment in 2018/2019 using only spectral bands. In this case, the RF model obtained 99% and TempCNN 94%. The Agriculture and Water bodies classes achieved 100% producer and user accuracy in all years, both with the RF model and TempCNN except for the first experiment, and the Tree/shrub vegetation class in experiments 04, 06 and 07. On the other hand, the Built-up Area/Urban class obtained the lowest percentage accuracy values for the producer and user in most experiments, reaching only 51% and 44% in the producer's accuracy in the second experiment of 2018/2019 in the RF and TempCNN models, respectively. It is concluded that combinations of bands, spectral indices and meteorological products can increase the performance of LULC classifications in the municipality of Petrolina. The Agricultural/Exposed Soil and Built-up Area/Urban classes, which are historically challenging in terms of separability in the literature, showed accuracies equal to or greater than 83% and 74% (for the producer's accuracy in 2019/2020 and 2017/2018, respectively) when using the TempCNN model in all experiments. In addition, all the classes analyzed showed satisfactory accuracies with the inclusion of meteorological data in all the experiments.

Keywords: Time series. Data cube. Machine learning. land use and land cover. Remote Sensing.





## LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Representação simplificada do modelo <i>Random Forest</i> (RF). . . . .	19
2.2 Representação simplificada do modelo <i>Temporal Convolutional Neural Network</i> (TempCNN). . . . .	23
3.1 Localização da área de estudo e distribuição de amostras de uso e cobertura da Terra obtidas pela plataforma TerraCollect. . . . .	25
3.2 Fluxograma das etapas metodológicas do trabalho. . . . .	26
3.3 Nível de estratégia de fusão. (A) Nenhuma fusão e (B) <i>Early fusion</i> . . . . .	30
4.1 Distribuição da precipitação para o período de 2017 a 2020 estimada pelo IMERG para o município de Petrolina-PE. (A) Anual e (B) Mensal . . . . .	36
4.2 Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2017/2018. . . . .	37
4.3 Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2018/2019. . . . .	43
4.4 Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2019/2020. . . . .	46
4.5 Porcentagem de confusão entre os grupos do agrupamento SOM com grade 15 × 15 para 2017/2018. . . . .	50
4.6 Porcentagem de confusão entre os grupos do agrupamento SOM com grade 15 × 15 para 2018/2019. . . . .	51
4.7 Porcentagem de confusão entre os grupos do agrupamento SOM com grade 15 × 15 para 2019/2020. . . . .	54
4.8 Classificações de uso e cobertura da Terra para 2017/2018 via RF para Petrolina-PE. . . . .	61
4.9 Classificações de uso e cobertura da Terra para 2018/2019 via RF para Petrolina-PE. . . . .	65
4.10 Classificações de uso e cobertura da Terra para 2019/2020 via RF para Petrolina-PE. . . . .	67
4.11 Classificações de uso e cobertura da Terra para 2017/2018 via TempCNN para Petrolina-PE. . . . .	69
4.12 Classificações de uso e cobertura da Terra para 2018/2019 via TempCNN para Petrolina-PE. . . . .	71
4.13 Classificações de uso e cobertura da Terra para 2019/2020 via TempCNN para Petrolina-PE. . . . .	73



## LISTA DE TABELAS

	<u>Pág.</u>
3.1 Bandas e índices espectrais utilizados nos experimentos. . . . .	28
3.2 Variáveis meteorológicas utilizadas nos experimentos. . . . .	29
3.3 Sequências dos experimentos realizados no estudo. . . . .	32
4.1 Validação <i>k-fold</i> nos conjuntos de treinamento dos experimentos com os modelos RF e TempCNN. . . . .	56
4.2 <i>F1-scores</i> no conjunto de treinamento dos experimentos com os modelos RF. . . . .	58
4.3 <i>F1-scores</i> no conjunto de treinamento dos experimentos com os modelos TempCNN. . . . .	60
4.4 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 01. . . . .	77
4.5 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 02. . . . .	79
4.6 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 03. . . . .	80
4.7 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 04. . . . .	81
4.8 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 05. . . . .	82
4.9 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 06. . . . .	84
4.10 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 07. . . . .	85
4.11 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 08. . . . .	86
4.12 Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 09. . . . .	88



## LISTA DE ABREVIATURAS E SIGLAS

ANN	–	<i>Artificial Neural Networks</i>
AP	–	Acurácia do produtor
ok AU	–	Acurácia do usuário
ok AG	–	Acurácia global
ok AI	–	<i>Artificial Intelligence</i>
AGDC	–	<i>Australian Geoscience Data Cube</i>
API	–	<i>Application Programming Interface</i>
ARD	–	<i>Analysis-Ready Data</i>
ARDC	–	<i>Africa Regional Data Cube</i>
BDC	–	<i>Brazil Data Cube</i>
BN	–	<i>Batch Normalization</i>
BSI	–	<i>Bare Soil Index</i>
CNNs	–	<i>Convolutional Neural Networks</i>
Conv	–	Camadas convolucionais
DL	–	<i>Deep Learning</i>
DO	–	<i>Dropout</i>
DT	–	<i>Decision Trees</i>
ECMWF	–	<i>European Centre for Medium-Range Weather Forecasts</i>
EO	–	<i>Earth Observation</i>
ERDAS	–	<i>Earth Resource Data Analysis System</i>
EVI	–	<i>Enhanced Vegetation Index</i>
FC	–	<i>Fully-Connected</i>
GFS	–	<i>Global Forecast System</i>
GNDVI	–	<i>Green Normalized Difference Vegetation Index</i>
GPU <sub>s</sub>	–	<i>Graphic Processing Units</i>
GPM	–	<i>Global Precipitation Measurement</i>
IMERG	–	<i>Integrated Multi-Satellite Retrievals for Global Precipitation Measurement</i>
INPE	–	Instituto Nacional de Pesquisas Espaciais
kNN	–	<i>k-Nearest Neighbor</i>
LULC	–	<i>Land Use and Land Cover</i>
ML	–	<i>Machine Learning</i>
ML	–	<i>Maximum Likelihood</i>
MNDWI	–	<i>Modified Normalized Difference Water Index</i>
MS	–	<i>Multi-Spectral</i>
MSAVI	–	<i>Modified Soil Adjusted Vegetation Index</i>
NCEP	–	<i>National Centers for Environmental Prediction</i>
NDVI	–	<i>Normalized Difference Vegetation Index</i>
NDBI	–	<i>Normalized Density Building Index</i>
NN	–	<i>Neural Networks</i>
ODC	–	<i>Open Data Cube</i>
P	–	Precipitação
ReLU	–	<i>Rectified Linear Unit</i>
RF	–	<i>Random Forest</i>

RNN	–	<i>Recurrent Neural Network</i>
SAVI	–	<i>Soil Adjusted Vegetation Index</i>
SDC	–	<i>Swiss Data Cube</i>
SOM	–	<i>Self-Organizing Maps</i>
SVM	–	<i>Support Vector Machines</i>
SITS	–	<i>Satellite Image Time Series</i>
TempCNN	–	<i>Temporal Convolutional Neural Network</i>
T ar	–	Temperatura do ar
T solo	–	Temperatura do solo
TPUs	–	<i>Tensor Processing Units</i>
UR ar	–	Umidade relativa do ar
U solo	–	Umidade do solo
VI	–	<i>Vegetation Index</i>

## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO</b> . . . . .	<b>1</b>
1.1 Objetivos . . . . .	4
1.2 Contribuições . . . . .	4
<b>2 REFERENCIAL TEÓRICO</b> . . . . .	<b>7</b>
2.1 O bioma Caatinga . . . . .	7
2.2 Cubo de dados de observação da Terra . . . . .	11
2.3 Dados de séries temporais de imagens de satélites de observação da Terra	12
2.3.1 Índices espectrais e variáveis meteorológicas . . . . .	14
2.4 Aprendizado de máquina . . . . .	17
2.4.1 <i>Random Forest</i> (RF) . . . . .	18
2.5 Aprendizado profundo . . . . .	21
2.5.1 <i>Temporal Convolutional Neural Network</i> (TempCNN) . . . . .	22
<b>3 METODOLOGIA</b> . . . . .	<b>25</b>
3.1 Localização e caracterização da área de estudo . . . . .	25
3.2 Coleta e processamento dos dados . . . . .	26
3.2.1 Amostras de uso e cobertura . . . . .	27
3.2.2 Dados de satélite . . . . .	27
3.2.3 Dados meteorológicos . . . . .	28
3.2.4 Estratégia de fusão . . . . .	30
3.2.5 Combinações de experimentos . . . . .	31
3.2.6 Avaliação de amostras de uso e cobertura . . . . .	31
3.2.7 Classificação e avaliação dos mapas . . . . .	33
<b>4 RESULTADOS E DISCUSSÃO</b> . . . . .	<b>35</b>
4.1 Precipitação estimada pelo IMERG . . . . .	35
4.2 Avaliação das amostras . . . . .	36
4.3 Confusão entre grupos . . . . .	49
4.4 Ajuste dos modelos . . . . .	55
4.5 Classificação dos mapas de LULC . . . . .	61
4.6 Avaliação dos mapas . . . . .	77

<b>5 CONCLUSÕES</b> . . . . .	<b>89</b>
5.1 Trabalhos futuros . . . . .	90
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> . . . . .	<b>91</b>
<b>APÊNDICE A</b> . . . . .	<b>111</b>
A.1 Uso de séries temporais para classificações de uso e cobertura da terra em Petrolina, Pernambuco . . . . .	111



# 1 INTRODUÇÃO

As mudanças de uso e cobertura da Terra impulsionam diferentes regiões do planeta a climas extremos, extinção de espécies, diminuição de serviços ecossistêmicos e degradação de recursos naturais. Esses impactos afetam principalmente a humanidade, especialmente pessoas de baixo poder aquisitivo e vulneráveis, que dependem dos recursos naturais para sua subsistência (ROY et al., 2022). Os principais causadores das mudanças de uso da Terra incluem o crescimento populacional, degradação florestal e desmatamento, expansão de áreas agrícolas, conversão de paisagens naturais e uso excessivo dos recursos naturais (KOBAYASHI et al., 2020; WANG; YANG, 2020; MALEDE et al., 2023).

Uma maneira de quantificar os efeitos desses impactos é por meio de mapas de uso e cobertura da Terra (em inglês, *Land Use and Land Cover*, LULC) atualizados e precisos (PELLETIER et al., 2019). Para isso, uma das ferramentas capazes de imagear grandes áreas da superfície terrestre são os sensores remotos acoplados em satélites. Eles fornecem produtos de forma contínua e consistente ao longo do tempo. Atualmente, tem-se o livre acesso ao crescente volume de dados de sensoriamento remoto provenientes de diversos satélites e sensores com diferentes resoluções espacial, espectral e temporal (PELLETIER et al., 2019; FERREIRA et al., 2020; SIMOES et al., 2021). Segundo Soille et al. (2018), apenas no ano de 2019, os satélites Landsat-7/8, Terra/Aqua, CBERS-4/4A e Sentinel-1/2/3 geraram aproximadamente cinco petabytes de imagens.

Para analisar dados de observação da Terra de uma mesma localização ao longo do tempo, imagens de satélites têm sido modeladas e organizadas como cubos de dados (NATIVI et al., 2017). Os cubos de dados de EO são matrizes multidimensionais com dimensões de espaço e tempo, compostos por imagens de satélites prontas para análise (em inglês, *analysis ready data*, ARD) (LEWIS et al., 2017). Tal organização permite a extração de séries temporais de imagens de satélite (em inglês, *Satellite Image Time Series*, SITS), sendo úteis para analisar a dinâmica ambiental de um determinado alvo ao longo do tempo permitindo uma visão ampla de LULC (PICOLI et al., 2018; FERREIRA et al., 2020; SIMOES et al., 2021).

Dados ARD são resultantes do processamento de imagens de satélite, desde a aquisição dos dados até a calibração radiométrica. Esse processo envolve etapas adicionais de conversão para determinar a refletância do topo da atmosfera (em inglês, *top-of-atmosphere*, TOA) e, em seguida, a refletância da superfície (GIULIANI et al., 2017). Atualmente, existem várias iniciativas para estruturar cubos de dados de EO

com diferentes escalas espaciais em várias regiões do mundo, como *Swiss Data Cube* (SDC) (GIULIANI et al., 2017), *Australian Geoscience Data Cube* (AGDC) (LEWIS et al., 2017), *Africa Regional Data Cube* (ARDC) (KILLOUGH, 2019) e *Brazil Data Cube* (BDC) (FERREIRA et al., 2020).

O projeto BDC está sendo desenvolvido e administrado na infraestrutura do Instituto Nacional de Pesquisas Espaciais (INPE) desde 2019. É a principal iniciativa para estruturar cubos de dados de EO no Brasil. Um dos objetivos do BDC é gerar coleções de imagens ARD e estruturar como cubo de dados multidimensionais integrando espaço, tempo e propriedades derivadas de espectros. O BDC utiliza e desenvolve tecnologias de *big data* para gerar, armazenar e processar cubos de dados estruturados a partir de imagens ARD. Além disso, gera mapas de LULC para o Brasil por meio de cubos de dados de EO aplicando procedimentos de processamento de imagens, análise de SITS e técnicas de aprendizado de máquina (em inglês, *Machine Learning*, ML) (FERREIRA et al., 2020).

Uma das implementações de *software* disponibilizada e gerenciada pelo BDC para processamento de cubo de dados é o pacote SITS, desenvolvido na linguagem *R* e de código aberto (SIMOES et al., 2021). O pacote SITS dispõe de um conjunto de funções para análise e classificação de SITS a partir de métodos de ML e aprendizado profundo (em inglês, *Deep Learning*, DL) utilizando a abordagem tempo-primeiro (em inglês, *Time-First*) e espaço-depois (em inglês, *Space-Later*) (CAMARA et al., 2016; SIMOES et al., 2021). Entre os modelos disponíveis neste pacote estão o *Random Forest* (RF) (BREIMAN, 2001) e o *Temporal Convolutional Neural Network* (TempCNN) (PELLETIER et al., 2019).

O modelo RF tem como principal vantagem seu alto grau de precisão na seleção de recursos para melhorar a classificação de mapas de LULC. Seu desempenho é semelhante a classificadores tradicionais, como *Decision Trees* (DT) (FRIEDL; BRODLEY, 1997) e *Support Vector Machines* (SVM) (VAPNIK, 1999), enquanto requer menos tempo computacional (INGLADA et al., 2015; BELGIU; DRĂGUȚ, 2016; PELLETIER et al., 2016). Embora o RF e outros algoritmos tradicionais demonstrem bons desempenhos em classificações de LULC, eles não utilizam ao máximo o domínio temporal do SITS (GÓMEZ et al., 2016; PELLETIER et al., 2019).

O modelo TempCNN, também amplamente reconhecido na geração de mapas de LULC a partir da classificação de SITS, é capaz de capturar padrões espaciais aplicando convoluções na dimensão temporal para aprender os recursos temporais e espectrais (BENGIO et al., 2013; ZHU et al., 2017; PELLETIER et al., 2019).

Com base em um conjunto de experimentos realizados por Pelletier et al. (2019), observou-se que esse modelo supera os resultados obtidos com os algoritmos RF e *Recurrent Neural Network* (RNN), com diferença de acurácia global de até 3%. De acordo com os autores, a análise visual dos mapas gerados utilizando este algoritmo demonstra seu potencial no mapeamento preciso de cobertura da Terra sem a super-representação de classes majoritárias.

As vantagens de utilizar os modelos de ML e DL, especialmente ao lidar com a complexidade e a variabilidade dos ambientes terrestres, são sua capacidade de extrair características complexas e padrões de grandes conjuntos de dados espaciais (GANJIRAD; BAGHERI, 2024). Para isso, exige aquisição e processamento de um volume significativo de imagens, porém, utilizando apenas bandas espectrais pode não ser suficiente para separar classes de uso e cobertura da terra. Dessa forma, a inclusão de índices espectrais pode realçar características específicas das diferentes classes e enriquecer o conjunto de dados de treinamento do algoritmo de classificação (GANJIRAD; BAGHERI, 2024).

Além disso, sabe-se que a variabilidade das variáveis meteorológicas como temperatura, precipitação, velocidade do vento, dentre outras, está associada às LULC (AIK et al., 2020; LI et al., 2018). De acordo com Ganjirad e Bagheri (2024), a redução das temperaturas locais é influenciada pela dinâmica das classes de uso e cobertura da Terra, como zonas úmidas, florestas caducifólias, corpos hídricos e áreas agrícolas. No entanto, áreas urbanizadas e zonas de tráfego contribuem para o aumento da temperatura. De acordo com este mesmo autor, a velocidade máxima do vento diário varia conforme o nível de desenvolvimento da área urbana. Partindo desta premissa, o uso de produtos meteorológicos pode ajudar a detectar e distinguir diferentes padrões de uso e cobertura da Terra.

Os produtos de uso e cobertura da Terra que estão atualmente disponíveis, como os desenvolvidos pelos projetos brasileiros MapBiomass e TerraClass, aplicam em suas metodologias o uso de séries temporais que incluem bandas e índices espectrais, bem como técnicas de ML. No entanto, a integração de dados meteorológicos a esses conjuntos de dados poderia auxiliar na detecção e diferenciação dos padrões de uso e cobertura da Terra. Portanto, esta pesquisa levanta a seguinte questão: *a combinação de bandas e índices espectrais de imagens de satélite, dados meteorológicos e técnicas de ML promovem melhor acurácia nos mapas de uso e cobertura da Terra?*

Para responder este questionamento utilizou-se o município de Petrolina-PE localizado na região Semiárida do Brasil onde está inserido o bioma Caatinga, o qual

tem sofrido uma intensa degradação ao longo dos últimos 400 anos em função do uso desordenado e predatório dos recursos naturais (BEZERRA et al., 2014; VITÓRIAS et al., 2015; ALTHOFF et al., 2018). Portanto, a hipótese dessa pesquisa é: a combinação de bandas, índices espectrais e produtos meteorológicos pode melhorar a acurácia dos mapas de uso e cobertura da Terra gerados por classificações de SITS usando aprendizado de máquina.

## 1.1 Objetivos

O objetivo geral desta pesquisa é avaliar o uso de diferentes combinações de bandas espectrais de satélite, índices espectrais e variáveis meteorológicas para melhorar a acurácia dos mapas de uso e cobertura da Terra para o município de Petrolina, Pernambuco, Brasil, usando séries temporais de imagens de satélite e algoritmos de ML e DL.

Os objetivos específicos deste trabalho são:

- a) Avaliar o ganho em utilizar séries temporais de imagens de satélite por meio de duas abordagens: (1) classificação de séries temporais de imagens de satélite e (2) classificação de uma imagem por ano de forma individual.
- b) Propor uma metodologia para representar dados meteorológicos como cubo de dados.
- c) Integrar cubos de dados de observação da Terra e produtos meteorológicos na classificação de séries temporais de imagens de satélite.
- d) Gerar mapas de uso e cobertura da Terra para o município de Petrolina, Pernambuco a partir de diferentes combinações de bandas, índices espectrais e dados meteorológicos.
- e) Avaliar os mapas classificados por meio das melhores práticas de Olofsson et al. (2014).

## 1.2 Contribuições

Nesta dissertação foram realizados vários testes no município de Petrolina no estado de Pernambuco, Brasil. Esse município se destaca pelo desenvolvimento de extensos projetos agrícolas irrigados em sua área, contribuindo para a geração e intensificação de mudanças no uso da terra. Portanto, para responder à questão científica levantada

nesta pesquisa realizaram-se 54 experimentos. Foram utilizadas diferentes combinações de bandas e índices espectrais derivados de imagens de satélites de EO e de produtos meteorológicos entre os períodos 2017/2018, 2018/2019 e 2019/2020. Para cada ano foram conduzidos 18 experimentos com os classificadores RF e TempCNN, sendo 9 combinações para cada modelo, das quais 7 utilizaram a abordagem *early fusion*.

As contribuições decorrentes do desenvolvimento desta pesquisa consistem na publicação de artigo e resumo em congresso nacional e workshop, bem como a implementação de códigos computacionais. A seguir, é apresentada uma lista das principais contribuições:

- a) Artigo publicado no Simpósio Brasileiro de Sensoriamento Remoto (SBSR 2023), com o objetivo de fazer classificações de LULC no município de Petrolina-PE comparando duas abordagens de classificação: séries temporais de NDVI e uma única imagem NDVI por ano. Esse estudo concluiu que, usando o modelo RF e séries temporais de NDVI potencializou a capacidade de diferenciar classes definidas e favoreceu o reconhecimento de mudanças na região de estudo ao longo do tempo (BRITO et al., 2023).
- b) Resumo simples publicado no Workshop de Computação Aplicada (WorCAP 2023) com o objetivo de avaliar o uso de diferentes índices espectrais de produtos meteorológicos para melhorar a acurácia de mapas de LULC usando séries temporais de imagens de satélite e algoritmos de ML e DL. Conclui-se que, os mapas classificados apenas com NDVI apresentaram maior confusão entre as classes Área urbana e Solo exposto. Os mapas usando NDVI, EVI e bandas RGB, identificou melhor regiões de Agricultura tanto com o RF quanto o TempCNN. Os mapas gerados com NDVI, temperatura e precipitação, apresentaram melhor separação entre as classes Solo exposto e Área urbana, porém, maior confusão entre as classes Solo exposto e Vegetação de caatinga (BRITO, 2023).
- c) Códigos implementados para o desenvolvimento dos resultados desta pesquisa. Link: <<https://github.com/Pedrobrito99/dissertation.git>>



## 2 REFERENCIAL TEÓRICO

Neste capítulo são apresentadas as Seções com os conceitos e fundamentações necessários para o entendimento desta pesquisa. Na Seção 2.1 encontram-se as principais características da vegetação caatinga presente no município de estudo. Na Seção 2.2 está o conceito e caracterização da técnica de cubo de dados de observação da Terra. Na Seção 2.3 são apresentados os conceitos de séries temporais de imagens de satélites de observação da Terra, índices espectrais e variáveis meteorológicas, bem como a relação desses índices e variáveis com as mudanças de uso e cobertura da Terra. Os conceitos de aprendizado de máquina e o modelo *Random Forests* (RF), assim como suas aplicações no sensoriamento remoto são mostrados na Seção 2.4. Por fim, na Seção 2.5 são apresentados os conceitos de aprendizado profundo e o modelo *Temporal Convolutional Neural Network* (TempCNN), juntamente com suas aplicações no sensoriamento remoto.

### 2.1 O bioma Caatinga

A Caatinga é o único bioma exclusivo do Brasil, disponibilizando uma área aproximada de  $912.529 \text{ km}^2$ , cobrindo parte dos estados da Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte, Sergipe e Minas Gerais, equivalente a 10,7% de todo o território brasileiro (SILVA et al., 2017; CARVALHO et al., 2023). É composto por um mosaico de florestas sazonalmente secas e arbustos espinhosos, sendo delimitado pelas florestas Amazônica e Atlântica e o bioma Cerrado, a oeste, leste e sul, respectivamente. Como ocorre frequentemente com diversos biomas, a Caatinga é separada dos biomas terrestres vizinhos por áreas de transição. Nessas áreas, encontram-se diferentes tipos de vegetação, cada um com suas próprias necessidades ecológicas, coexistindo lado a lado (EITEN, 1972; LEAL et al., 2005; SILVA; LACHER, 2020).

Muitas espécies vegetais do bioma Caatinga desenvolveram mecanismos de adaptação em consequência da sazonalidade climática. Por exemplo, deciduidade, mudanças morfológicas nas folhas, que envolvem a presença de espinhos, espaçamento foliar e modificações no processo de metabólitos secundários (CARVALHO et al., 2023). O território deste bioma é caracterizado por terrenos cristalinos com baixa permeabilidade e regiões sedimentadas que dispõem de reservas subterrâneas de água. Os solos, nas suas poucas exceções, são ricos em minerais, pedregosos, pouco desenvolvidos e espessos, e tem baixa capacidade de reter água, quesito que limita a produção primária bruta na região (ALVES et al., 2009).

O bioma Caatinga apresenta na sua grande maioria clima semiárido caracterizado pelo baixo volume de precipitação, com médias entre 400 e 1200 mm ao ano, e altas temperaturas, que variam em média entre 25° e 30°C. No entanto, em áreas de planalto, a precipitação anual pode alcançar 1800 mm e as temperaturas médias podem ser inferiores, o que mostra como a Caatinga encontra-se distante de ser homogênea (TABARELLI et al., 2018). Segundo Monteiro et al. (2015), o clima do Bioma Caatinga é anômalo servindo como um importante laboratório de estudo de adaptação das plantas à quantidade de chuva com alta variabilidade e ao estresse ambiental. Por outro lado, Oliveira et al. (2021) afirma que, por causa da sua sensibilidade às variações climáticas e diversidade biótica, é muito complexo monitorar este bioma.

É um desafio entender as mudanças espaço-temporais do uso e cobertura da Terra que ocorrem na região Semiárida. Pois as vegetações naturais conseguem recuperar-se rapidamente, neste sentido necessita-se de estudos que possam trazer resultados mais próximos da realidade (CASTELLETTI et al., 2003; SOUSA et al., 2021). Devido à escassez de estudos em comparação com outros biomas brasileiros, a Caatinga vem sofrendo com a degradação de suas espécies nativas e a exploração excessiva de seus recursos naturais (FREITAS; ARANHA, 2024). Sua vegetação tem sofrido impactos de diferentes ações antrópicas, como a extração de lenhas, geração de pastos e campos agrícolas (ALTHOFF et al., 2018; MACÊDO et al., 2024).

Embora o bioma Caatinga tenha importância fundamental na biodiversidade do planeta Terra com um terço de suas plantas e 15% dos seus animais sendo espécies exclusivas, inexistente em qualquer outra região do mundo, apenas 1% encontra-se sob proteção de conservação (COUTRIM; SOUZA, 2018). Apesar de ser frequentemente associado à seca e à pobreza, o bioma Caatinga possui condições edafoclimáticas essenciais do ponto de vista ambiental, social e econômico - fatores que tornam importantes as ações de monitoramento (MORAIS et al., 2021).

De acordo com Silva et al. (2020a), diversos estudos têm mostrado que a ocorrência de precipitação antes da passagem dos satélites exerce influência no comportamento espectral da cobertura vegetal do bioma Caatinga. Além disso, Silva et al. (2019) identificaram relação direta entre a sazonalidade das chuvas e o crescimento da Caatinga durante o ano. Uma forma de identificar, analisar e inferir mudanças nas coberturas vegetais geradas por processos naturais e artificiais é realizando monitoramento com imagens orbitais (LEITE et al., 2020; SILVA et al., 2020a).

No estado de Pernambuco onde há presença de vegetação caatinga, está inserido o município de Petrolina, com predominância de áreas rurais, onde os habitantes ob-



têm sustento principalmente da agricultura irrigada, de sequeiro e do extrativismo vegetal (SILVA et al., 2020a). Em termos administrativos, o município é composto pela Sede, que se divide em quatro zonas: Norte, Oeste, Leste e Central. Além disso, inclui os distritos de Curral Queimado e Rajada e os povoados de Cristália, Nova Descoberta, Tapera, Izacolândia, Pedrinhas, Uruás, Lagoa dos Carneiros, Caatinginha, Caititu, Cruz de Salinas, Pau Ferro, Atalho, Caiçara, Barreto, Lagoa dos Cavalos, Barreiro, Varzinha, Lagoa da Pedra, Lajedo, Jatobá, Amargosa, Aranzel e Angico Alto (BARBOZA et al., 2018; IBGE, 2024).

No município de Petrolina, encontram-se diversas paisagens, incluindo a presença de vários mosaicos florestais de vegetação arbórea e arbustiva de caatinga. A distribuição dessa vegetação é influenciada por variações climáticas e ambientais, especialmente a intensidade e frequência das chuvas, além das características geológicas e do solo (FILHO et al., 2019; SILVA et al., 2020b; JARDIM et al., 2022). As espécies da Caatinga são majoritariamente caducifólias, adaptadas às condições de escassez de água, e apresentam uma produção de biomassa significativa durante os períodos chuvosos devido ao clima local (SANTOS et al., 2017). A cobertura de dossel das plantas da Caatinga é descontínua, permitindo a exposição do solo em épocas secas, e há também a presença de um estrato herbáceo, cactos e arbustos (BARBOSA et al., 2019).

No município, os rios são predominantemente temporários, com exceção do Rio São Francisco, que possui 52% de sua bacia hidrográfica localizada no Semiárido brasileiro. Esse rio tem sido fundamental para o desenvolvimento de grandes projetos agrícolas irrigados em sua extensão (TEIXEIRA et al., 2008), contribuindo para intensificar e promover a mudança no uso da Terra. Petrolina está localizada na Bacia Hidrográfica do Rio São Francisco e é drenada tanto pela Bacia do Rio Pontal quanto por diversos pequenos rios internos que são afluentes do São Francisco (GEOLÓGICO; BRASIL-CPRM, 2005; BARBOZA et al., 2018). Até 2016, o município contava com cerca de quarenta e cinco açudes de pequeno e médio porte, construídos para armazenar água da chuva e atender às necessidades das famílias (SANTOS; PEREIRA, 2016).

Neste município encontram-se extensas áreas agrícolas irrigadas voltadas para a exploração comercial de diversas frutas. As principais culturas são: manga (43%), uva (31%), coco (9%), goiaba (5%), banana (4%), acerola (4%) e outros (4%), totalizando 220,00  $km^2$ . Essas produções evoluem para patamares cada vez maiores na economia em consequência da exportação, desempenhando um papel crucial nos indicadores

sociais e econômicos da região. O sucesso produtivo dessas culturas é atribuído ao uso da irrigação, às condições edafoclimáticas favoráveis ao longo do ano e à adoção de tecnologias agrícolas avançadas (DINC, 2022).

Os solos deste município se caracterizam por sua profundidade e desenvolvimento. As classes predominantes são os Argissolos Vermelho-Amarelos e Amarelos, que podem ser Eutróficos e apresentar características plínticas ou não, com variações abruptas ou não em sua textura e presença ou ausência de concreções. Além desses, há ocorrências menos frequentes de Latossolos Vermelho-Amarelos Eutróficos, Neossolos Quartzarênicos Distróficos, Neossolos Litólicos Eutróficos, Planossolos Nátricos e Planossolos Solódicos (SILVA et al., 2006). Em Petrolina, também se encontra a extração de gnaiss para ser utilizado na construção civil. Segundo Silva et al. (2023), essa exploração tem causado várias alterações na paisagem local.

Em função desta heterogeneidade de fatores de solo, clima e planta intrínsecos ao bioma, várias pesquisas estão sendo realizadas neste município para buscar resultados mais próximos da realidade. Por exemplo, a pesquisa realizada por Brito et al. (2020), que teve como objetivo analisar a variação espaço-sazonal da cobertura vegetal na região utilizando os índices de vegetação da diferença normalizada (em inglês, *Normalized Difference Vegetation Index*, NDVI) e de vegetação ajustado ao solo (em inglês, *Soil Adjusted Vegetation Index*, SAVI). Os resultados mostraram valores de SAVI variando de 0,2 a 0,3 em áreas de caatinga no início do período seco (30/05/2013) e de 0,1 a 0,2 no período seco (03/09/2013). Nas áreas agrícolas, esse índice evidenciou bem as áreas de agricultura irrigada.

De acordo com o mesmo autor (BRITO et al., 2017), ao comparar duas imagens em áreas de vegetação caatinga - uma no início do período seco (22/05/2016) e outra no período chuvoso (17/01/2017) - concluiu que os valores obtidos por meio do NDVI foram mais elevados no início do período seco, com predominância de valores entre 0,4 e 0,6. Isso ocorreu devido às chuvas nos meses anteriores, quando comparada à imagem do período seco. Por outro lado, nas áreas irrigadas, este índice atingiu valores superiores a 0,6 em ambos os períodos. Neste mesmo trabalho, também foi avaliado o índice de água de diferença normalizada (em inglês, *Normalized Difference Water Index*, NDWI), que apresentou valores negativos em ambos os períodos, com variações interanuais devido à ocorrência de chuvas. O estudo concluiu que, em áreas irrigadas, o índice NDWI evidenciou áreas maiores quando comparado ao NDVI.

## 2.2 Cubo de dados de observação da Terra

Cubos de dados de observação da Terra (em inglês, *Earth Observation*, EO), ou simplesmente cubos de dados, são geralmente definidos como matrizes multidimensionais com dimensões de espaço e tempo (LU et al., 2018; APPEL; PEBESMA, 2019). Por outro lado, de acordo com Appel e Pebesma (2019) e Simoes et al. (2021), um cubo de dados *raster* regular é uma matriz quadridimensional com dimensões  $x$  (longitude ou leste),  $y$  (latitude ou norte), tempo e propriedades presentes nas bandas espectrais que respeitam as seguintes especificidades:

- a) As dimensões espaciais referem-se a um único sistema de referência espacial (em inglês, *spatial reference system*, SRS).
- b) As células de um cubo de dados têm um tamanho espacial constante em referência ao SRS.
- c) A dimensão temporal é constituída por um conjunto de intervalos igualmente espaçados e contínuos.
- d) Uma célula tem um único valor para cada combinação de dimensões.

Em um cubo de dados todas as células dispõem da mesma dimensão espaço-temporal. A resolução espacial das células é a mesma nas dimensões  $x$  e  $y$ , os intervalos temporais são iguais e cada uma das células contém um conjunto válido de medidas. Para cada posição no espaço tem-se um conjunto de séries temporais válidas e, para cada intervalo temporal, o cubo de dados regular deve fornecer uma imagem bidimensional válida (SIMOES et al., 2021).

A abordagem de cubos de dados de EO tem recebido cada vez mais atenção como um novo recurso para armazenar, organizar, gerenciar e analisar dados de EO de uma maneira não disponível antes. Consequentemente, o seu objetivo é alcançar todo o potencial de repositórios de dados de EO abordando os desafios de volume, variedade e velocidade, disponibilizando acesso a dados espaço-temporais extensos em um formulário pronto para análise (BAUMANN, 2017; LEWIS et al., 2017; GIULIANI et al., 2017).

No momento atual estão disponíveis diversas iniciativas de estruturar cubos de dados com diferentes escalas espaciais, como o *Swiss Data Cube* (SDC) (GIULIANI et al., 2017), *Australian Geoscience Data Cube* (AGDC) (LEWIS et al., 2017), *Africa*

*Regional Data Cube* (ARDC) (KILLOUGH, 2019) e *Brazil Data Cube* (BDC) (FERREIRA et al., 2020). De acordo com Ferreira et al. (2020), o projeto BDC dispõe de quatro objetivos predominantes, sendo eles:

- a) Gerar coleções de imagens de dados prontos para análise (em inglês, *Analysis-Ready Data*, ARD) a partir de imagens de sensoriamento remoto com resolução espacial de 10 a 64 metros, para todas as regiões do Brasil.
- b) Estruturar as imagens ARD geradas, como cubos de dados multidimensionais integrando espaço, tempo e propriedades derivadas de espectros.
- c) Utilizar, propor e produzir tecnologias de *big data*, para gerar, armazenar e processar cubos de dados estruturados a partir de imagens ARD.
- d) Desenvolver produtos informativos de LULC para o território brasileiro, por meio de cubos de dados de EO, aplicando procedimentos de processamento de imagens, análise de séries temporais de imagens de satélite (em inglês, *Satellite Image Time Series*, SITS) e técnicas de aprendizado de máquina (em inglês, *Machine Learning*, ML).

Além disso, essas iniciativas disponibilizam diferentes implementações de *softwares* de código aberto. Como o *Open Data Cube* (ODC), para o AGDC, que possibilita a catalogação de conjuntos de dados EO, seu acesso e uso por meio de uma coleção de ferramentas de linhas de comando a uma interface de programação (em inglês, *Application Programming Interface*, API) *Python* (GOMES et al., 2020). Dentro da interface do BDC, o pacote *sits* desenvolvido na linguagem *R* apresenta um conjunto de funções para análise de SITS por meio de métodos de ML e aprendizado profundo (em inglês, *Deep Learning*, DL), usando uma abordagem de tempo-primeiro (em inglês, *Time-First*) e espaço-depois (em inglês, *Space-Later*) (CAMARA et al., 2016; SIMOES et al., 2021).

### **2.3 Dados de séries temporais de imagens de satélites de observação da Terra**

Dados de séries temporais é um termo amplo usado para denotar dados de diferentes áreas associados a uma componente de tempo. Alqahtani et al. (2021) definem dados de séries temporais como uma coleção ordenada de observações ou sequências de pontos de dados gerados ao longo do tempo, em geral, em períodos de tempo regulares. O uso de dados de séries temporais têm sido aplicado por pesquisadores em

diferentes estudos para monitoramento de desmatamento (MATOSAK et al., 2022), identificação de distúrbios florestais (MA et al., 2022), intensidade e produtividade de terras agrícolas (LÖW et al., 2018) e mapeamento de LULC (CHAVES et al., 2021).

Na atualidade, uma maneira de processar esse tipo de dado é utilizando pacotes desenvolvidos para este fim, a exemplo de SITS, que possui medidas calibradas e comparáveis em momentos diferentes para uma mesma região da Terra. Essas medidas podem ser obtidas de um único sensor, como também, da associação de diversos sensores (SIMOES et al., 2021). O curto período de revisitação dos satélites nas regiões possibilita as imagens de alta resolução espacial serem efetivamente organizadas em SITS, que se mostra como ferramenta prática de monitoramento de mudanças na superfície terrestre ao longo do tempo, suportando as diversas aplicações (IENCO et al., 2019).

Por meio das constelações de satélites atuais é possível obter SITS com altas resoluções temporal, espacial e espectral (PELLETIER et al., 2019). Segundo Pelletier et al. (2019), as SITS de alta resolução formam uma grande fonte de dados para mapear a cobertura da terra, em especial, a vegetação e culturas agrícolas (MATTON et al., 2015; VUOLO et al., 2018), tanto em áreas continentais (INGLADA et al., 2017), quanto regionais (IMMITZER et al., 2016). Com o uso de séries temporais, especialistas conseguem adquirir melhor entendimento de padrões e processos ecológicos, observando as mudanças em intervalos de tempo contínuos, ao invés de comparar imagens únicas de diferentes datas (WOODCOCK et al., 2020; SIMOES et al., 2021). Camara et al. (2016), definiram essa técnica como *Time-First, Space-Later*.

O conceito *time-first, space-later* para classificação de imagens de satélite, tem como passo inicial no estudo de dados de sensoriamento remoto, a análise de séries temporais. As informações espaciais são agregadas após as classificações de todas as séries temporais. Assim, cada posição no espaço (latitude e longitude), está relacionada a uma série temporal (SIMOES et al., 2021). De acordo com Lynnes e Huang (2022), a classificação da cobertura da Terra normalmente é beneficiada por essa abordagem, em que a história temporal de cada pixel é utilizada na classificação de LULC. Além disso, diversos trabalhos afirmam que a abordagem *space-first, time-later* é menos apropriada para o monitoramento de mudanças contínuas, indicando o uso do *time-first, space-later* (CAMARA et al., 2016; PELLETIER et al., 2019; WOODCOCK et al., 2020; SIMOES et al., 2021).

Um dos trabalhos atuais na literatura que utiliza a abordagem *time-first, space-later* é dos autores [Chaves et al. \(2021\)](#), para identificar classes de mudanças de uso e cobertura da Terra em paisagens heterogêneas e dinâmicas usando bandas espectrais, VIs e técnicas de reconhecimento de padrões. Os autores também defendem o uso de VIs, afirmando que, embora seja razoável considerar melhores resultados de classificações das mudanças utilizando recurso de todas as bandas espectrais ao longo das séries temporais, ao invés de um índice espectral associado ao verde derivado de uma expressão matemática operando duas bandas por imagem ([PELLETIER et al., 2016](#)). Os VIs são indicadores do vigor da vegetação e sua atividade fotossintética, empregues como séries temporais para analisar as variações em estados fisiológicos e propriedades biofísicas da vegetação ([CORTI et al., 2018](#)).

### 2.3.1 Índices espectrais e variáveis meteorológicas

Em consequência das propriedades físicas e suas interações com a radiação eletromagnética, os materiais da superfície terrestre dispõem de assinaturas espectrais características. Como essas interações alteram de acordo com os efeitos das variáveis ambientais, informações de diferentes processos da superfície conseguem ser obtidas a partir de suas assinaturas espectrais. Apesar disso, frequentemente é necessário realizar a combinação entre regiões espectrais específicas de interesse, isto é, bandas espectrais, para diminuir efeitos de confusão ao analisar um material ou fenômeno específico. Essas combinações de bandas, ou seja, índices espectrais, produzem informações condensadas sobre materiais, subsistemas ou processos específicos ([MONTERO et al., 2023](#)).

Um índice espectral é uma operação matemática aplicada a um conjunto de duas ou mais observações espectrais por pixel para obter uma evidência considerável ([GOVAERTS et al., 1999](#); [PRASAD et al., 2022](#)). Essas operações destacam os pixels de uma imagem para não apenas apresentar a abundância relativa de uma cobertura de terra de interesse, mas também destacar uma função do ecossistema ([DENG et al., 2015](#); [XUE; SU, 2017](#); [TRAN et al., 2022](#)). Várias equações de índices espectrais ideais para diversos campos de aplicação foram produzidas, e constantemente são desenvolvidos e testados novos índices espectrais ([MONTERO et al., 2023](#)).

Atualmente, a biosfera terrestre é o domínio de aplicação mais comum de índices espectrais, em que o foco principal é o monitoramento da vegetação por meio de VIs ([MONTERO et al., 2023](#)). Os VIs mais utilizados são NDVI e índice de vegetação melhorado (em inglês, *Enhanced Vegetation Index*, EVI) que, quando combinados a outros VIs apresentam ganhos no reconhecimento de mudanças sutis na vegeta-

ção, mostrando ser útil para melhorar classificações de LULC (CORTI et al., 2018; CHAVES et al., 2020; CHAVES et al., 2021; MONTERO et al., 2023).

Segundo Devkota et al. (2023), além dos VIs, outros índices espectrais podem ser utilizados para identificar padrões finos de classes de uso do solo específicas, como o índice de diferença normalizada de áreas construídas (em inglês, *Normalized Density Building Index*, NDBI) (ZHA et al., 2003), índice de diferença normalizada da água modificado (em inglês, *Modified Normalized Difference Water Index*, MNDWI) (XU, 2006) e índice de solo exposto (em inglês, *Bare Soil Index*, BSI) (DIEK et al., 2017), para nível de urbanização, características de água e características de solo exposto, respectivamente.

O NDVI é sensível à presença de clorofila, sendo ideal para mapear a saúde e a densidade da vegetação. A clorofila nas plantas absorve de modo eficaz os espectros azul e vermelho enquanto reflete o espectro verde. Além disso, as plantas possuem alta refletância no espectro NIR. Essa combinação de alta absorção no espectro vermelho e alta refletância no NIR torna o NDVI um índice excelente para monitorar a vegetação. Ele é particularmente útil na identificação de classes como Vegetação arbórea/arbustiva e Agricultura (GANJIRAD; BAGHERI, 2024; SAMSUDIN et al., 2016).

O índice de vegetação de diferença normalizada verde (em inglês, *Green Normalized Difference Vegetation Index*, GNDVI) (GITELSON et al., 1996) é semelhante ao NDVI, mas utiliza a banda verde em vez da banda vermelha, oferecendo sensibilidade adicional para algumas culturas específicas. Ele complementa o NDVI, proporcionando uma visão mais detalhada de diferentes tipos de vegetação, sendo particularmente útil na identificação de áreas agrícolas (CÁRDENAS et al., 2018; TUSSUPOV et al., 2024). O EVI, por sua vez, melhora a sensibilidade em regiões de vegetação densa e é menos afetado pela saturação do NDVI em áreas de alta biomassa. Ele também complementa o NDVI, proporcionando uma análise mais precisa, especialmente em áreas com vegetação densa e moderada, como as classes de Vegetação arbórea/arbustiva e Agricultura (HUETE et al., 2002; QIU et al., 2018).

O SAVI reduz a influência do solo na detecção da vegetação, sendo ideal para áreas com cobertura vegetal menos densa. Ele complementa o NDVI e o EVI, oferecendo uma visão aprimorada em áreas de vegetação esparsa, como as classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva. (HUETE, 1988; SANTOS et al., 2014; FADL et al., 2024). O índice de vegetação ajustado ao solo modificado (em

inglês, *Modified Soil Adjusted Vegetation Index*, MSAVI), ajustado para minimizar a influência do solo na vegetação esparsa, é particularmente útil em áreas semiáridas. Ele também complementa o NDVI e o EVI em regiões onde a cobertura vegetal é esparsa e o solo é mais visível, podendo melhorar a classificação das classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva (QI et al., 1994; HUETE et al., 1999; THINLEY et al., 2024).

O índice de solo seco exposto (em inglês, *Dry Bare Soil Index*, DBSI) é eficaz na detecção de solo exposto, quando utilizado em conjunto com o NDVI, pode ajudar a distinguir solo exposto de vegetação, aumentando a precisão na classificação. (RASUL et al., 2018). O NDBI, por sua vez, é destinado à detecção de áreas construídas, sendo fundamental para mapear Área urbana/construída. Quando utilizado com NDVI e MNDWI, o NDBI ajuda a distinguir entre áreas construídas, vegetação e corpos d'água (ZHA et al., 2003; GANJIRAD; BAGHERI, 2024). O MNDWI é utilizado na detecção de corpos d'água e é eficaz em separar a água de áreas urbanas e vegetadas, sendo essencial para identificar a classe Corpos hídricos. Quando integrado a índices como o NDBI, ele facilita a distinção entre áreas urbanas e corpos d'água, o que é crucial para uma classificação precisa em ambientes urbanos (XU, 2006; YANG et al., 2017; YILMAZ et al., 2023).

Pesquisas na literatura tem mostrado que existe relações entre índices espectrais e variáveis meteorológicas, como temperatura da superfície terrestre (ALEXANDER, 2020; GHANBARI et al., 2023), precipitação (SCHULTZ; HALPERT, 1995; DAHAM et al., 2018) e evapotranspiração (MASROOR et al., 2022). Além disso, assim como os índices espectrais, os dados meteorológicos têm apresentado relações com as mudanças de uso do solo (FU; WENG, 2016; LI et al., 2017; MURO et al., 2018; KADAM et al., 2019; BOYAJ et al., 2020; ANJALI; ROSHNI, 2022; HUSSAIN; KARUPPANNAN, 2023).

As variáveis meteorológicas caracterizam as propriedades físicas da atmosfera para compreensão de processos que explicam as suas mudanças, importante para prever estados futuros. Por outro lado, a meteorologia tem objetivo mais amplo, não ocupando apenas a atmosfera, mas também, a sua interação com outros componentes do sistema climático, como oceano, solo e cobertura vegetal. Além disso, é uma ciência quantitativa, construída firmemente a partir de modelos físicos-matemáticos que descrevem o comportamento do fluido atmosférico (MIRANDA, 2013).

Atualmente, o uso de dados de SITS de EO, vem sendo bastante utilizado como entrada para modelos de ML e DL para gerar produtos informativos de mudanças



de uso e cobertura da Terra (PICOLI et al., 2018; PELLETIER et al., 2019; IENCO et al., 2019; SIMOES et al., 2020; CHAVES et al., 2021), e a verificação de outras abordagens que apresentam relações com mudanças de uso da terra, como variáveis meteorológicas, podem melhorar a qualidade desses produtos.

## 2.4 Aprendizado de máquina

O aprendizado de máquina é um subcampo da inteligência artificial (em inglês, *Artificial Intelligence*, AI) que produz algoritmos aptos a tomar decisões fundamentadas em dados, ao contrário de outros modelos que seguem regras de programação estática (BOUCHEFRY; SOUZA, 2020). Segundo Masini et al. (2023), o ML é uma combinação de algoritmos de computador automatizado com métodos estatísticos capazes de aprender padrões intrínsecos em conjuntos de dados. Assim, a área de ML preocupa-se com as técnicas e algoritmos que possibilitam os computadores aprenderem a partir de exemplos (LARY, 2022).

Nos últimos anos, o uso de ML tem sido utilizado em diferentes estudos científicos ambientais, como sequestro de carbono (ZHANG et al., 2019), hidrologia (LANGE; SIPPEL, 2020; MOSAFFA et al., 2022), desastres naturais (POURGHASEMI et al., 2023) e mapeamento de mudanças de uso da terra (SHRESTHA et al., 2023). Por outro lado, de acordo com Lary et al. (2016), por quase duas décadas métodos de ML são aplicados em problemas de ciência e engenharia, porém, em geociência e sensoriamento remoto, sua aplicação é relativamente recente e restrita.

O avanço do ML nos últimos dez anos teve como impulso três tendências tecnológicas (HONG et al., 2020): (1) a grande quantidade de dados disponibilizados para uso acadêmico e industrial, por causa do rápido avanço da detecção e da Internet das Coisas, (2) o melhor acesso a recursos computacionais poderosos e acessíveis, devido elaboração e criação de *chips* orientados para ML, como *Tensor Processing Units* (TPUs) e *Graphic Processing Units* (GPUs) e (3) o desenvolvimento e validação de algoritmos avançados de ML.

Segundo Abraham et al. (2019), as técnicas de ML podem ser divididas em quatro categorias fundamentais de acordo com a abordagem aplicada, o tipo de dado usado na entrada e saída dos modelos e natureza do problema a ser resolvido. Essas categorias são:

- a) Aprendizado supervisionado: que tem como objetivo inferir uma função ou mapeamento baseada em dados de treinamento rotulados. Os dados

de treinamento constituem um vetor de entrada  $\vec{x}$  e de saída  $\vec{y}$ . O vetor  $\vec{y}$  é rotulado de acordo com a experiência disponível anteriormente. Duas categorias de algoritmos com base em aprendizado supervisionado são: classificação e regressão.

- b) Aprendizado não-supervisionado: que é usado para trabalhar com dados sem rótulos. Seu objetivo é identificar um padrão ou estrutura intrínsecas nesses dados. Um exemplo de método de aprendizado não-supervisionado, é agrupamento, utilizado para gerar grupos em um determinado conjunto de dados. Esses grupos são constituídos por objetos que dispõem de alguma característica mutuamente e não possui qualquer similaridade com objetos de outros grupos.
- c) Aprendizado semi-supervisionado: que combina dados rotulados e não rotulados para produzir um modelo adequado para classificar os dados.
- d) Aprendizado por reforço: que utiliza observações coletadas da interação com o ambiente para aplicar uma série de decisões que minimizem o risco ou maximizam a recompensa. Um exemplo de algoritmo de aprendizagem por reforço, é o Q-learning.

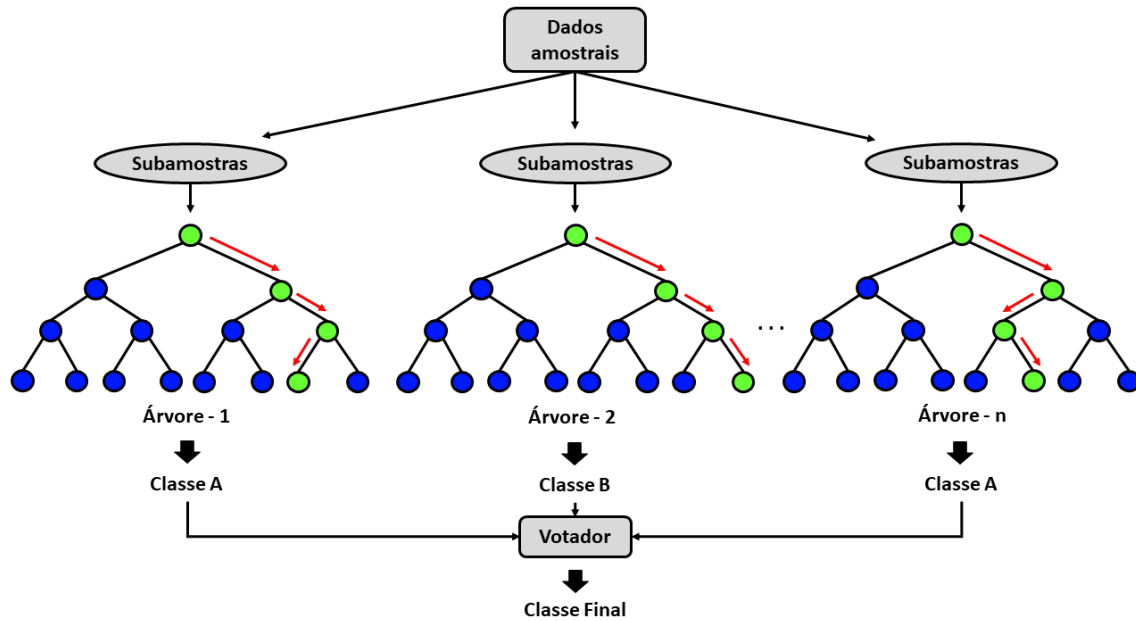
De acordo com Pelletier et al. (2016), métodos supervisionados como *Neural Networks* (NN) (ATKINSON; TATNALL, 1997), *Maximum Likelihood* (ML) (BARTOLUCCI; SCRUCCA, 2010), *Decision Trees* (DT) (FRIEDL; BRODLEY, 1997) e *Support Vector Machines* (SVM) (VAPNIK, 1999) apresentaram melhores resultados do que métodos não-supervisionados em estudos presentes na literatura (SZUSTER et al., 2011; KHATAMI et al., 2016). Além disso, o método supervisionado *Random Forest* (RF) (BREIMAN, 2001) apresentou um desempenho satisfatório para gerar mapas de classificação de uso da Terra precisos (BELGIU; DRĂGUȚ, 2016), semelhantes a classificadores tradicionais como DT e SVM, dispondo de menor tempo computacional (INGLADA et al., 2015; PELLETIER et al., 2016).

#### 2.4.1 *Random Forest* (RF)

O modelo *Random Forest* (Figura 2.1) desenvolvido por Breiman (2001) é um algoritmo de ML supervisionado não paramétrico, que combina os resultados de diversas árvores de decisão para obter um único resultado (MO et al., 2021; ABDULHAFEDH, 2022; CHANG et al., 2022; GUO et al., 2022). As árvores de decisão tem o objetivo de identificar a melhor separação de subconjunto dos dados de entrada e são

usadas tanto para problemas de regressão quanto classificação (ABDULHAFEDH, 2022).

Figura 2.1 - Representação simplificada do modelo *Random Forest* (RF).



Cada uma das árvores são criadas a partir de um algoritmo de aprendizado individual, que separa o conjunto de dados amostrais de entrada em conjuntos menores (subamostras) baseado em um teste de valor de atributo. Os resultados das diversas árvores de decisão são combinados para obter um único resultado.

Fonte: Próprio autor.

O RF é categorizado como uma técnica de *ensemble*, que pode ser definida como a combinação de diferentes modelos, pois combinam vários algoritmos de ML para tomar decisões (UMOH et al., 2022). A combinação de diversos modelos possibilita que o erro tido de um único modelo seja compensado pelos demais produzindo desempenhos melhores, quando comparados a modelos únicos (SAGI; ROKACH, 2018). Além disso, as principais razões de erros no aprendizado são decorrentes do viés, ruído e variância, e as técnicas *ensemble* contribuem para amenizá-los. (ABDULHAFEDH, 2022).

As técnicas *ensemble* podem ser classificadas como *boosting* e *bagging* (também denominada *bootstrap aggregation*) (FREUND et al., 1996; BREIMAN, 1996; FRIEDMAN et al., 2000; FRIEDMAN, 2001), sendo o RF uma das técnicas de *bagging*

(SRIVASTAVA et al., 2023), também definida por alguns autores como uma extensão da mesma (ABDULHAFEDH, 2022).

A *bagging* é um método de produzir um novo conjunto de dados a partir da substituição de um conjunto de dados existentes (UMOH et al., 2022). Nessa técnica, é selecionada uma amostra do conjunto de treinamento de modo aleatório, com substituição. Ou seja, cada elemento (dado) do conjunto de treinamento pode ser escolhido mais do que uma única vez. Após serem geradas diferentes amostras de dados os modelos *bagging* são treinados separadamente. De acordo com o objetivo do modelo, isto é, regressão ou classificação, a média ou predominância das previsões gera uma maior precisão na estimativa (MOHINDRU et al., 2022; ABDULHAFEDH, 2022). Segundo Mohindru et al. (2022) a agregação de *bagging* pode ser utilizada para diminuir a variância utilizando a previsão coletiva do grupo.

Assim, no modelo RF, cada uma das árvores são criadas a partir de um algoritmo de aprendizado individual, que separa o conjunto de variáveis de entrada em conjuntos menores baseado em um teste de valor de atributo, como o coeficiente de Gini (PELLETIER et al., 2016; ABDULHAFEDH, 2022). Diferente das DT tradicionais, as árvores do RF são criadas sem podas e de maneira aleatória selecionando um subconjunto de variáveis de entrada em cada um dos nós (PELLETIER et al., 2016). O número de variáveis usadas na divisão dos nós RF equivale a raiz quadrada do total de variáveis de entrada (LIAW et al., 2002; PELLETIER et al., 2016). Limitando o tamanho do conjunto de variáveis utilizadas para a divisão, o algoritmo tem uma menor complexidade computacional e também diminui a correlação entre as árvores. Assim, a metodologia aplicada pelo RF para a divisão descorrelaciona as diversas árvores, obtendo resultados mais confiáveis e menos variação (PELLETIER et al., 2016).

O procedimento descrito acima é repetido continuamente em cada subconjunto derivado, até que cada um dos nós contenham amostras muito semelhantes ou a divisão não agregue mais valor às previsões. Além disso, a construção das árvores pode ser parada ao chegar em uma profundidade máxima (*max\_depth*) ou a quantidade de amostras do nó seja inferior a um limite mínimo estabelecido (*min\_samples*). Ao longo do processo de decisão, cada uma das árvores realiza a classificação dos dados do recurso de entrada. Por fim, o modelo RF apresenta o rótulo da classe que adquiriu o maior número de votos (PELLETIER et al., 2016).

O algoritmo RF requer o ajuste de quatro parâmetros principais, sendo eles: (i) o número de árvores ( $k$ ), (ii) o número de características selecionadas aleatoriamente

em cada nó ( $n$ ), (iii) a profundidade máxima de cada árvore ( $max\_depth$ ) e (iv) o número mínimo de amostras por nó ( $min\_samples$ ).

## 2.5 Aprendizado profundo

O aprendizado profundo, também conhecido como aprendizado neural profundo e rede neural profunda, é um tipo de aprendizado de representação que utiliza diversas camadas de representações para identificar padrões a partir da análise de dados usados na entrada dos modelos (ESTEVA et al., 2019; CHEN et al., 2023). O mesmo imita a formação e processamento de dados no cérebro humano com a finalidade de tomar decisões (ARORA et al., 2022).

As inúmeras camadas neurais interconectadas processam e extraem recursos de dados brutos, e aprendem gradativamente a reconhecer padrões mais complexos. É possível ajustar as conexões entre os neurônios e as camadas, possibilitando a rede aprender e realizar previsões a partir dos dados brutos de entrada (CHEN et al., 2023). Ainda de acordo com Chen et al. (2023), o ML engloba um conjunto maior de algoritmos utilizados para modelar e analisar dados, enquanto o DL, é um subconjunto especializado do ML, que dispõe do treinamento de redes neurais profundas.

Segundo Arora et al. (2022), o uso das técnicas de DL são geralmente preferidos, pois:

- a) Tem um maior desempenho em comparação a todas as outras abordagens tradicionais quando se tem uma grande quantidade de dados. Contudo, o uso dos métodos tradicionais de ML são preferidos quando se tem um número de dados limitados. Além disso, as técnicas de DL necessitam de uma infraestrutura robusta para funcionar em um intervalo de tempo não excessivo.
- b) Continuam com melhor desempenho em comparação a outros métodos, quando se tem ausência de conhecimento e experiência para introspecção de recursos, uma vez que, a engenharia de recursos é menos preocupante. Mesmo em tarefas consideradas complexas, como reconhecimento de fala, processamento de linguagem natural e classificação de imagens, o DL consegue apresentar êxito.

Nos últimos anos, algoritmos de DL alcançaram um crescimento maciço de popularidade na análise de imagens de sensoriamento remoto (MA et al., 2019), e

tornaram-se o atual modelo para processamento de dados na área (DIGRA et al., 2022). Além disso, Ma et al. (2019) afirmaram que, desde o ano de 2014, a comunidade de sensoriamento remoto direcionou interesse para o DL, e os algoritmos de ML tem apresentado grande sucesso nos diferentes trabalhos de análise de imagem, tais como, detecção de objetos e classificação de cenas de uso e cobertura da terra (CHEN et al., 2014; ZOU et al., 2015; CHENG et al., 2016; YU et al., 2017; VETRIVEL et al., 2018; NIJHAWAN et al., 2019; LI et al., 2020; TANG et al., 2021; CHENG et al., 2022; TEMENOS et al., 2023).

Uma das aplicações dos modelos de DL nos dias atuais é a classificação de séries temporais extraídas de imagens de satélite a partir de cubos de dados de EO para gerar mapas de classificação de uso e cobertura da Terra. Um exemplo é o trabalho realizado por Pelletier et al. (2019), em que, exploraram o uso de *Temporal Convolutional Neural Network* (TempCNN ou 1D-TempCNN) na classificação de SITS, e mostraram que o mesmo teve melhores resultados do que o RF e RNN.

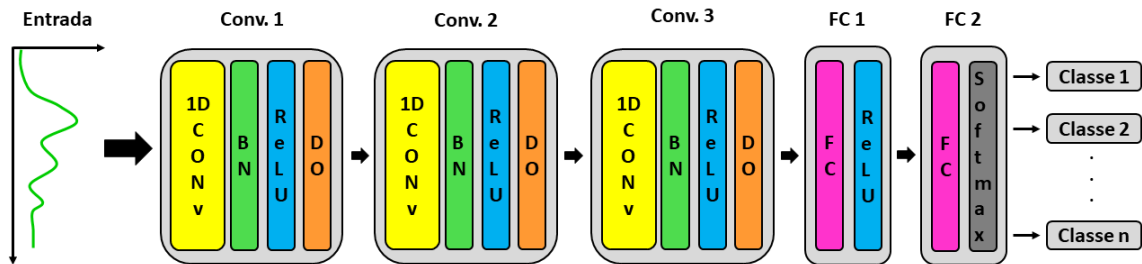
### 2.5.1 *Temporal Convolutional Neural Network* (TempCNN)

A *Temporal Convolutional Neural Network* (TempCNN ou 1D-TempCNN) (Figura 2.2), proposta por Pelletier et al. (2019), é uma rede neural convolucional unidimensional, onde as convoluções são aplicadas no domínio do tempo. A arquitetura de base do modelo TempCNN, recebe uma série temporal multivariada e é constituída por três camadas convolucionais (Conv) aplicadas consecutivamente, duas camadas totalmente conectadas (em inglês, *Fully-Connected*, FC) e uma camada *Softmax* de saída, que gera a distribuição de classe preditiva. Para controlar o *overfitting*, as camadas convolucionais contém normalização em lote (em inglês, *Batch Normalization*, BN), unidade linear retificada (em inglês, *Rectified Linear Unit*, ReLU) e *Dropout* (DO).

As *Convolutional Neural Networks* (CNNs) são tipos de rede neural profunda especial (KISNER et al., 2022), originalmente desenvolvida para análise de imagens, em particular, dígitos manuscritos (ZHU et al., 2018; BELYADI; HAGHIGHAT, 2021). A arquitetura normalmente utilizada nas CNNs é constituída por três camadas, sendo elas, a camada de entrada, a camada oculta e a camada de saída. Os dados de treinamento alimentam a camada de entrada, e em seguida, são transferidos para a camada oculta onde ocorrem as extrações de recursos. A camada oculta é formada por convoluções, *pooling* e camadas totalmente conectadas (HUNG, 2023).

A principal particularidade dos modelos CNNs é a extração de características nas

Figura 2.2 - Representação simplificada do modelo *Temporal Convolutional Neural Network* (TempCNN).



A entrada da rede é uma série temporal multivariada. Três filtros convolucionais são utilizados consecutivamente, seguidos de duas camadas totalmente conectadas (em inglês, *Fully-Connected*, FC) e uma camada *Softmax* de saída, que gera a distribuição de classe preditiva.

Fonte: Próprio autor.

suas camadas convolucionais e *pooling*. As camadas convolucionais conseguem extrair automaticamente recursos dos dados de entrada aplicando convoluções com uma janela deslizante de filtro de tamanho fixo. A camada de *pooling* tem o objetivo de reduzir a dimensionalidade dos mapas de recursos mantendo os recursos considerados relevantes. A camada totalmente conectada é normalmente a última em um modelo CNN, e faz o nivelamento dos mapas de recursos para realizar a classificação (HUNG, 2023).

Segundo Pelletier et al. (2019), os modelos CNNs têm sido utilizado em diversas aplicações de sensoriamento remoto, entre elas a classificação da cobertura da terra a partir de imagens de alta resolução espacial (MAGGIORI et al., 2016; POSTADJIAN et al., 2017), segmentação semântica (VOLPI; TUIA, 2016), detecção de objetos (AUDEBERT et al., 2017), reconstrução de dados ausentes (ZHANG et al., 2018) e *pansharpening* (MASI et al., 2016). Em ambos os trabalhos, as CNNs utilizam a estrutura espacial dos dados ao máximo, aplicando convoluções nas dimensões  $x$  e  $y$ . Contudo, as CNNs podem ser implementadas em diferentes dimensões. (KISNER et al., 2022).

No campo do sensoriamento remoto, a principal aplicação e bem-sucedida das CNNs continua sendo na classificação de imagens hiperespectrais. Testes foram conduzidos com as 1D-CNNs (unidimensionais) em toda a dimensão espectral (HU et al., 2015), 2D-CNNs (bidimensional) em toda a dimensão espacial e 3D-CNNs (tridimensional)

nas dimensões espectral e espacial (LIANG; LI, 2016; LI et al., 2017; HAMIDA et al., 2018). Por outro lado, nas classificações de dados multifonte e multitemporais, às 1D-CNNs e 2D-CNNs estavam sendo aplicadas sem empregar a dimensão temporal, efetuando as convoluções apenas no domínio espectral e espacial, deixando a ordem das imagens ausentes de influência no algoritmo e seus resultados (PELLETIER et al., 2019).

Os modelos 1D-CNNs (TempCNNs), nos quais as convoluções são efetuadas no domínio do tempo, têm provado a sua competência em trabalhar com a dimensão temporal para classificação de séries temporais (PELLETIER et al., 2019). Segundo Kisner et al. (2022), em consequência dos modelos 1D-CNNs conseguirem aprender com os valores dos recursos e suas ordens temporais, os mesmos podem produzir classificações mais precisas. Diante disso, as arquiteturas TempCNN podem aproveitar ao máximo a estrutura temporal das SITS, sendo exploradas em sensoriamento remoto, em que as operações convolucionais são aplicadas apenas no domínio do tempo (MAURO et al., 2017; ZHONG et al., 2019; PELLETIER et al., 2019).



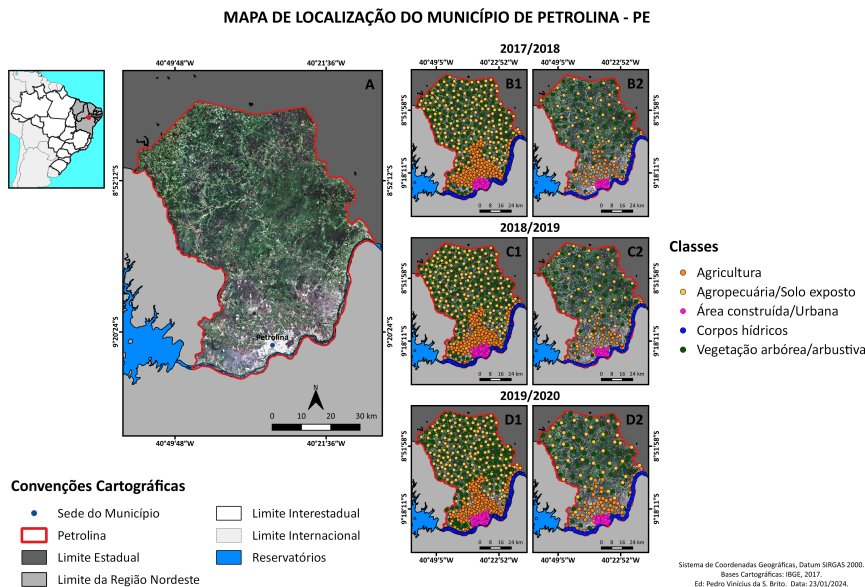
### 3 METODOLOGIA

Neste Capítulo, apresenta-se a metodologia adotada neste trabalho. A localização e caracterização da área de estudo é apresentada na Seção 3.1. Na Seção 3.2 são mostradas as etapas de coleta e processamento dos dados.

#### 3.1 Localização e caracterização da área de estudo

O estudo foi realizado no município de Petrolina (entre 7°37'35" e 9°43'56" de latitude sul e 41°12'11" e 39°7'27" de longitude oeste), região Semiárida Pernambucana do Nordeste brasileiro, às margens do rio São Francisco (Figura 3.1) (SILVA et al., 2019). Petrolina está localizada na área do bioma Caatinga e é reconhecida como o principal polo de fruticultura irrigada do semiárido brasileiro, graças ao fácil acesso à água do rio São Francisco, que abastece os perímetros irrigados da região. O município possui uma área territorial de 4.561,870 km<sup>2</sup>, urbanizada de 86,39 km<sup>2</sup> e uma população estimada em 386.791 habitantes (JARDIM et al., 2022; IBGE, 2022).

Figura 3.1 - Localização da área de estudo e distribuição de amostras de uso e cobertura da Terra obtidas pela plataforma TerraCollect.



(A) Localização geográfica do município de Petrolina, Pernambuco. (B1, C1 e D1) Amostras utilizadas no treinamento dos classificadores RF e TempCNN. (B2, C2, D2) Amostras utilizadas para validar os mapas classificados a partir da técnica de área ponderada.

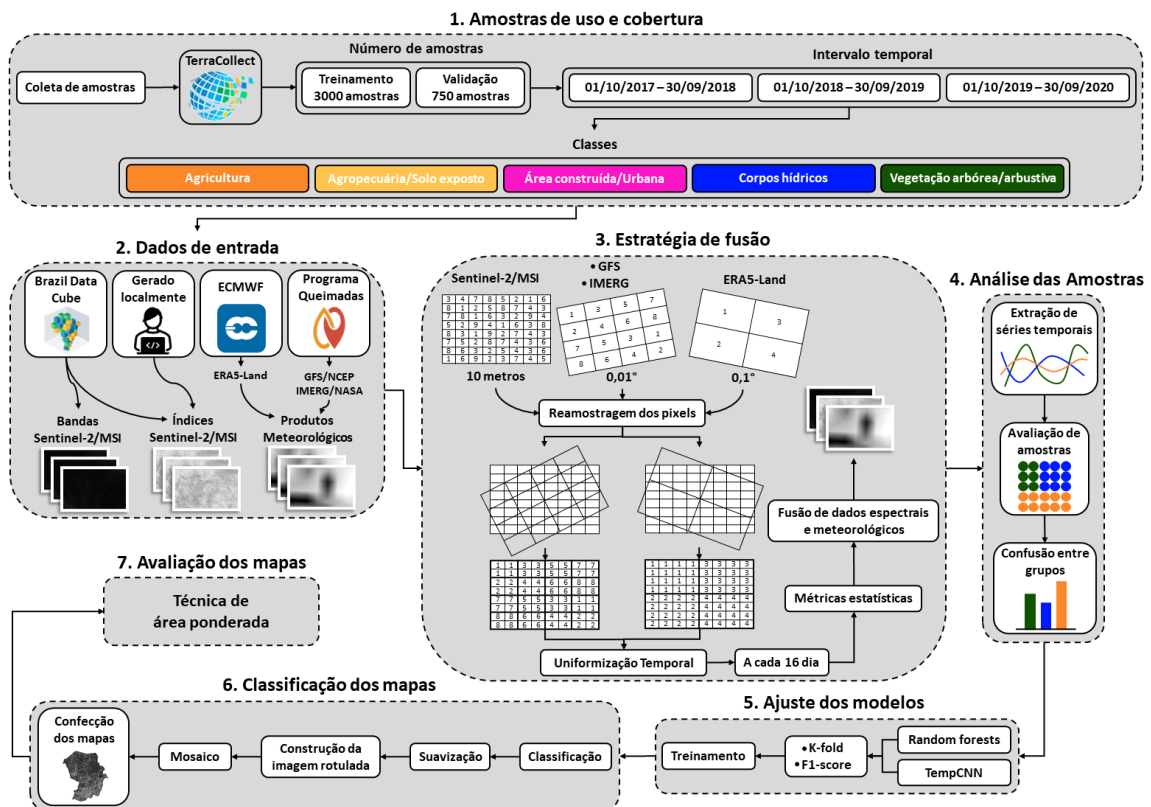
Fonte: Próprio autor.

O clima é classificado como BSw'h'- Semiárido quente e seco, de acordo com a classificação de Köppen-Geiger, com temperatura média do ar de  $26,3^{\circ}C$ , umidade relativa média de 58% e precipitação de aproximadamente 567 mm ao ano (ALVARES et al., 2013; SILVA et al., 2019). Devido à sua localização geográfica e à influência da Zona de Convergência Intertropical (ZCIT), o padrão de precipitação em Petrolina é irregular durante o ano, com maior volume das chuvas ocorrendo de fevereiro a maio (MARENGO et al., 2017; OLIVEIRA et al., 2017; JARDIM et al., 2022).

### 3.2 Coleta e processamento dos dados

As etapas metodológicas deste trabalho envolvem uma série de experimentos de classificação de séries temporais. Em cada experimento foram avaliados diferentes dados de entrada nos modelos de ML e DL. (Figura 3.2).

Figura 3.2 - Fluxograma das etapas metodológicas do trabalho.



Fonte: Próprio autor.

### 3.2.1 Amostras de uso e cobertura

Amostras de LULC foram coletadas via interpretação visual de imagens Sentinel-2/MSI, com auxílio de séries temporais de NDVI usando a plataforma *TerraCollect* desenvolvida no escopo do projeto BDC (FERREIRA et al., 2020). Um total de 3.000 amostras foram coletadas para treinar os classificadores (1.000 por ano/200 por classe), cada uma contendo longitude, latitude, data inicial, final e classe (Figura 3.1 B1, C1 e D1). As classes amostradas foram Agricultura (áreas que fazem uso de irrigação durante todo o ano), Agropecuária/Solo exposto (Agropecuária abrange áreas que transitam entre o uso de agricultura de sequeiro e irrigada, principalmente para pastagem animal), Área construída/Úrbana, Corpos hídricos e Vegetação arbórea/arbustiva. Para validar os mapas classificados, foram realizadas novas coletas, um total de 750 amostras (250 por ano/50 por classe) (Figura 3.1 B2, C2 e D2).

Neste estudo, optou-se por obter o mesmo número de amostras para cada classe do mapa, pois um número igualitário de amostras pode assegurar que todas as classes sejam adequadamente representadas durante o treinamento dos algoritmos de classificação. Isso evita o viés comum em métodos de amostragem proporcional, onde classes minoritárias podem ser sub-representadas, comprometendo a precisão da classificação dessas classes. (KHAN et al., 2017).

### 3.2.2 Dados de satélite

Os dados de sensoriamento remoto incluíram cubos de dados anuais de bandas e índices espectrais gerados pelo satélite Sentinel-2/MSI, coleção BDC S2 SEN2COR\_10\_16D\_STK-1. Esses dados possuem imagens com resolução espacial de 10 metros projetadas e recortadas na grade BDC considerando função de composição temporal de 16 dias, usando a abordagem de empilhamento de imagens com menos nuvem de um período temporal. Os dados foram baixados gratuitamente do servidor *Brazil Data Cube* (BDC) (FERREIRA et al., 2020) abrangendo o período de 01/10/2017 a 30/09/2020 e cobrindo toda a área de estudo. As imagens utilizadas foram de 2017 a 2020 porque eram as disponibilizadas pelo BDC no momento inicial da pesquisa. No total, utilizou-se 8 bandas e 9 índices espectrais contendo 23 imagens por ano para cada um dos atributos (Tabela 3.1).

Os índices EVI e NDVI foram disponibilizados pelo BDC e os demais foram gerados localmente. O índice SAVI foi gerado para as constantes de ajuste  $L = 0,5$  e  $L = 1$ , em conformidade com as características da vegetação de caatinga. De acordo com SANTOS et al. (2014), a constante  $L$  tem a função de minimizar o efeito do solo

Tabela 3.1 - Bandas e índices espectrais utilizados nos experimentos.

Bandas espectrais			
Disponibilidade	Banda	Descrição	
BDC	B02	Blue	
	B03	Green	
	B04	Red	
	B08	Nir	
	B11	Swir16	
	B12	Swir22	
	B8A	Nir08	
	SCL	Cloud	
Índices espectrais			
Disponibilidade	Índice	Fórmula	Referência
BDC	EVI	$2,5 \cdot \frac{nir-red}{nir+6 \cdot red+7,5 \cdot blue+1}$	Huete et al. (1999)
	NDVI	$\frac{nir-red}{nir+red}$	Rouse et al. (1974)
Computado localmente	DBSI	$\frac{swir-green}{swir+green} - NDVI$	Rasul et al. (2018)
	GNDVI	$\frac{nir-green}{NDVI+green}$	Gitelson et al. (1996)
	MNDBI	$\frac{swir22-nir}{swir22+nir}$	Faridatul e Wu (2019)
	MNDWI	$\frac{green-swir}{green+swir}$	Xu (2006)
	MSAVI	$2 \cdot nir+1 - \sqrt{(2 \cdot nir+1)^2 - 8 \cdot (nir-red)}$	Qi et al. (1994)
	NDBI	$\frac{swir16-nir}{swir16+nir}$	Zha et al. (2003)
	SAVI	$\frac{nir-red}{nir+red+L} \cdot (1+L)$	Huete (1988)

BDC = Brazil Data Cube; EVI= Enhanced Vegetation Index; NDVI = Normalized difference vegetation index; DBSI = Dry Bare Soil Index; GNDVI = Green Normalized difference vegetation; MNDBI = Modified Normalized Difference Bare Land Index; MNDWI = Modification of Normalized Difference Water Index; MSAVI = Modified Soil Adjusted Vegetation Index; NDBI = Normalized Density Building Index; SAVI= Soil Adjusted Vegetation Index.

Fonte: Próprio autor.

no sinal da vegetação, em especial, em regiões menos densas. Para  $L = 0$  o SAVI se mantém equivalente ao NDVI,  $L = 0,25$  é indicado para áreas de vegetação de grande densidade,  $L = 0,5$  vegetação intermediária e  $L = 1$  vegetação de baixa densidade. O Sistema de Referência de Coordenadas (SRC) foi ajustado nos dados de 2019 e 2020 para realizar os processos posteriores. A banda SCL foi usada para remover pixels marcados como nuvem e sombra de nuvem.

### 3.2.3 Dados meteorológicos

Os cubos de dados de temperatura do ar e umidade relativa do ar foram gerados a partir do produto resultante do modelo de previsão do tempo *Global Forecast System* (GFS) do *National Centers for Environmental Prediction* (NCEP) (Tabela 3.2). O

GFS gera dados para dezenas de variáveis atmosféricas e do solo, incluindo concentração de ozônio na atmosfera, umidade do solo, vento, precipitação e temperaturas. Os cubos de dados de precipitação foram construídos utilizando o produto gerado do algoritmo *Integrated Multi-Satellite Retrievals for Global Precipitation Measurement*, GPM (IMERG). O mesmo combina informações da constelação de satélites GPM para estimar a precipitação na maior parte da superfície terrestre.

Tabela 3.2 - Variáveis meteorológicas utilizadas nos experimentos.

Disponibilidade	Variáveis	Variáveis meteorológicas			Fonte
		Métricas	Resolução espacial	Resolução temporal	
	Precipitação ( $P$ , $mm\ dia^{-1}$ )	Máxima Média Median Soma			IMERG/NASA
Programa Queimadas	Temperatura do ar ( $T_{ar}$ , $K$ )	Máxima Média Mediana Mínima	$0,01^\circ \times 0,01^\circ$	1 dia	GFS/NCEP
	Umidade relativa do ar ( $UR_{ar}$ , %)	Máxima Média Mediana Mínima			GFS/NCEP
	Evaporação do solo descoberto ( $E_{solo}$ , $m$ de água equivalente)	Máxima Média Mediana Mínima Soma			
ECMWF	Temperatura do solo ( $T_{solo}$ , $k/0 - 7cm$ )	Máxima Média Mediana Mínima			
	Evaporação total ( $E_{total}$ , $m$ de água equivalente)	Máxima Média Mediana Mínima Soma	$0,1^\circ \times 0,1^\circ$	1 dia às 11:00	ERA5-Land/ECMWF
	Umidade do solo ( $U_{solo}$ , $m^3m^{-3}/0 - 7cm$ )	Máxima Média Mediana Mínima Soma			

Fonte: Próprio autor.

Ambos os produtos de temperatura do ar, umidade relativa do ar e precipitação, foram disponibilizados gratuitamente pelo Programa Queimadas, do Instituto Nacional de Pesquisas Espaciais (INPE) (INPE, 2023), e tem resolução espacial de  $0,01^\circ \times 0,01^\circ$  e temporal de 1 dia. Por outro lado, os cubos de dados de evaporação do solo descoberto, temperatura do solo, evaporação total e umidade do solo foram gerados a partir do conjunto de dados de reanálise ERA5-Land (MUÑOZ-SABATER,

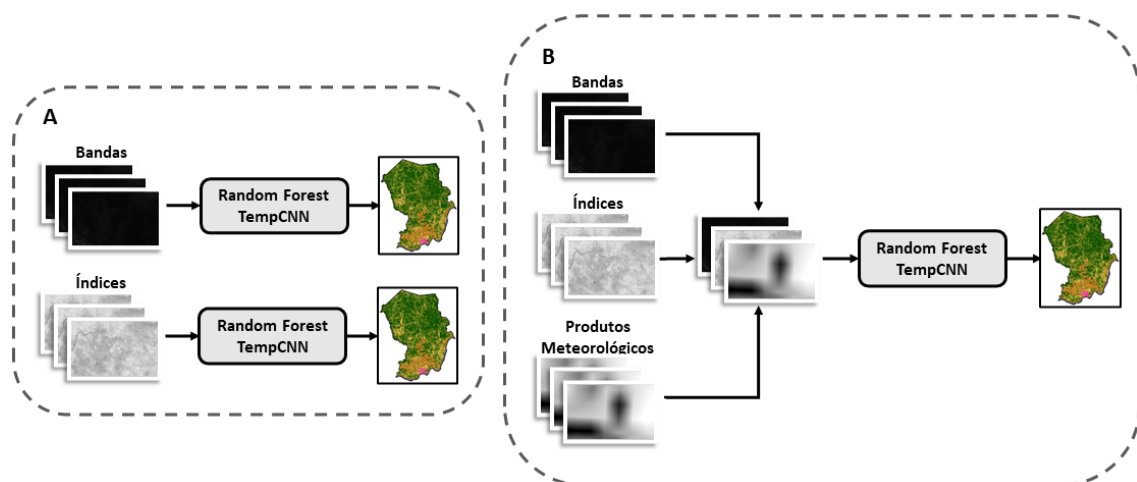
2019), que fornece uma visão consistente da evolução das variáveis terrestres ao longo de várias décadas.

A reanálise combina dados de modelos com observações de todo o mundo em um conjunto de dados globalmente completo e consistente usando as leis da física. O ERA5-Land foi produzido reproduzindo o componente terrestre da reanálise climática ERA5 do *European Centre for Medium-Range Weather Forecasts* (ECMWF). Os seus dados têm resolução espacial de  $0,1^\circ \times 0,1^\circ$  e temporal de 1 dia. Como são disponibilizados a cada uma hora, optou-se por utilizá-los das 11:00, próximo ao horário da passagem do satélite Sentinel-2.

### 3.2.4 Estratégia de fusão

A interação entre as bandas e índices do Sentinel-2 e produtos meteorológicos é analisada empregando a técnica de fusão *early fusion* (Figura 3.3 B), classificada como *input-level* conforme as definições padrão de *deep learning* (OFORI-AMPOFO et al., 2021). Para aplicar a abordagem *early fusion*, as bandas e índices espectrais derivados do Sentinel-2 foram integrados ao longo da dimensão do canal com os dados meteorológicos. Posto que essa concatenação só pode ocorrer quando as imagens do Sentinel-2 e variáveis meteorológicas possuem a mesma dimensão espacial, número e tamanho dos pixels, e número de aquisições, os produtos meteorológicos foram recortados no tamanho da grade *small\_grid* do BDC.

Figura 3.3 - Nível de estratégia de fusão. (A) Nenhuma fusão e (B) *Early fusion*.



Fonte: Próprio autor.

Em seguida, foi feita uma reamostragem com as imagens Sentinel-2 para igualar o número e tamanho dos pixels das imagens com os produtos. Optou-se por reduzir a frequência temporal das variáveis meteorológicas de diária para intervalos de 16 dias. Isso foi feito com o objetivo de uniformizar a dimensão temporal com a do Sentinel-2. A estratégia adotada consistiu em calcular as temperaturas do ar, umidades relativas do ar e temperaturas do solo máxima, média, mediana e mínima; evaporação do solo descoberto, evaporações totais e umidade do solo máxima, média, mediana, mínima e soma; e precipitação máxima, média, mediana e soma, totalizando 23 imagens para cada estatística por ano.

### 3.2.5 Combinações de experimentos

Neste estudo, foram elaborados um total de 54 experimentos, distribuídos igualmente ao longo de três anos distintos: 2017/2018, 2018/2019 e 2019/2020 (Tabela 3.3). Para cada ano, foram realizados 18 experimentos abrangendo os classificadores RF e TempCNN, com 9 combinações para cada modelo e 7 utilizando a técnica *early fusion* (Figura 3.3 B). Inicialmente foram geradas as classificações para cada um dos conjuntos de atributos (Figura 3.3 A), sendo eles, bandas e índices espectrais. Em seguida, gerou-se as classificações usando as bandas, índices espectrais e variáveis meteorológicas aplicando a técnica *early fusion*.

Decidiu-se empregar períodos interanuais, após teste comparativo com classificações anuais. Observou-se que ao considerar todo o período de chuva na região de estudo, a precisão dos mapas classificados foi maior em comparação com mapas anuais.

### 3.2.6 Avaliação de amostras de uso e cobertura

Nos três anos, foram extraídas as séries temporais das amostras a partir dos cubos de dados. As amostras foram filtradas pelo método proposto por Santos et al. (2021), que usa mapas auto-organizáveis (em inglês, *Self-Organizing Maps*, SOM) (KOHONEN, 1990) para agrupar, via inferência *Bayesiana*, séries temporais semelhantes. Nesta etapa, foram gerados dezesseis testes para cada um dos anos variando o tamanho da grade de neurônios para o conjunto amostral. Como base na escolha dos testes, inicialmente adotou-se os parâmetros indicados por Santos et al. (2021), que foram  $5\frac{\sqrt{N}}{2}$  neurônios, em que,  $N$  é o número de amostras, taxa de aprendizado decrescente, de 0,50 a 0,01, e métrica euclidiana.

Para se ter uma maior amplitude entre as confusões das amostras nos agrupamentos SOM, os tamanhos de grades analisadas foram  $4 \times 4$ ,  $5 \times 5$ ,  $6 \times 6$ ,  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$ ,

Tabela 3.3 - Sequências dos experimentos realizados no estudo.

Experimento	Atributo	Período	Modelos
01	Bandas		
02	Índices		
03	Bandas Índices Temperatura do ar		
04	Bandas Índices Temperatura do ar Umidade relativa do ar		
05	Bandas Índices Temperatura do ar Umidade relativa do ar Precipitação		
06	Bandas Índices Temperatura do ar Umidade relativa do ar Precipitação Evaporação do solo descoberto		
07	Bandas Índices Temperatura do ar Umidade relativa do ar Precipitação Evaporação do solo descoberto Temperatura do solo	2017/2018 2018/2019 2019/2020	<i>Random forest</i> TempCNN
08	Bandas Índices Temperatura do ar Umidade relativa do ar Precipitação Evaporação do solo descoberto Temperatura do solo Evaporação total		
09	Bandas Índices Temperatura do ar Umidade relativa do ar Precipitação Evaporação do solo descoberto Temperatura do solo Evaporação total Umidade do solo		

Fonte: Próprio autor.

$10 \times 10$ ,  $11 \times 11$ ,  $12 \times 12$ ,  $13 \times 13$ ,  $14 \times 14$ ,  $15 \times 15$ ,  $16 \times 16$ ,  $17 \times 17$ ,  $18 \times 18$  e  $19 \times 19$  com taxa de aprendizado de 0,50 a 0,01, 200 interações e distância euclidiana. A validação cruzada *k-fold*, com  $k = 5$ , foi aplicada para selecionar os melhores modelos de classificação RF diferenciando-se pelo número de árvores (100, 500, 1000, 1500 e 2000). Por outro lado, para os modelos TempCNN foi aplicado o *model tuning*. O



procedimento de escolha do melhor conjunto de hiperparâmetros para uma aplicação particular. Simoes et al. (2021), afirma ser fortemente recomendada esta etapa para classificação de imagens, pois permite ao algoritmo um ajuste mais preciso aos dados de treinamento. A nível de comparação com o RF foi aplicado o *k-fold*, com  $k = 5$ , para o TempCNN no modelo escolhido no *model tuning*.

Para ajustar os modelos variou-se o número de filtros convolucionais 1D por camada (8, 16, 32, 64, 128, 256), tamanho dos *kernels* convolucionais 1D (1, 3, 5, 7, 9, 11), taxas de *dropout* para filtros convolucionais 1D (0, 15; 0, 1; 0, 2; 0, 3; 0, 4; 0, 5), taxa de aprendizado ( $10^{-4}$  a  $10^{-2}$ ) e redução de peso ( $10^{-8}$  a  $10^{-2}$ ). Realizou-se 50 tentativas aleatórias na busca randômica. A arquitetura das TempCNNs foram treinadas com tamanho do *batch* como 64, 256 nós nas camadas totalmente conectadas e 50 épocas. Um pouco antes dessa quantidade de épocas a perda (em inglês, *loss*) atingiu o patamar. Para validação dos modelos usou-se 20% dos dados de treinamento. Após a seleção dos modelos RF e TempCNN mais precisos por ano, foi aplicada a avaliação *F1-score* para avaliar a precisão das classes de uso e cobertura da Terra.

### 3.2.7 Classificação e avaliação dos mapas

Os modelos RF e TempCNN selecionados foram treinados, as imagens classificadas, e por fim, refinadas por suavização *Bayesiana*. Avaliou-se a precisão dos mapas classificados utilizando a técnica de área ponderada, de acordo com as melhores práticas de Olofsson et al. (2014). Foram usadas suas técnicas estatísticas robustas para estimar a precisão dos mapas classificados, por ter uma base sólida e clara na avaliação. Qualitativamente os mapas foram avaliados e comparados um a um por interpretação visual a conhecimento sobre a região. As etapas de classificação e avaliação foram todas realizadas a partir do pacote SITS.



## 4 RESULTADOS E DISCUSSÃO

Os resultados apresentados a seguir serão divididos em seções, conforme as etapas dos experimentos realizados nesta pesquisa. Na seção 4.1 é apresentada a distribuição da precipitação no município de Petrolina-PE estimada pelo IMERG. Na Seção 4.2 encontra-se os resultados obtidos na avaliação das amostras utilizando o método desenvolvido por Santos et al. (2021). Na Seção 4.3 são apresentados os percentuais de pureza por grupo após a remoção das amostras consideradas ruidosas usando o método de Santos et al. (2021). Na Seção 4.4 está os resultados obtidos por meio da validação cruzada *k-fold* e *F1-score* para os modelos RF e TempCNN. Na Seção 4.5 encontra-se os resultados dos mapas de uso e cobertura da Terra classificados utilizando os modelos RF e TempCNN. Por fim, na Seção 4.6 são apresentados os resultados de confiabilidade dos mapas classificados utilizando as melhores práticas de Olofsson et al. (2014).

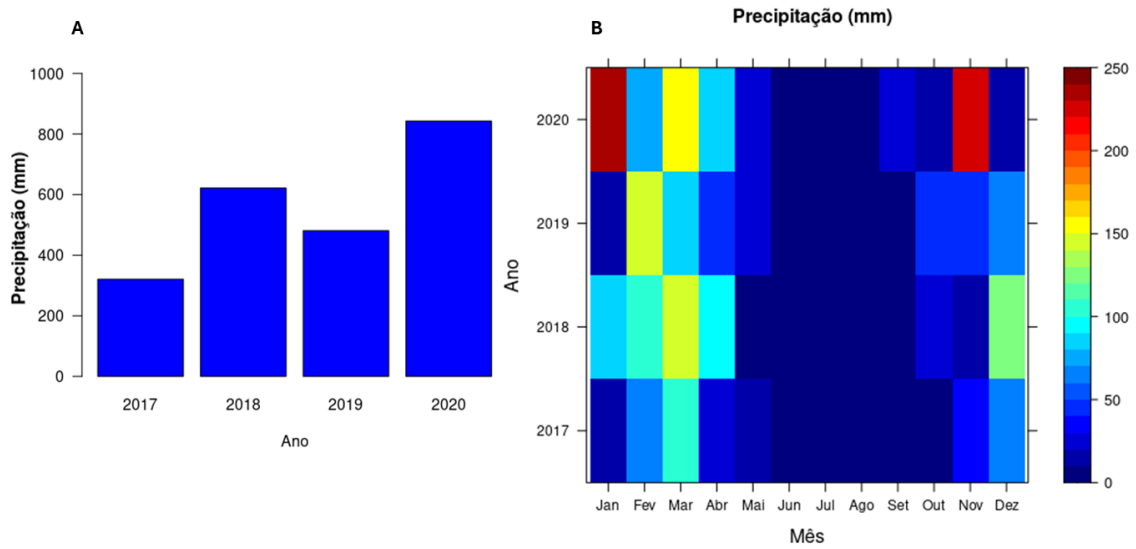
### 4.1 Precipitação estimada pelo IMERG

Para uma melhor compreensão do comportamento e da dinâmica, principalmente das classes de vegetação nos resultados seguintes e na análise dos mapas de classificação, apresenta-se na Figura 4.1 (A e B) a distribuição da precipitação no município de Petrolina. Uma variável de grande importância para entender as mudanças de uso e cobertura da Terra. O ano de 2020 foi o mais chuvoso, com precipitação de 842.96 mm, seguido de 2018 com 621.73 mm, 2019 com 481.01 mm e 2017 com 320.79 mm.

De acordo com Alves et al. (2009), na Caatinga, quando ocorre chuva, a paisagem muda rapidamente em razão da regeneração das plantas, dando novamente um aspecto verde à vegetação. O que antes parecia morto e feio, agora se enche de vida e beleza. Portanto, é importante analisar o comportamento das chuvas em curtos intervalos de tempo na área de estudo, por exemplo, em intervalos mensais.

Os maiores valores de precipitação foram registrados em janeiro, novembro e março de 2020 (com 231.57 mm, 228.81 mm e 150.52 mm, respectivamente), fevereiro de 2019 (com 145.17 mm) e em março e dezembro de 2018 (com 140.48 mm e 125.47 mm, respectivamente) (Figura 4.1 B), que, como mostra a Figura 4.1 (A), foram também os anos de maior volume de chuva. Valores mais altos são esperados nesses meses do ano, pois estão entre os considerados de chuva na região (novembro a abril).

Figura 4.1 - Distribuição da precipitação para o período de 2017 a 2020 estimada pelo IMERG para o município de Petrolina-PE. (A) Anual e (B) Mensal



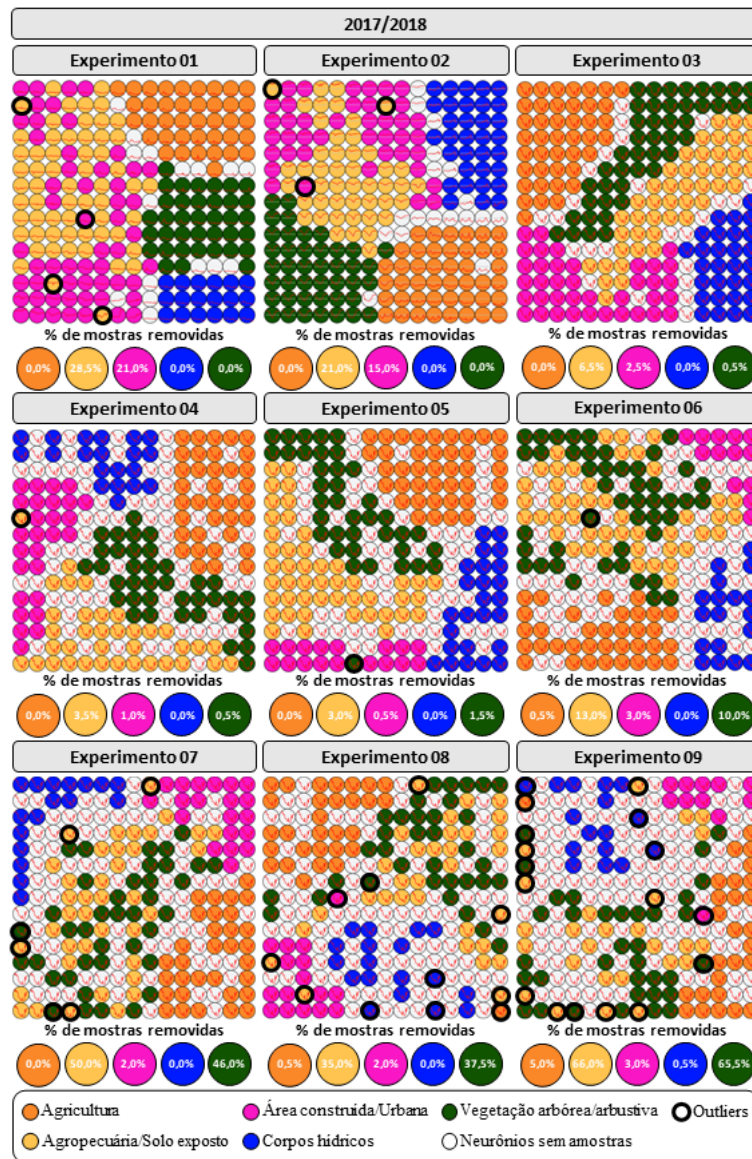
Fonte: Próprio autor.

## 4.2 Avaliação das amostras

Na Figura 4.2 são apresentados os agrupamentos via SOM para os nove tipos de combinações das classificações para o período 2017/2018. Verifica-se que, os agrupamentos utilizando os experimentos 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P) e 06 (bandas, índices, T ar, UR ar, P e E solo) apresentaram menor quantidade de neurônios *outliers*. Segundo Santos et al. (2021), *outliers* são neurônios cuja classe predominante difere das classes da vizinhança. Tendo a ausência dos mesmos no experimento 03.

Comparado aos dois primeiros experimentos que utilizaram apenas bandas e índices espectrais, respectivamente, a adição de variáveis meteorológicas nos experimentos 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P) e 06 (bandas, índices, T ar, UR ar, P e E solo) podem enriquecer o conjunto de dados com informações mais amplas e detalhadas. Isso pode resultar em uma representação mais precisa e abrangente das classes de uso e cobertura da Terra ao longo do tempo. Essa abordagem apresenta resultados mais robustos, especialmente ao lidar com uma quantidade maior de dados (RÊGO et al., 2022).

Figura 4.2 - Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2017/2018.



Fonte: Próprio autor.

Por outro lado, observa-se um aumento no número de neurônios nos experimentos 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), chegando a 17 neurônios *outliers* no experimento 09. Como observado por Dahhani et al. (2022) e Cao et al. (2017), a inclusão de um grande número de variáveis pode resultar em multicolinearidade e redundância de informações, impactando negativamente a capacidade de distinguir as classes de LULC,

podendo ter identificação de mais neurônios *outliers*.

O experimento 01, que utilizou apenas bandas, resultou em mais neurônios *outliers* em comparação com o uso exclusivo de índices (experimento 02) (Figura 4.2). Cada índice espectral destaca uma característica específica da superfície terrestre que está sendo considerada. Além disso, podem fornecer informações sobre outras características da paisagem. Isso ocorre porque os valores dos índices são calculados com base em combinações de diferentes bandas espectrais, e essas combinações podem capturar nuances ou variações em outras características da superfície terrestre (TEFFERA et al., 2018). Essas diversas características capturadas pelos índices espectrais podem fornecer informações mais específicas e distintivas das classes para o modelo, resultando em uma melhor identificação das classes e, conseqüentemente, em um menor número de neurônios *outliers*.

Além disso, os resultados obtidos por Teffera et al. (2018) em sua área de estudo mostraram que, os conjuntos de índices espectrais têm uma habilidade notável em gerar cores visualmente distintas, facilitando a distinção entre diversas características da superfície terrestre, como corpos de água, vegetação e regiões áridas, sem exigir um esforço excessivo. Resultados similares foram alcançados por Piyoosh e Ghosh (2022), que obtiveram uma excelente precisão na classificação de imagens ao utilizar índices espectrais, os quais melhoraram a diferenciação entre várias combinações de classes.

A presença desses neurônios não afirma erros de rotulagem de amostras, mas podem retratar amostras que possuem diferentes padrões de classes de LULC no tempo ou espaço ou inseparáveis utilizando SITS (SANTOS et al., 2021). Houve uma disparidade de 43 neurônios sem amostras entre os experimentos 03 (bandas, índices e Tar) e 04 (bandas, índices, Tar e UR ar), sendo esta a maior diferença registrada entre os experimentos realizados em 2017/2018. No SOM, cada neurônio é rotulado com a classe predominante das amostras associadas a ele. Em determinados casos, nenhuma amostra é associada a um neurônio específico, resultando na atribuição do rótulo “neurônios sem amostras” a esse neurônio vazio (SANTOS et al., 2020). Isso significa que não há exemplos representativos o suficiente de certas classes nos dados para serem atribuídos a esses neurônios específicos.

Com exceção do experimento 05 (bandas, índices, Tar, UR ar e P), à medida que mais variáveis meteorológicas foram adicionadas, aumentaram o número de neurônios sem amostras (Figura 4.2). Nos três últimos experimentos, observa-se uma maior quantidade de neurônios sem amostras, totalizando 78, 94 e 110 em 07 (ban-

das, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), respectivamente. Conforme discutido anteriormente, a inclusão de muitas variáveis pode resultar em multicolinearidade e redundância de informações (DAHANI et al., 2022; CAO et al., 2017). Em razão da alta correlação entre variáveis, torna-se desafiador para o modelo distinguir entre elas e atribuir amostras a neurônios específicos. Além disso, se várias variáveis fornecerem informações semelhantes ou redundantes sobre as classes, isso pode dificultar a diferenciação entre elas, levando à ocorrência de neurônios sem amostras.

Observa-se que os neurônios das classes Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva apresentaram agrupamentos homogêneos nos quatro primeiros experimentos (Figura 4.2). A principal razão para os melhores agrupamentos nessas classes pode ser atribuída aos comportamentos característicos de suas assinaturas espectrais, o que facilita ao modelo distingui-las. Resultados semelhantes foram encontrados por Brito et al. (2023) ao utilizar séries temporais de NDVI. Os padrões espectrais revelaram que as classes Agricultura, Corpos hídricos e Vegetação de caatinga exibem perfis particulares.

Além disso, ao avaliar o potencial dos algoritmos *k-Nearest Neighbor* (kNN), RF, SVM e *Artificial Neural Networks* (ANN) na identificação das fitofisionomias da Caatinga em imagens OLI/Landsat 8, Alba et al. (2022) concluíram que, as classes de Áreas agrícolas e Água não apresentaram confusão ao utilizar o modelo RF. Isso evidencia a facilidade de mapeamento dessas classes por meio da técnica de inteligência artificial e de dados com média resolução espacial. Eles alcançaram um percentual de 96,11% dos pixels classificados como vegetação florestal.

De todos os experimentos referente ao ano 2017/2018, o terceiro (bandas, índices e T ar) foi o que apresentou melhores agrupamentos dos neurônios, tendo grupos homogêneos para todas as classes. A temperatura do ar, em conjunto com as bandas espectrais e índices, provavelmente fornece informações complementares que auxiliam o modelo na identificação de padrões distintos nas classes de uso e cobertura da Terra. Por outro lado, nota-se que houve maiores confusões entre os neurônios das classes Agropecuária/Solo exposto e Área construída/Urbana nos dois primeiros experimentos (bandas e índices, respectivamente).

No estudo conduzido por Brito et al. (2023), constatou-se que, apenas as classes de Área urbana e Solo exposto exibiram perfis semelhantes, ao passo que as categorias de Agricultura, Corpos hídricos e Vegetação de Caatinga demonstraram diferenças

distintas. Além disso, conforme apontado por [Souza e Folharini \(2019\)](#), a dificuldade em distinguir essas classes é uma ocorrência comum na região Semiárida do Nordeste do Brasil devido à proximidade de suas assinaturas espectrais, que às vezes se sobrepõem.

Nos quatro últimos experimentos (Figura 4.2), as maiores confusões ocorreram entre os neurônios das classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva. Embora a principal causa dessas confusões podem estar atribuídas à multicolinearidade e redundância de informações devido ao grande número de variáveis meteorológicas inseridas. De acordo com [Cao et al. \(2017\)](#) e [Dahhani et al. \(2022\)](#), é relevante destacar que essas confusões podem ter sido influenciadas pelas características sazonais e fenológicas da vegetação. Segundo [Silva et al. \(2019a\)](#), durante períodos de seca mais intensa, a distinção entre as classes de solo exposto e vegetação pode ser mais desafiadora, uma vez que a escassez de chuvas dificulta o processo de recuperação da vegetação, que depende diretamente da precipitação.

Verifica-se que as classes Agricultura e Corpos hídricos tiveram os menores percentuais de amostras removidas em quase todos os experimentos, ambos registrando 0,0% nos cinco primeiros e no 07 (bandas, índices, T ar, UR ar, P, E solo e Tsolo) (Figura 4.2). Esses resultados corroboram os encontrados por [Alba et al. \(2022\)](#), os quais afirmaram que, as classes de Áreas Agrícolas e Água em Petrolina podem ser facilmente mapeadas utilizando técnicas de inteligência artificial e dados com média resolução espacial.

Nos dois primeiros experimentos (bandas e índices, respectivamente), não houve remoção de amostras da classe Vegetação arbórea/arbustiva. Isto demonstra que a classe de Vegetação arbórea/arbustiva pode ser bem identificada utilizando dados espectrais. Em relação ao uso de bandas espectrais, de acordo com [Oldeland et al. \(2010\)](#), a região do infravermelho próximo (VNIR) fornece informações específicas sobre pigmentos foliares e estrutura da vegetação, enquanto as bandas na região do infravermelho de onda curta (SWIR) aprimoram a caracterização da vegetação, especialmente em regiões Semiáridas, oferecendo detalhes sobre componentes lenhosos e conteúdo de água da vegetação.

Além disso, [Sims e Gamon \(2002\)](#), sugerem a utilização de todas as bandas espectrais disponíveis, considerando que cada uma delas podem contribuir com uma quantidade específica de informação para explicar a variação dos dados. Essas proposições podem respaldar os resultados satisfatórios na identificação de padrões da classe de Vegetação arbórea/arbustiva no primeiro experimento, uma vez que foram



empregadas diversas bandas das regiões do VNIR e SWIR.

Por outro lado, é notável que as chuvas no Semiárido brasileiro exerçam uma influência significativa nos índices de vegetação, especialmente devido à contribuição da umidade do solo que afeta a rápida capacidade de recuperação da vegetação no bioma Caatinga (SILVA et al., 2022; SILVA et al., 2023). Segundo Silva e Galvêncio (2013) e Silva et al. (2019b), ao avaliarem índices de vegetação como NDVI e SAVI nesta mesma região, constataram alta aplicabilidade e confiabilidade desses índices nas pesquisas sobre a cobertura vegetal da Caatinga. Essas características e resultados ressaltam que os índices de vegetação podem desempenhar um papel crucial na identificação precisa da classe de Vegetação arbórea/arbustiva, com destaque para o NDVI e SAVI.

O percentual de amostras removidas da classe de Vegetação arbórea/arbustiva aumentou progressivamente ao adicionar mais variáveis meteorológicas, variando de 0,5% (experimento 03, com bandas, índices e T ar) a 65,5% (experimento 09, com bandas, índices, T ar, UR ar, P, E solo, Tsolo, E total e U solo). Nota-se uma diminuição nos valores do experimento 07 (bandas, índices, T ar, UR ar, P, E solo e Tsolo) para 08 (bandas, índices, T ar, UR ar, P, E solo, Tsolo e E total), com uma diferença de 8,5%. Porém, no experimento 09, o número de amostras removidas voltou a aumentar.

O aumento gradual da porcentagem de amostras removidas da classe Vegetação arbórea/arbustiva é, principalmente, resultado da multicolinearidade e da redundância de informações gerada pela inclusão de um grande volume de dados (DAHANI et al., 2022; CAO et al., 2017). Ademais, como observado anteriormente, as confusões entre os neurônios das classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva, foram aumentando à medida que mais variáveis foram adicionadas. Isso indica que a remoção das amostras foi majoritariamente motivada pela confusão com a classe Agropecuária/Solo exposto.

Ainda na Figura 4.2, observa-se que nas classes de Agropecuária/Solo exposto e Área construída/Urbana, o número de amostras removidas diminuiu do experimento 01 (bandas) - com percentuais de remoção de 28,5% e 21,0%, respectivamente - para o experimento 05 (bandas, índices, T ar, UR ar e P), em que os percentuais diminuíram para 3,0% e 0,5%, respectivamente. Embora as classes de Agropecuária/Solo exposto e Área construída/Urbana apresentem perfis espectrais semelhantes (BRITO et al., 2023; SOUZA; FOLHARINI, 2019), a inclusão de variáveis meteorológicas como Temperatura do ar, Umidade relativa do ar e Precipitação possibilitou a me-

hora na separabilidade entre essas classes.

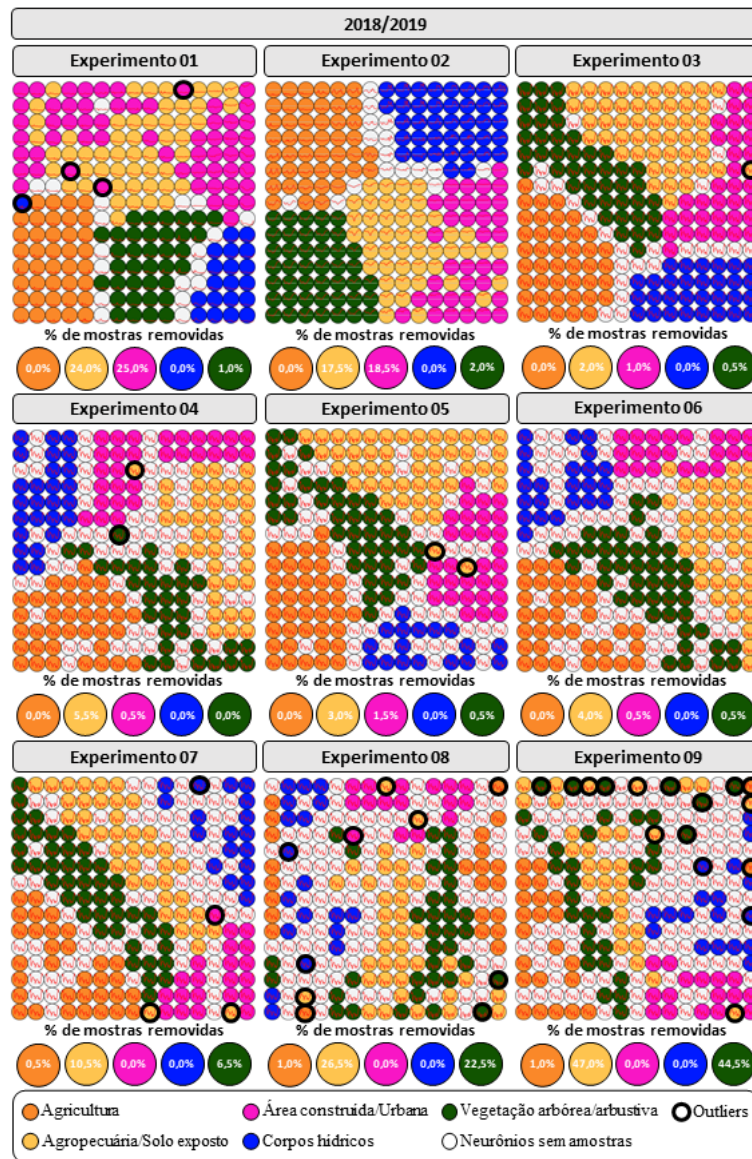
Entre os experimentos 06 (bandas, índices, T ar, UR ar, P e E solo) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), a quantidade de amostras removidas da classe Agropecuária/Solo exposto aumentou de 13,0% para 66,0%, respectivamente. Contudo, nesses mesmos experimentos, o número de amostras removidas da classe de Área construída/Urbana oscilou entre 2,0% e 3,0%. Os resultados demonstram que a inclusão de variáveis meteorológicas contribuíram para a identificação da classe Área construída/Urbana em todos os experimentos realizados. No entanto, a adição de variáveis meteorológicas além das utilizadas no quinto experimento elevou a confusão entre as classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva. Possivelmente influenciada pelas características sazonais e fenológicas da vegetação (DAHANI et al., 2022; CAO et al., 2017).

Analisando esse mesmo conjunto de dados para o período 2018/2019, Figura 4.3 percebe-se que, na maioria dos experimentos o número de neurônios *outliers* foi menor comparado ao período anterior. Havendo uma diferença de 3, 1, 2, 1 e 1 neurônios *outliers* nos experimentos 02 (índices), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), respectivamente. Por outro lado, observa-se um aumento no número desses neurônios em três dos experimentos (03 - bandas, índices e T ar; 04 - bandas, índices, T ar e UR ar; e 05 - bandas, índices, T ar, UR ar e P) de 2018/2019, cada um acrescido de 1 neurônio *outlier*.

Nos experimentos 02 e 06 de 2018/2019, nota-se ausência de neurônios *outliers*, apresentando mais experimentos sem eles do que em 2017/2018 (Figura 4.2). A principal causa da discrepância no número de neurônios *outliers* entre os períodos 2017/2018 e 2018/2019 é na complexa dinâmica ambiental da caatinga. Essa dinâmica é caracterizada pela alta variabilidade espacial e temporal da precipitação, tanto em termos intra-anuais quanto interanuais (BARBOSA et al., 2019). Como mencionado anteriormente, a sensibilidade do bioma às flutuações climáticas e sua rica diversidade biótica dificulta seu monitoramento (OLIVEIRA et al., 2021). Essas mudanças podem resultar em uma diferença na distribuição dos dados, levando a variações no número de *outliers*. Além disso, é importante considerar que algumas amostras podem conter ruídos ou inconsistências, decorrentes de fenômenos atmosféricos. Esses fatores podem aumentar a possibilidade de ocorrência de neurônios *outliers*.

Pode-se observar ainda na Figura 4.3 que, com exceção dos experimentos 01 (bandas)

Figura 4.3 - Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2018/2019.



Fonte: Próprio autor.

e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), o número de neurônios sem amostras diminuiu em 2018/2019 em comparação com 2017/2018 (Figura 4.2). Durante esse período, é possível que tenha ocorrido diversas variações nas condições ambientais na região afetando a distribuição dos dados. Isso pode ter contribuído para a redução do número de neurônios sem amostras, já que 2018/2019 pode ter apresentado condições favoráveis para uma representação mais clara das classes de interesse.

Do mesmo modo que em 2017/2018 (Figura 4.2), em 2018/2019 (Figura 4.3), a maior discrepância no número de neurônios ocorreu entre os experimentos 03 (bandas, índices e T ar) e 04 (bandas, índices, T ar e UR ar), com uma diferença de 39 neurônios. Porém, ao contrário do que ocorreu em 2017/2018, em 2018/2019, houve uma redução no número de neurônios sem amostras do experimento 01 (bandas) para o 02 (índices). Nesse caso, é possível que o modelo SOM tenha sido capaz de identificar e mapear de maneira mais precisa características e variações nos dados dos índices espectrais desse período específico. Além disso, em 2018/2019 podem ter ocorrido condições climáticas mais favoráveis para os índices espectrais identificarem características específicas das classes de interesse.

Essas possíveis respostas se devem ao fato de que os índices espectrais são desenvolvidos para capturar características específicas da superfície da Terra, como água, solo e vegetação. Essas características podem ser menos complexas de serem diferenciadas pelo modelo SOM, resultando na redução do número de neurônios sem amostras. Porém, conforme mencionado por Santos et al. (2021), a identificação de neurônios não necessariamente indica erros de rotulagem das amostras.

Em relação às classes, a Figura 4.3, referente ao período 2018/2019, mostra que os neurônios associados às classes de Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva tiveram agrupamentos homogêneos nos experimentos 02 (índices), 03 (bandas, índices e T ar), 05 (bandas, índices, T ar, UR ar e P) e 06 (bandas, índices, T ar, UR ar, P e E solo). Apesar de um neurônio *outlier* da classe Agropecuária/Solo Exposto estar presente no experimento 03, este experimento pode ser considerado como aquele com os melhores agrupamentos de neurônios, tanto em 2017/2018 (Figura 4.2) quanto em 2018/2019. Isso se deve ao fato de também possuir uma quantidade reduzida de neurônios sem amostras (apenas 8 neurônios), em comparação com os experimentos subsequentes no mesmo período.

Observando 2018/2019 (Figura 4.3), nota-se confusões entre as classes Agropecuária/Solo Exposto e Área Construída/Urbana nos dois primeiros experimentos (bandas e índices, respectivamente) permanecendo semelhante ao que aconteceu em 2017/2018 (Figura 4.2). Além da semelhança espectral entre essas classes, como documentado por Souza e Folharini (2019), Ali e Johnson (2022) e Brito et al. (2023) em estudos realizado em regiões semiáridas, sendo importante considerar que a resolução espacial dos dados do Sentinel-2 pode não ser suficiente para distinguir essas classes com precisão, especialmente em áreas próximas ou com sobreposição. Assim como em 2017/2018, as principais confusões nos quatro últimos experimentos ocor-

reram entre as classes Agropecuária/Solo Exposto e Vegetação Arbórea/arbustiva.

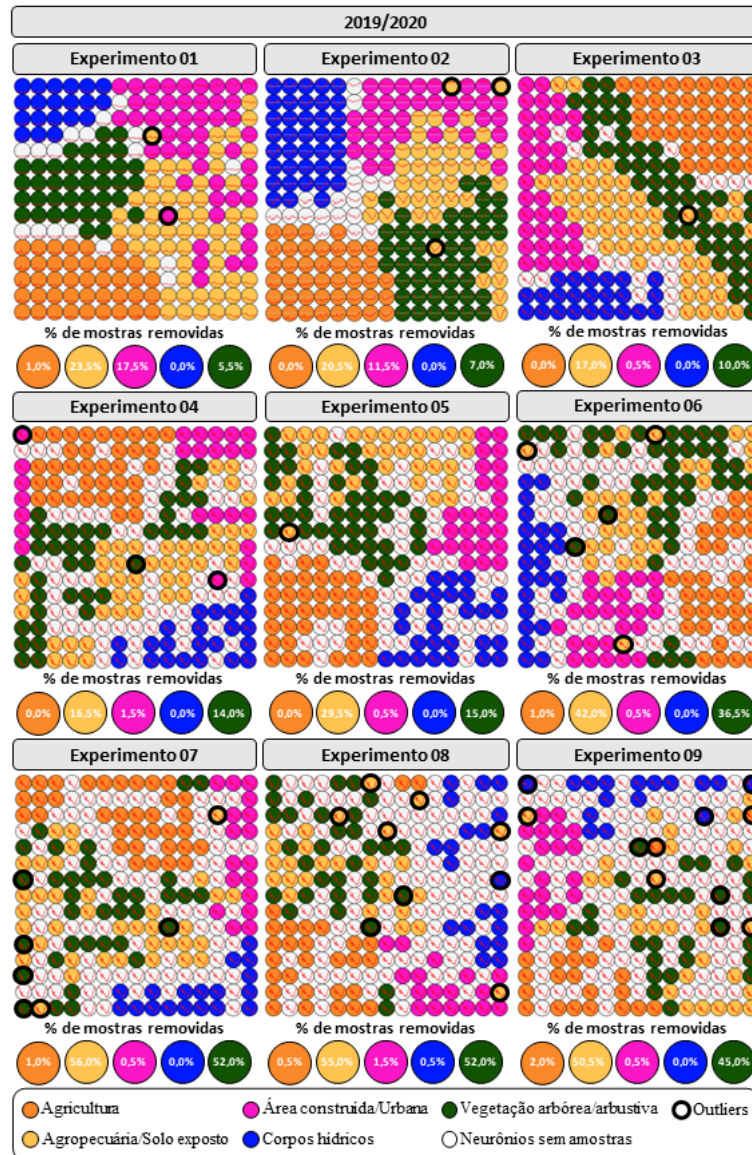
O número de amostras em cada experimento referente à Figura 4.3 mostra que as classes Agricultura e Corpos hídricos se manteve com menor quantidade de amostras removidas na maioria dos experimentos, semelhante a 2017/2018 (Figura 4.2). Em 2018/2019, essas classes registraram 0,0% de amostras removidas nos seis primeiros experimentos. Diferentemente da classe Corpos hídricos que apresentou esse mesmo resultado em todos os experimentos. A classe Vegetação arbórea/arbustiva também registrou 0,0% de amostras removidas no experimento 04 (bandas, índices, T ar, UR ar), porém, com predominância de 0,5% nos experimentos 03 (bandas, índices e T ar), 05 (bandas, índices, T ar, UR ar e P) e 06 (bandas, índices, T ar, UR ar, P e E solo). Além disso, observa-se um aumento na quantidade de amostras removidas a partir do experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), atingindo 6,5%, e chegando a 44,5% no experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo).

Ainda na Figura 4.3 observa-se que, apesar da diminuição no número de amostras removidas das classes Agropecuária/Solo Exposto e Área Construída/Urbana entre os experimentos 01 (bandas) - com percentuais de remoção de 24,0% e 25,0%, respectivamente - e o experimento 03 (bandas, índices e T ar) - com percentuais de 2,0% e 1,0%, respectivamente - houve um aumento novamente no experimento 04 (bandas, índices, T ar e UR ar) para a classe Agropecuária/Solo exposto, atingindo 5,5%. O maior percentual foi registrado no experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), com 47,0%. No entanto, a classe Área Construída/Urbana manteve seus valores entre 0,0% nos experimentos 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09, e 1,5% no experimento 05 (bandas, índices, T ar, UR ar e P).

Avaliando o mesmo conjunto de experimentos para o período de 2019/2020 (Figura 4.4), pode se observar que, nos experimentos 01 (bandas), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) realizados nesse período, houveram menor quantidade de neurônios *outliers* comparado com anos anteriores. Isso pode estar associado novamente a sazonalidade climática da região, de acordo com os resultados de [Becerra et al. \(2015\)](#). Essa diferença é evidenciada ao observar que, em relação a 2017/2018 (Figura 4.2), houve uma redução de 2, 2 e 6 neurônios *outliers*, respectivamente, e em relação a 2018/2019 (Figura 4.3), a redução foi de 2, 1 e 5 neurônios, respectivamente. No entanto, nos experimentos 04 (bandas, índices, T ar e UR ar), 06 (bandas, índices,

T ar, UR ar, P e E solo) e 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), o número desses neurônios aumentaram em comparação com 2017/2018, com uma diferença de 2, 4 e 1 respectivamente, e também em relação a 2018/2019, com uma diferença de 1, 5 e 3 neurônios, respectivamente.

Figura 4.4 - Agrupamentos das séries temporais dos experimentos na grade 15x15 para 2019/2020.



Fonte: Próprio autor.

Diferentemente de 2017/2018 (Figura 4.2) e 2018/2019 (Figura 4.3), esse período de 2019/2020 não apresentou experimentos sem a presença de neurônios *outliers*

(Figura 4.4). Conforme discutido anteriormente, a variação no número de *outliers* pode ser atribuída a mudanças interanuais, como por exemplo o volume de chuvas, bem como à presença de ruídos e inconsistências em algumas amostras. Além disso, é possível que amostras coletadas em 2019/2020 tenham apresentado maior variabilidade espacial e temporal ou heterogeneidade em comparação aos anos anteriores (2017/2018 e 2018/2019). Essas características podem ter contribuído para o aumento no número de neurônios *outliers*, uma vez que dados muito complexos ou heterogêneos podem dificultar a identificação e compreensão de padrões evidentes e representativos pelo algoritmo SOM.

Na Figura 4.4, observa-se que o número de neurônios sem amostras em 2019/2020 foi menor nos experimentos 03 (bandas, índices e T ar) e 04 (bandas, índices, T ar e UR ar) comparado a 2017/2018 (Figura 4.2) (com uma diferença de 5 e 8 neurônios, respectivamente) e 2018/2019 (Figura 4.3) (com uma diferença de 4 e 3 neurônios, respectivamente). No entanto, nos experimentos 05 (bandas, índices, T ar, UR ar e P), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) de 2019/2020, observou-se um aumento no número de neurônios sem amostras em comparação com os mesmos experimentos realizados em 2017/2018 (com uma diferença de 1, 4 e 7 neurônios, respectivamente) e 2018/2019 (com uma diferença de 10, 13 e 7 neurônios, respectivamente).

Semelhante ao que aconteceu nos períodos de 2017/2018 (Figura 4.2) e 2018/2019 (Figura 4.3), em 2019/2020 (Figura 4.4), a maior diferença no número de neurônios sem amostras foi observada entre os experimentos 02 (índices) e 03 (bandas, índices e T ar), totalizando 40 neurônios. Da mesma forma que em 2017/2018, houve um aumento no número desses neurônios do experimento 01 (bandas) para o 02 em 2019/2020, com ambos os períodos registrando 18 neurônios sem amostras no segundo experimento. Como mencionado anteriormente, essa variação no número de neurônios sem amostras entre os anos pode estar diretamente relacionada às variações nas condições ambientais da região.

Em 2019/2020 (Figura 4.4) os neurônios das classes Agricultura e Corpos hídricos formaram agrupamentos homogêneos na maioria dos experimentos, com exceção dos dois últimos. Por outro lado, a classe Vegetação arbórea/arbustiva, que apresentou homogeneidade em quatro dos experimentos nos períodos de 2017/2018 (Figura 4.2) e 2018/2019 (Figura 4.3), mostrou-se homogênea apenas no primeiro experimento em 2019/2020. Observa-se que em 2019/2020, a classe Vegetação arbórea/arbustiva demonstrou confusão com a classe Agropecuária/Solo Exposto a partir do segundo

experimento, e essa confusão aumentou à medida que mais variáveis meteorológicas foram adicionadas.

Como mencionado anteriormente, as confusões entre as classes de Vegetação arbórea/arbustiva e Agropecuária/Solo Exposto podem ser influenciadas pelas características sazonais e fenológicas da vegetação (SILVA et al., 2019a). Essas confusões, mais uma vez podem ser intensificadas em função da multicolinearidade e redundância de informações decorrentes da inclusão de muitas variáveis meteorológicas (CAO et al., 2017; DAHHANI et al., 2022). Ao analisar os experimentos 01 (bandas) e 02 (índices) dos períodos anteriores, ficam evidentes que as confusões entre os neurônios das classes Agropecuária/Solo Exposto e Área Construída/Urbana foram menos frequentes em 2019/2020. As condições climáticas de 2019/2020 podem ter provocado impacto positivo nos padrões espectrais, gerando comportamentos espectrais menos semelhantes entre as classes Agropecuária/Solo Exposto e Área Construída/Urbana.

Comparando os percentuais de amostras removidas em 2019/2020 com os anos anteriores (Figura 4.4), observa-se que, assim como em 2017/2018 (Figura 4.2) e 2018/2019 (Figura 4.3), as classes Agricultura e Corpos hídricos tiveram o menor número de amostras removidas na maioria dos experimentos. Ambas as classes registraram, em 2019/2020, 0,0% de amostras removidas nos experimentos 02 (índices), 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar) e 05 (bandas, índices, T ar, UR ar e P). Além disso, a classe Corpos hídricos também obteve o mesmo percentual nos experimentos 01 (bandas), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), enquanto no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) registrou 0,5% de amostras removidas. Neste mesmo ano no experimento 09 a classe Agricultura teve a sua maior quantidade de amostras removidas, com 2,0%, resultado similar ao observado em 2017/2018, com 5,0%.

Por outro lado, a classe Vegetação arbórea/arbustiva teve uma quantidade maior de amostras removidas nos oito primeiros experimentos, comparado com 2017/2018 (Figura 4.2) e 2018/2019 (Figura 4.3). No entanto, no experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), 2017/2018 apresentou uma quantidade maior, com 65,5% das amostras removidas em comparação com 2019/2020 (Figura 4.4), que removeu apenas 45,5%. Semelhante ao que ocorreu em 2018/2019, houve uma redução na quantidade de amostras removidas em 2019/2020 entre os experimentos 01 (bandas) - com percentuais de remoção de 24,0% e 23,5%,



nos respectivos anos - e o experimento 03 (bandas, índices e T ar) com percentuais de 1,0% e 0,5%, nos respectivos anos - para ambas as classes Agropecuária/Solo Exposto e Área Construída/Urbana. No entanto, ao contrário de 2018/2019, em 2019/2020 a classe Área Construída/Urbana teve aumento já no experimento 04 (bandas, índices, T ar e UR ar) - com percentuais de 1,5%.

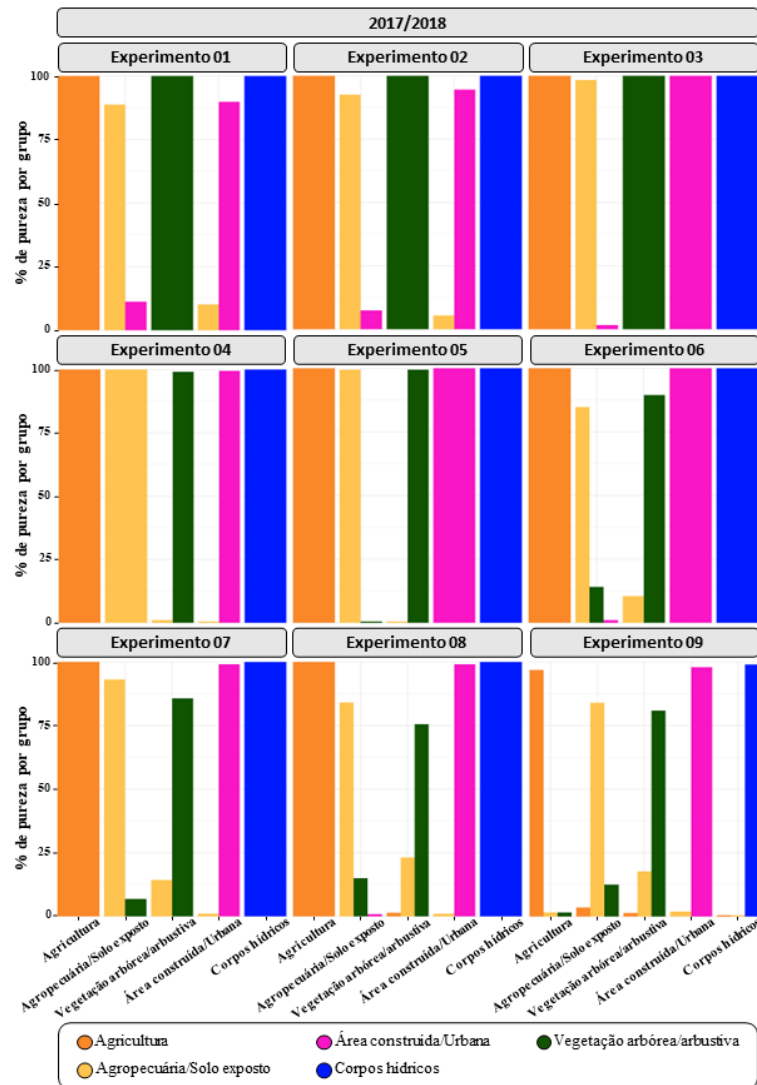
Ainda em 2019/2020 (Figura 4.4), nota-se que, a partir do experimento 05 (bandas, índices, T ar, UR ar e P) - com um percentual de remoção de 29,5% - o número de amostras removidas da classe Agropecuária/Solo Exposto aumentou gradualmente até o experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), com uma remoção de 56,0%, onde teve o maior número de amostras removidas. Nos experimentos subsequentes, esses valores diminuiriam novamente, alcançando 50,5% no experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo). Quanto à classe Área Construída/Urbana, as taxas de remoção de amostras variaram de 0,5% no experimento 05 (bandas, índices, T ar, UR ar e P) a 1,5% no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), a partir do quinto experimento.

### 4.3 Confusão entre grupos

Na Figura 4.5, são exibidos os resultados de outra etapa, mostrando o percentual de pureza por grupo após a remoção das amostras consideradas ruidosas por meio da aplicação SOM, etapa anterior (Figura 4.2). Observa-se que ambos os experimentos identificaram agrupamentos com grau de pureza igual ou superior a 75,52% (experimento 08, utilizando bandas, índices, T ar, UR ar, P, E solo, T solo e E total, na classe Vegetação arbórea/arbustiva). Nos experimentos 01 (bandas) e 02 (índices), os grupos com menor percentual de pureza foram das classes Agropecuária/Solo Exposto (com 88,81% e 92,54%, respectivamente) e Área Construída/Urbana (com 89,87% e 94,61%, respectivamente). No entanto, nos experimentos 03 (bandas, índices e T ar), 05 (bandas, índices, T ar, UR ar e P) e 06 (bandas, índices, T ar, UR ar, P e E solo), a classe Área Construída/Urbana apresentou 100% de pureza, enquanto Agropecuária/Solo exposto atingiu esse mesmo valor apenas no experimento 04 (bandas, índices, T ar e UR ar).

Ambas as classes Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva alcançaram 100% de pureza nos experimentos 01 (bandas), 02 (índices) e 03 (bandas, índices e T ar) (Figura 4.5). No entanto, esse percentual se manteve apenas para as classes Agricultura e Corpos Hídricos nos experimentos 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e

Figura 4.5 - Porcentagem de confusão entre os grupos do agrupamento SOM com grade  $15 \times 15$  para 2017/2018.



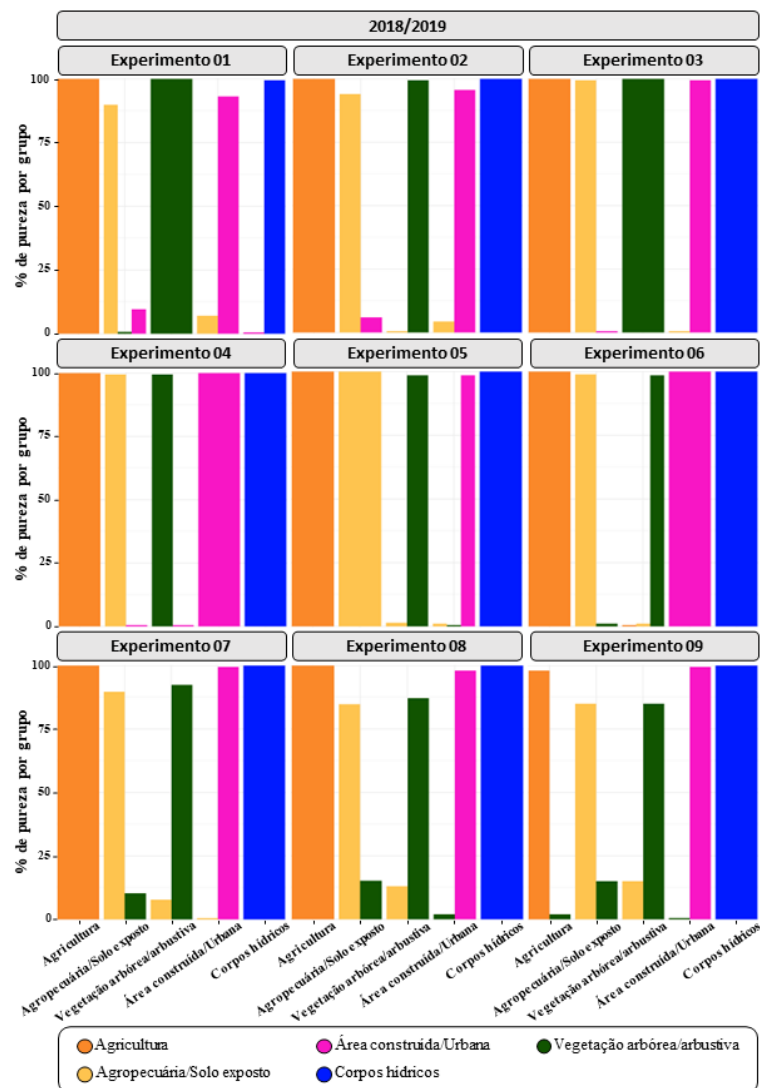
Fonte: Próprio autor.

E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total). A partir do experimento 05, nota-se que os grupos de Agropecuária/Solo Exposto quanto Vegetação arbórea/arbustiva apresentaram os menores percentuais de pureza em todos os experimentos subsequentes, variando de 83,92% (experimento 09, utilizando bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) a 99,48% (experimento 05) e de 75,52% (experimento 08) a 99,49% (experimento 05), respectivamente.

Os resultados do percentual de pureza por grupo para 2018/2019 (Figura 4.6), mos-

tram que, semelhante a 2017/2018 (Figura 4.5), o grupo com menor percentual de pureza em 2018/2019 foi observado no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), com 84,76%. Por outro lado, diferentemente do que ocorreu anteriormente em 2017/2018, esse baixo percentual foi encontrado na classe Agropecuária/Solo Exposto, não na classe Vegetação arbórea/arbustiva. Vale destacar que como discutido anteriormente, as maiores confusões tendem a ocorrer nessas classes.

Figura 4.6 - Porcentagem de confusão entre os grupos do agrupamento SOM com grade  $15 \times 15$  para 2018/2019.



Fonte: Próprio autor.

Além dos experimentos 01 (bandas) e 02 (índices) como observado em 2017/2018 (Figura 4.5), os menores valores percentuais de pureza do experimento 03 (bandas, índices e T ar) também foram registrados nos grupos das classes Agropecuária/Solo Exposto e Área Construída/Urbana, em 2018/2019 (Figura 4.6). Os percentuais de pureza para estas classes foram 89,87% e 93,05% no experimento 01, 94,01% e 95,62% no experimento 02, e 99,48% e 99,49% no experimento 03, respectivamente. No entanto, é possível observar que em 2018/2019, no experimento 01, o grupo da classe Agropecuária/Solo exposto apresentou confusão com o grupo da classe Vegetação arbórea/arbustiva (de 0,63%). Além disso, no segundo experimento de 2018/2019, o grupo da classe Vegetação arbórea/arbustiva mostra confusão com o de Agropecuária/Solo exposto (de 0,50%), que ocorreu em 2017/2018 a partir do experimento 04 (bandas, índices, T ar e UR ar).

O grupo da classe Área Construída/Urbana em 2018/2019 (Figura 4.6) alcançou percentual máximo de pureza em menos experimentos em comparação com 2017/2018 (Figura 4.5), ocorrendo apenas nos experimentos 04 (bandas, índices, T ar e UR ar) e 06 (bandas, índices, T ar, UR ar, P e E solo). Por outro lado, o grupo da classe Agropecuária/Solo Exposto manteve a mesma quantidade de experimentos com percentual máximo de pureza, ocorrendo apenas no experimento 05 (bandas, índices, T ar, UR ar e P). No experimento 03 (bandas, índices e T ar) realizado em 2018/2019 (Figura 4.6), os grupos das classes Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva obtiveram 100% de pureza. Entretanto, esse mesmo resultado foi observado exclusivamente nas classes Agricultura e Corpos hídricos nos experimentos 02 (índices), 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total).

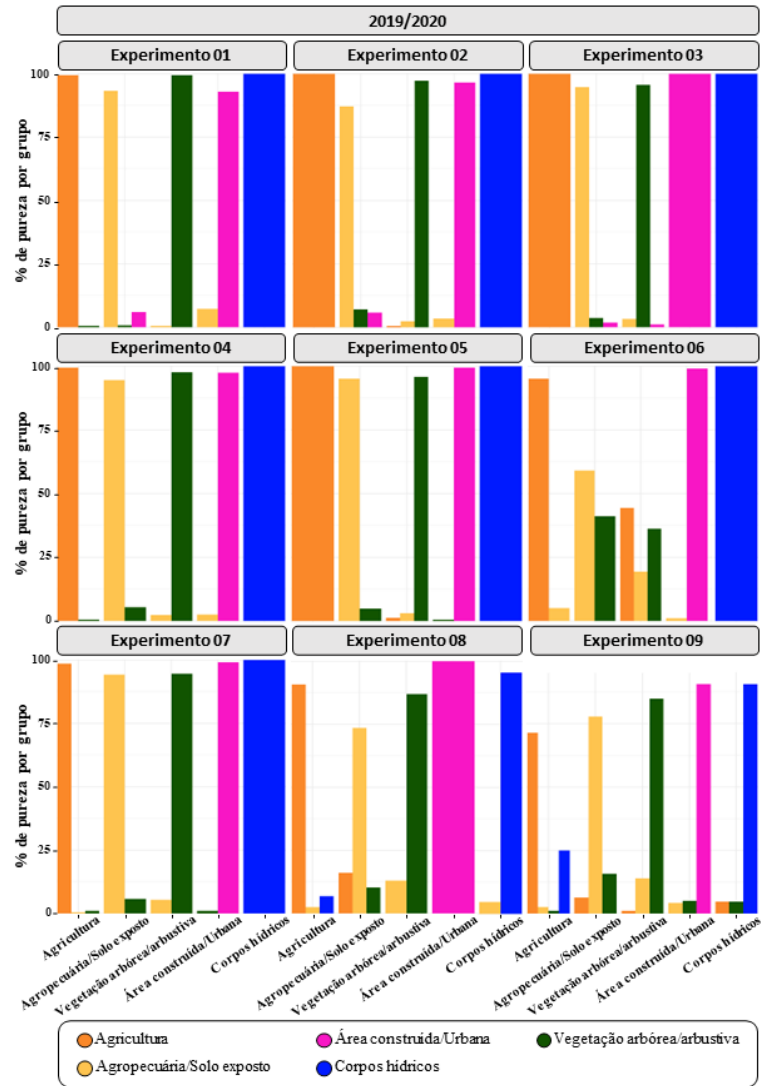
Além disso, em 2018/2019 (Figura 4.6), foi apenas no experimento 01 (bandas) que a classe corpos hídricos não alcançou o percentual máximo de pureza por grupo. Semelhante ao observado em 2017/2018 (Figura 4.5), a partir do experimento 06 (bandas, índices, T ar, UR ar, P e E solo), os valores percentuais de pureza dos grupos das classes Agropecuária/Solo exposto e Vegetação arbórea/arbustiva foram os menores, porém com uma menor variação que 2017/2018, variando de 84,76% (experimento 08, utilizando bandas, índices, T ar, UR ar, P, E solo, T solo e E total) a 98,95% (experimento 06), e de 84,90% (experimento 09, utilizando bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) a 98,50% (experimento 06), respectivamente.

Avaliando o mesmo conjunto de dados para o período 2019/2020 (Figura 4.7), o menor valor percentual de pureza por grupo foi registrado no experimento 06 (bandas, índices, T ar, UR ar, P e E solo), alcançando 36,29%. Da mesma forma que em 2017/2018 (Figura 4.5), este valor foi atribuído ao grupo da classe Vegetação arbórea/arbustiva. Nos experimentos 01 (bandas) e 02 (índices) de 2019/2020, os grupos das classes Agropecuária/Solo exposto (com 93,33% e 87,13%, respectivamente) e Área Construída/Urbana (com 92,85% e 96,53%, respectivamente) registraram os menores percentuais de pureza, seguindo o que ocorreu em 2017/2018 e 2018/2019 (Figura 4.6). No entanto, ao contrário dos anos anteriores, o grupo da classe Vegetação arbórea/arbustiva não alcançou valor máximo de pureza em nenhum experimento de 2019/2020.

Observou-se que nos três primeiros experimentos (Figura 4.7), houve confusão entre o grupo da classe Agropecuária/Solo exposto e os grupos das classes Área Construída/Urbana, com percentuais de 6,00%, 5,84% e 1,77% nos experimentos 01 (bandas), 02 (índices) e 03 (bandas, índices, T ar), respectivamente, e Vegetação arbórea/arbustiva, com percentuais de 0,66%, 7,01% e 3,55% nos os experimentos 01, 02 e 03, respectivamente. Diferente do que ocorreu em 2017/2018 e 2018/2019, a classe Vegetação arbórea/arbustiva não alcançou 100% de pureza em nenhum dos experimentos, registrando seu maior percentual no primeiro experimento, com 99,46%. Além disso, ao contrário dos anos anteriores, no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), o grupo da classe Corpos hídricos não alcançou uma pureza de 100%, mostrando confusão com o grupo da classe Agropecuária/Solo exposto (de 4,62%). No entanto, nos seis primeiros experimentos, esse grupo demonstrou o mais alto nível de pureza, sendo o que mais apresentou esse resultado entre todos os experimentos de 2019/2020.

Os resultados dos percentuais de pureza de 2017/2018, 2018/2019 e 2019/2020 evidenciam a dificuldade em distinguir as classes Agropecuária/Solo Exposto e Área Construída/Urbana, devido às suas semelhanças espectrais (SOUZA; FOLHARINI, 2019; ALI; JOHNSON, 2022; BRITO et al., 2023), mesmo quando aplicado o método de avaliação de amostras desenvolvido por (SANTOS et al., 2021). Por outro lado, indicam que a inclusão de variáveis meteorológicas pode aprimorar a identificação dessas classes. No entanto, é relevante destacar que foram observados percentuais de pureza mais baixos para a classe Agropecuária/Solo Exposto e Vegetação arbórea/arbustiva ao adicionar mais variáveis meteorológicas, especialmente a partir do experimento 06. Isso possivelmente, novamente se deve à multicolinearidade e redundância de informações resultantes da inclusão de muitas variáveis, assim como

Figura 4.7 - Porcentagem de confusão entre os grupos do agrupamento SOM com grade  $15 \times 15$  para 2019/2020.



Fonte: Próprio autor.

à influência das características sazonais e fenológicas da vegetação de caatinga, conforme discutido anteriormente na primeira etapa do SOM (CAO et al., 2017; SILVA et al., 2019a; DAHHANI et al., 2022). A variação nos percentuais de pureza entre os anos 2017/2018, 2018/2019 e 2019/2020 pode ser atribuída a mudanças interanuais, bem como à presença de ruídos e inconsistências em algumas amostras, conforme mencionado anteriormente.

#### 4.4 Ajuste dos modelos

Após a avaliação do percentual de confusão para cada classe, são apresentados os resultados obtidos por meio da validação cruzada *k-fold* ( $k = 5$ ) nos conjuntos de treinamento dos experimentos utilizando os modelos RF e TempCNN, Tabela 4.1. Observa-se que, o modelo RF alcançou uma precisão de 100% em dois experimentos com 1000 árvores em 2017/2018: 06 (bandas, índices, T ar, UR ar, P e E solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), já no experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) houve variação do número de árvores, 100, 500, 1500 e 2000, no período 2018/2019. Por outro lado, nota-se que os menores valores de precisão para o RF foram registrados no primeiro experimento (95,5%, 95,8% e 95,9%, com 2000, 1000 e 500 árvores, respectivamente) em 2019/2020.

A partir do experimento 3 (Tabela 4.1), primeiro a adicionar dados meteorológicos, todos os experimentos subsequentes apresentaram valores de precisão iguais ou superiores a 98,6% (experimento 04, com bandas, índices, T ar e UR ar, em 2019/2020). Além disso, após o terceiro experimento, todos os valores de precisão abaixo de 99,3% foram registrados especificamente em 2019/2020. Dentre os sete últimos experimentos, apenas o experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) alcançou precisão igual ou superior a 99,2% em todos os anos e com todos os números de árvores utilizados. É notável que, para todos os anos e diferentes números de árvores, todos os modelos RF apresentaram acurácia satisfatória.

É importante destacar que a validação cruzada *k-fold* não representa a precisão geral do mapa em dados reais, pois não leva em consideração fatores externos ao processo de treinamento, como a variabilidade espacial e temporal da cobertura da Terra (SIMOES et al., 2021). Porém, fornece uma estimativa do erro de previsão intrínseco do modelo, ou seja, o erro que o modelo cometeria se fosse aplicado aos mesmos dados de treinamento novamente (HASTIE et al., 2009). No estudo conduzido por Brito et al. (2023) no município de Petrolina, foram obtidos valores *k-fold* entre 97,4% e 98,5% em todos os casos ao utilizar séries temporais. Na mesma região, conforme apresentado por Alba et al. (2022), o modelo RF alcançou o maior valor de Kappa em todas as 30 repetições realizadas, apresentando um coeficiente médio de 0,9841.

Na Tabela 4.1, também é possível verificar resultados utilizando a técnica *k-fold* usando o modelo TempCNN, que alcançaram precisão máxima de 99,4%. Esse valor foi observado nos experimentos 05 (composto por bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo) e 07 (bandas, índices, T ar,

Tabela 4.1 - Validação *k-fold* nos conjuntos de treinamento dos experimentos com os modelos RF e TempCNN.

Nº árvores	Random Forest					TempCNN
	Acurácia (%)					
	100	500	1000	1500	2000	
<b>Experimento 01</b>						
2017/2018	96,4	96,1	96,9	96,4	96,4	92,7
2018/2019	97,9	97,8	98,2	98,1	98,3	90,2
2019/2020	96,1	95,8	95,9	96,0	95,5	95,8
<b>Experimento 02</b>						
2017/2018	97,3	97,3	98,0	97,4	97,3	97,3
2018/2019	97,9	98,2	98,1	98,4	97,8	97,4
2019/2020	97,1	97,3	97,2	96,7	97,0	95,9
<b>Experimento 03</b>						
2017/2018	99,9	99,8	99,9	99,8	99,8	98,8
2018/2019	99,4	99,3	99,4	99,5	99,4	98,2
2019/2020	98,9	98,9	99,2	99,0	98,9	98,4
<b>Experimento 04</b>						
2017/2018	99,9	99,8	99,9	99,9	99,9	98,2
2018/2019	99,8	99,7	99,8	99,7	99,9	98,3
2019/2020	98,8	98,6	98,9	98,9	99,0	98,3
<b>Experimento 05</b>						
2017/2018	99,9	99,9	99,9	99,9	99,9	99,1
2018/2019	99,5	99,5	99,4	99,4	99,5	99,4
2019/2020	99,0	99,1	99,1	98,9	99,1	98,4
<b>Experimento 06</b>						
2017/2018	99,9	99,9	100	99,9	99,7	99,4
2018/2019	99,8	99,8	99,8	99,8	99,8	98,6
2019/2020	99,2	99,0	99,2	99,0	98,7	98,9
<b>Experimento 07</b>						
2017/2018	99,6	99,6	99,3	99,3	99,8	98,5
2018/2019	99,7	99,8	99,8	99,8	99,7	99,4
2019/2020	99,2	99,4	99,4	99,5	99,5	98,0
<b>Experimento 08</b>						
2017/2018	99,9	99,8	100	99,6	99,8	96,2
2018/2019	99,9	99,9	99,9	99,9	99,6	96,4
2019/2020	98,8	99,1	99,1	99,1	99,0	98,6
<b>Experimento 09</b>						
2017/2018	99,4	99,6	99,7	99,4	99,6	98,6
2018/2019	100	100	99,8	100	100	98,8
2019/2020	99,0	99,0	99,0	98,8	99,1	95,9

Fonte: Próprio autor.

UR ar, P, E solo e T solo), realizados nos anos 2018/2019, 2017/2018 e 2018/2019, respectivamente. Da mesma forma que ocorreu com o RF, os valores de precisão mais baixos do TempCNN também foram registrados no primeiro experimento (90,2%, 92,7% e 95,8% em 2018/2019, 2017/2018 e 2019/2020, respectivamente).

No entanto, a precisão dos modelos TempCNN foi maior do terceiro ao sétimo ex-



perimento em todos os anos, em comparação com os dois primeiros experimentos (Tabela 4.1). O menor valor registrado nesse intervalo ocorreu no experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), com uma precisão de 98,0% em 2019/2020. No experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), observa-se que o percentual de 95,9% computado em 2019/2020 foi idêntico ao do experimento 02 (índices) para o mesmo ano. Por outro lado, no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), os percentuais registrados em 2017/2018 e 2018/2019 foram inferiores aos do experimento 02 nos mesmos anos. Percebe-se que esse modelo é sensível no quesito adição dos parâmetros.

Apesar dos valores de validação cruzada *k-fold* obtidos com o modelo TempCNN serem satisfatórios, é importante notar que, na maioria dos experimentos, os resultados do TempCNN foram inferiores aos do RF. No entanto, estudos anteriores demonstraram resultados superiores ao RF em outras regiões. A pesquisa conduzida por Pelletier et al. (2019), utilizaram os modelos RF e TempCNN para analisar o uso de dados do Sentinel-2 no mapeamento do estado de Victoria, na Austrália, revelou que o TempCNN alcançou os maiores valores de Acurácia Global (AG), Média de F-score e Acurácia Média (AM), com diferenças respectivas de 1,65%, 1,98% e 3,42% em comparação ao RF. Em estudo anterior no Sudoeste da França, Pelletier et al. (2019) concluiu que o modelo TempCNN superou os algoritmos RF e RNN em uma margem de 1% a 3% na precisão geral.

Após a avaliação do *k-fold* em cada experimento, a seguir são apresentados os resultados de precisão para cada classe por meio da técnica *F1-score*. No modelo RF observa-se que, os menores percentuais de precisão foram registrados no primeiro experimento para as classes Agropecuária/Solo exposto (89,1% e 90,8%, em 2019/2020 e 2017/2018, respectivamente) e Corpos hídricos (91,9% em 2019/2020) (Tabela 4.2). Por outro lado, observa-se que os experimentos 06 (bandas, índices, T ar, UR ar, P e E solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) alcançaram uma precisão de 100% para todas as classes em 2017/2018, 2017/2018 e 2018/2019, respectivamente.

Um percentual de precisão de 100% foi alcançado por ambas as classes Área construída/Urbana e Vegetação arbórea/arbustiva nos experimentos 03 (bandas, índices e T ar) a 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) (Tabela 4.2). Além disso, a classe Área

Tabela 4.2 - *F1-scores* no conjunto de treinamento dos experimentos com os modelos RF.

Random Forest			
Acurácia (%)			
Classes	2017/2018	2018/2019	2019/2020
<b>Experimento 01</b>			
Agricultura	99,3	95,1	97,9
Agropecuária/Solo exposto	90,8	100	89,1
Vegetação arbórea/arbustiva	99,5	100	99,8
Área construída/Urbana	100	100	99,7
Corpos hídricos	92,1	94,9	91,9
<b>Experimento 02</b>			
Agricultura	94,2	95,4	92,1
Agropecuária/Solo exposto	100	100	96,5
Vegetação arbórea/arbustiva	99,7	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	94,5	95,4	96,6
<b>Experimento 03</b>			
Agricultura	99,7	99,5	98,1
Agropecuária/Solo exposto	100	99,0	97,6
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	99,7	99,0	99,7
<b>Experimento 04</b>			
Agricultura	100	99,7	98,0
Agropecuária/Solo exposto	99,7	100	97,3
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	99,7	99,7	99,5
<b>Experimento 05</b>			
Agricultura	100	99,5	98,5
Agropecuária/Solo exposto	99,7	99,0	97,2
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	99,7	99,0	99,2
<b>Experimento 06</b>			
Agricultura	100	99,7	97,6
Agropecuária/Solo exposto	100	99,5	97,4
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	100	99,7	99,5
<b>Experimento 07</b>			
Agricultura	99,0	99,7	98,3
Agropecuária/Solo exposto	100	99,4	97,9
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	99,5	99,8	99,7
<b>Experimento 08</b>			
Agricultura	100	100	96,7
Agropecuária/Solo exposto	100	99,7	100
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	96,8
Corpos hídricos	100	99,8	99,5
<b>Experimento 09</b>			
Agricultura	98,5	100	96,8
Agropecuária/Solo exposto	100	100	97,0
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	100
Corpos hídricos	99,5	100	99,7

Fonte: Próprio autor.

construída/Urbana obteve precisão de 100% no segundo experimento, enquanto a Vegetação arbórea/arbustiva alcançou o mesmo percentual no experimento 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total). É observado que, somente no ano de 2019/2020, no experimento 01 (bandas), nenhuma classe alcançou precisão máxima, com o valor percentual mais alto sendo obtido pela classe Vegetação arbórea/arbustiva de 99,8%. No trabalho realizado por Brito et al. (2023), na cidade de Petrolina, foi constatado que o uso de séries temporais e RF resultou em valores de *F1-score* acima de 90%. Esses resultados são consistentes com os obtidos neste trabalho, com exceção da classe Agropecuária/Solo exposto em 2019/2020.

Utilizado o modelo TempcNN do *F1-score* (Tabela 4.3), verifica-se que, semelhante aos resultados do RF (Tabela 4.2), as menores precisões foram registradas no primeiro experimento, com percentuais de 68,2% para Corpos hídricos em 2018/2019, 75,1% para Agricultura, também em 2018/2019 e 78,3% para Agropecuária/Solo exposto e Corpos hídricos, em 2017/2018. Porém, diferentemente do que foi observado nos resultados com o modelo RF, nenhum dos experimentos alcançou precisão máxima em todas as classes.

Observa-se na Tabela 4.3, que as classes Área construída/Urbana (experimento 08 - bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e Vegetação arbórea/arbustiva (experimentos 03 - bandas, índices e T ar; 04 - bandas, índices, T ar e UR ar; 05 - bandas, índices, T ar, UR ar e P; 06 - bandas, índices, T ar, UR ar, P e E solo; e 09 - bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) foram as únicas classes a alcançar precisão de 100% em todos os anos. Nota-se também que, usando o modelo TempCNN, nenhuma das classes alcançou precisão de 100% nos experimentos 01 (bandas) e 02 (índices) em 2018/2019 e 2019/2020, respectivamente. Os maiores valores de precisão foram registrados para as classes Vegetação arbórea/arbustiva, atingindo 99,5% em 2018/2019 e Área construída/Urbana, com 99,3% em 2019/2020.

Embora o modelo TempCNN tenha uma arquitetura específica para lidar com séries temporais, capturando padrões e relações temporais através das convoluções (PELLETIER et al., 2019), os resultados desta pesquisa indicam que, em comparação com o RF, o modelo TempCNN pode encontrar mais desafios na identificação das classes na área de estudo, especialmente nos experimentos que não utilizam variáveis meteorológicas. No estudo conduzido por Pelletier et al. (2019), a análise das métricas por classe revelou que o TempCNN alcançou os maiores valores de *F1-score* para todas as classes em comparação com o RF, exceto para solo exposto e áreas

Tabela 4.3 - *F1-scores* no conjunto de treinamento dos experimentos com os modelos TempCNN.

TempCNN			
Acurácia (%)			
Classes	2017/2018	2018/2019	2019/2020
<b>Experimento 01</b>			
Agricultura	100	75,1	87,0
Agropecuária/Solo exposto	78,3	99,2	98,2
Vegetação arbórea/arbustiva	100	99,5	100
Área construída/Urbana	99,7	99,2	99,8
Corpos hídricos	78,3	68,2	91,1
<b>Experimento 02</b>			
Agricultura	92,6	99,7	92,2
Agropecuária/Solo exposto	99,5	93,3	97,0
Vegetação arbórea/arbustiva	99,5	99,5	97,5
Área construída/Urbana	100	100	99,3
Corpos hídricos	93,3	92,9	92,4
<b>Experimento 03</b>			
Agricultura	96,7	95,7	97,5
Agropecuária/Solo exposto	100	100	95,7
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	99,8	100
Corpos hídricos	97,0	95,4	98,3
<b>Experimento 04</b>			
Agricultura	95,1	96,0	95,4
Agropecuária/Solo exposto	100	99,8	98,0
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	99,5	99,7
Corpos hídricos	95,6	96,0	97,7
<b>Experimento 05</b>			
Agricultura	97,6	98,5	94,9
Agropecuária/Solo exposto	100	100	98,2
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	100	99,8
Corpos hídricos	97,8	98,5	97,8
<b>Experimento 06</b>			
Agricultura	98,2	97,1	98,4
Agropecuária/Solo exposto	100	99,5	96,0
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	100	99,5	100
Corpos hídricos	98,5	96,7	98,8
<b>Experimento 07</b>			
Agricultura	100	98,3	97,4
Agropecuária/Solo exposto	94,3	99,7	91,5
Vegetação arbórea/arbustiva	99,8	100	100
Área construída/Urbana	100	100	99,8
Corpos hídricos	97,0	98,8	97,0
<b>Experimento 08</b>			
Agricultura	88,7	90,2	96,9
Agropecuária/Solo exposto	100	100	94,8
Vegetação arbórea/arbustiva	99,5	99,8	99,5
Área construída/Urbana	100	100	100
Corpos hídricos	91,7	91,5	98,7
<b>Experimento 09</b>			
Agricultura	100	95,0	96,5
Agropecuária/Solo exposto	99,7	100	84,6
Vegetação arbórea/arbustiva	100	100	100
Área construída/Urbana	92,9	99,6	99,7
Corpos hídricos	97,5	97,8	93,3

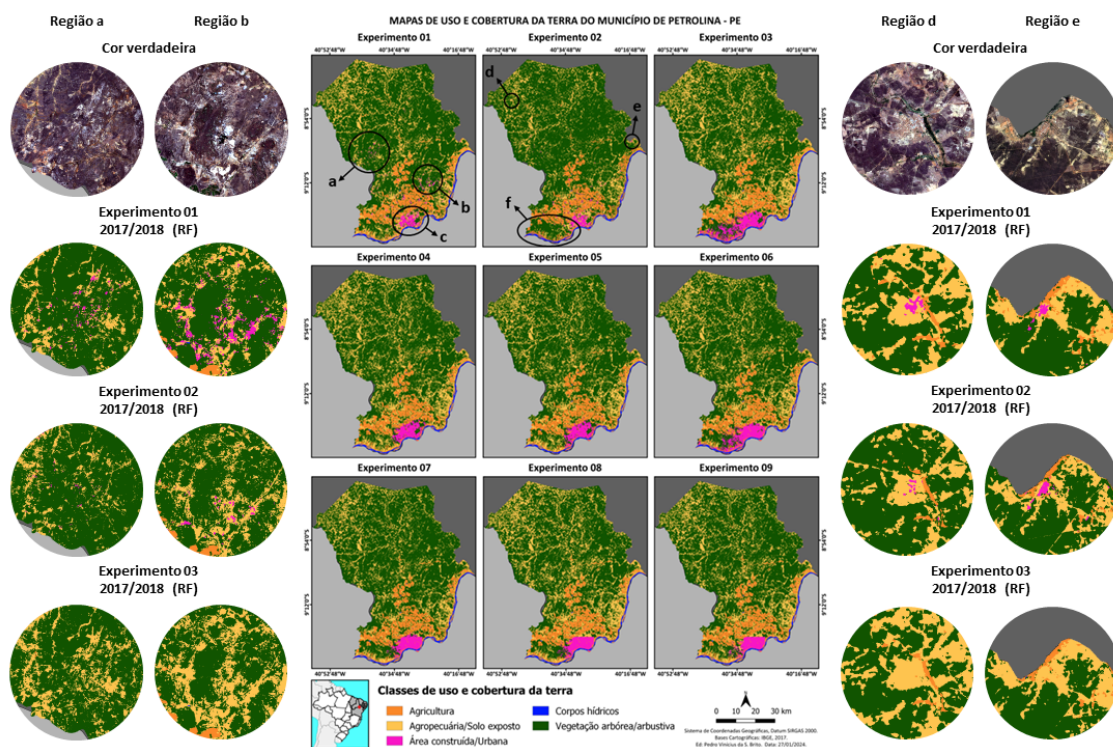
Fonte: Próprio autor.

agrícolas de canola, onde as diferenças foram em torno de 0,5%.

#### 4.5 Classificação dos mapas de LULC

Após avaliar e analisar cada experimento e classe em cada etapa anterior, a seguir são apresentados os mapas de uso e cobertura da Terra de 2017/2018, utilizando o modelo RF na Figura 4.8. Os resultados indicam que as classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco, foram melhor identificadas em todos os experimentos. Resultados semelhantes foram encontrados na mesma área de estudo ao utilizar o modelo RF, por Alba et al. (2022) e Brito et al. (2023).

Figura 4.8 - Classificações de uso e cobertura da Terra para 2017/2018 via RF para Petrolina-PE.



Fonte: Próprio autor.

Entretanto, no primeiro experimento (bandas) observa-se que algumas regiões de Agropecuária/Solo exposto e Vegetação arbórea/arbustiva foram erroneamente classificadas como Área construída/urbana (regiões a e b). Como mencionado ante-

riormente, é comum encontrar confusões entre essas classes na região de estudo, conforme discutido por [Silva et al. \(2019\)](#), [Souza e Folharini \(2019\)](#) e [Brito et al. \(2023\)](#). Na região (b), algumas áreas de Solo exposto, classificadas como Área construída/urbana, correspondem, em alguns casos, a locais de expansão urbana. Isso destaca a capacidade do modelo em identificar áreas propensas à futura urbanização.

No segundo experimento (Figura 4.8), observa-se uma redução das confusões entre Agropecuária/Solo exposto e Vegetação arbórea/arbustiva com Área construída/urbana nas regiões (a e b) em comparação com o primeiro experimento. Como observado por [Ganjirad e Bagheri \(2024\)](#) é possível que o uso exclusivo de bandas espectrais não seja suficiente para classificar algumas regiões, sendo essencial integrar índices espectrais para realçar características específicas e fornecer dados mais informativos ao algoritmo de classificação. Em consequência disso, gerar classificações mais precisas de diferentes classes de LULC.

Além disso, no segundo experimento (Figura 4.8), as áreas de mineração na região (b) foram precisamente identificadas utilizando índices espectrais (experimento 02) comparado ao uso de bandas (experimento 01). O modelo RF classificou essas áreas como Área construída/urbana devido à sua similaridade espectral com regiões construídas e urbanas, já que são utilizadas para extração de pedras e outros materiais destinados à construção, incluindo o processo de britamento desses materiais ([SILVA et al., 2023](#)).

Na Figura 4.8, observa-se outra região (c) e percebe-se uma subestimativa da classe Área construída/Urbana na área urbana da cidade de Petrolina nos dois primeiros experimentos, com 81,00 e 63,95  $km^2$ , respectivamente. De acordo com os dados do [IBGE \(2019\)](#), a Área urbanizada da cidade de Petrolina é de aproximadamente 86,39  $km^2$ . Contudo, na mesma região, a área urbana foi superestimada quando variáveis meteorológicas foram adicionadas nos experimentos 03 (bandas, índices e T ar), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total), com 139,62; 126,00; 88,71 e 91,25  $km^2$ , respectivamente.

Comparados com outros experimentos em 2017/2018 região (c) (Figura 4.8), os experimentos 04 (bandas, índices, T ar e UR ar) e 05 (bandas, índices, T ar, UR ar e P) apresentaram resultados mais condizentes com a realidade para área urbana da cidade de Petrolina, com 83,93 e 86,12  $km^2$ , respectivamente. A partir do experimento 06, nota-se que, ao adicionar mais variáveis meteorológicas os detalhes finos da classe Área construída/Urbana foram removidos (região c), principalmente nos

experimentos 08 e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) ( $81,76 \text{ km}^2$ ). Isso ocorreu possivelmente devido à baixa resolução espacial dos produtos meteorológicos.

Por outro lado, nos dois primeiros experimentos, o distritos de Rajada (região d) e o povoado de Izacolândia (região e) foram identificados, apesar de suas áreas relativamente pequenas e de estarem distantes da área urbana da cidade (Figura 4.8). Isso demonstra a precisão dos modelos RF na identificação da classe de Área construída/Urbana. Além disso, observa-se que no distrito de Rajada, o primeiro experimento identificou uma área maior dessa classe em comparação com o segundo experimento. No entanto, no povoado de Izacolândia, o segundo experimento conseguiu identificar mais regiões da classe Área construída/Urbana e também demonstrou maior precisão em separar as regiões de Área construída/Urbana e Agropecuária/Solo exposto quando próximas uma da outra, comparado ao primeiro experimento.

A partir do terceiro experimento, o distrito de Rajada e o povoado de Izacolândia não foram mais identificados (Figura 4.8). Além disso, é observado que o segundo experimento teve um desempenho superior na identificação das áreas de Vegetação arbórea/arbustiva em comparação com o primeiro, especialmente nas áreas de Vegetação arbustiva (região f). Esses resultados podem ser atribuídos principalmente à utilização de índices espectrais, em particular o SAVI, que reduz o efeito do solo no sinal da vegetação (SANTOS et al., 2014).

Nessa mesma região no experimento 02 (Figura 4.8), nota-se que nos experimentos 03 (bandas, índices e T ar), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), áreas de Vegetação arbustiva foram erroneamente classificadas como Área construída/Urbana, enquanto nos experimentos 04 (bandas, índices, T ar e UR ar) e 05 (bandas, índices, T ar, UR ar e P) parte dessas áreas foram classificadas como Agropecuária/Solo exposto, classificação semelhante ao primeiro experimento. Como visto anteriormente, as confusões entre essas classes são frequentemente encontradas na área de estudo (SILVA et al., 2019; SOUZA; FOLHARINI, 2019; BRITO et al., 2023).

Na Figura 4.8, as áreas de Agropecuária/Solo exposto e Vegetação arbórea/arbustiva foram erroneamente classificadas como Área construída/urbana nas regiões a e b, não foram mais identificadas a partir do experimento 03 (bandas, índices e T ar).

Embora áreas de Vegetação arbustiva esparsa estejam sendo erroneamente classificadas como Agropecuária/Solo exposto em algumas áreas das regiões a e b, fica evidente que o uso de variáveis meteorológicas pode aprimorar a identificação da classe Agropecuária/Solo exposto em diferentes regiões do município. Este resultado é possivelmente atribuído aos padrões de refletância semelhantes das classes em determinadas épocas do ano na região, conforme mencionado por [Alba et al. \(2022\)](#).

Avaliando a classificação para o período 2018/2019 (Figura 4.9), são apresentados os mapas de uso e cobertura do solo classificados utilizando o modelo RF. Observa-se que, assim como em 2017/2018 (Figura 4.8), as classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco foram melhor identificadas em todos os experimentos. Além disso, no primeiro experimento de 2018/2019 (Figura 4.9), o modelo RF conseguiu identificar uma pequena área habitada e isolada que pertence ao distrito de Rajada (região g). Isso evidencia mais uma vez a capacidade do modelo em identificar áreas construídas mesmo que sejam pequenas e estejam afastadas da área urbanas da cidade.

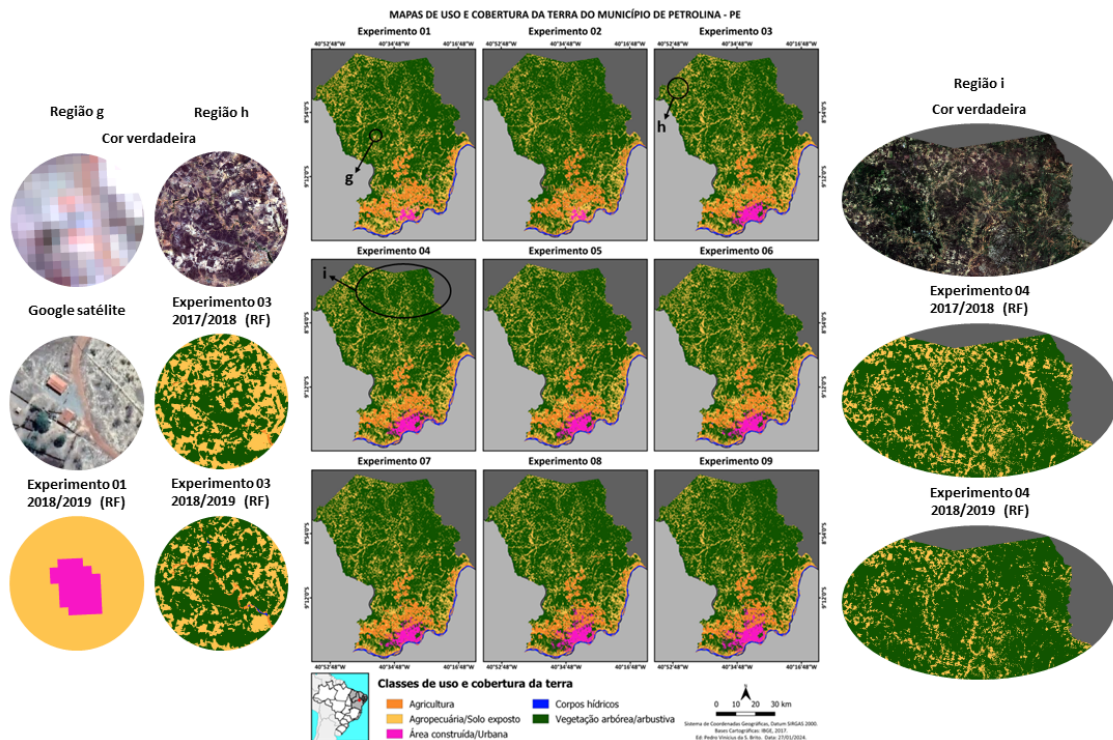
Realizando monitoramento da expansão urbana em Kumasi, localizada ao norte da capital de Gana, entre 1986 e 2015, e prevendo futuras áreas urbanas/construídas para 2025 utilizando imagens de satélite Landsat e outros conjuntos de dados relevantes, [Frimpong e Molkenthin \(2021\)](#) observaram que, o classificador RF demonstrou uma performance superior em comparação aos métodos paramétricos tradicionais, conforme evidenciado por seus resultados de precisão. Os autores também enfatizaram a eficácia do RF como uma ferramenta eficiente para rastrear e prever áreas urbanas/construídas.

No período de 2018/2019 (Figura 4.9), em todos os experimentos foram identificadas mais áreas de Corpos hídricos (região h) comparado ao período anterior (Figura 4.8). Além disso, destaca-se uma maior identificação de áreas da classe Vegetação arbórea/arbustiva nos experimentos de 2018/2019 em comparação com 2017/2018, principalmente nas direções Norte e Nordeste do município (região i). A principal causa na maior identificação de áreas de Corpos hídricos e da classe Vegetação arbórea/arbustiva em 2018/2019 pode ser a maior ocorrência de chuvas no município entre os anos de 2018 e 2019, resultando no aumento do nível dos reservatórios da região e mudança na vegetação, um vez que as vegetações naturais da caatinga conseguem recuperar-se rapidamente ([SOUSA et al., 2021](#)).

As confusões que ocorreram nos experimentos 01 (bandas) e 02 (índices) nas re-



Figura 4.9 - Classificações de uso e cobertura da Terra para 2018/2019 via RF para Petrolina-PE.



Fonte: Próprio autor.

giões a e b em 2017/2018 (Figura 4.8), onde áreas de Agropecuária/Solo exposto e Vegetação arbustiva esparsa foram erroneamente classificadas como Área construída/urbana, foram menores nos mesmos experimentos em 2018/2019 (Figura 4.9), chegando praticamente ser ausente. Esse resultado pode estar relacionado à variabilidade interanual da precipitação, pois com a maior ocorrência de chuva, a vegetação tende a desenvolver suas folhas e, conseqüentemente, se diferenciar melhor de outras classes, como Agropecuária/Solo exposto e Área construída/urbana (BARBOSA et al., 2019; SOUSA et al., 2021).

Como em 2017/2018 (região b) (Figura 4.8), os dois primeiros experimentos de 2018/2019 identificaram as áreas de mineração, classificando-as erroneamente como Área construída/urbana, porém o experimento 01 (bandas) conseguiu mapear mais dessas áreas com maior precisão em comparação ao experimento 02 (índices), no mesmo ano (Figura 4.9). Além disso, ainda na região b, observa-se que no experimento 02 de 2018/2019, o modelo não conseguiu identificar todos os reservatórios

como Corpos hídricos, ao contrário dos outros experimentos no mesmo ano. Ademais, nos dois primeiros experimentos de 2018/2019, algumas margens de reservatórios foram erroneamente classificadas como Área construída/Úrbana (região b), mas a partir do terceiro experimento, com o uso de variáveis meteorológicas, esses erros não ocorreram mais. Mais uma vez, esse resultado demonstra a eficiência do modelo RF e das variáveis meteorológicas para distinguir Solo exposto de Área construída/Úrbana em diferentes regiões do município.

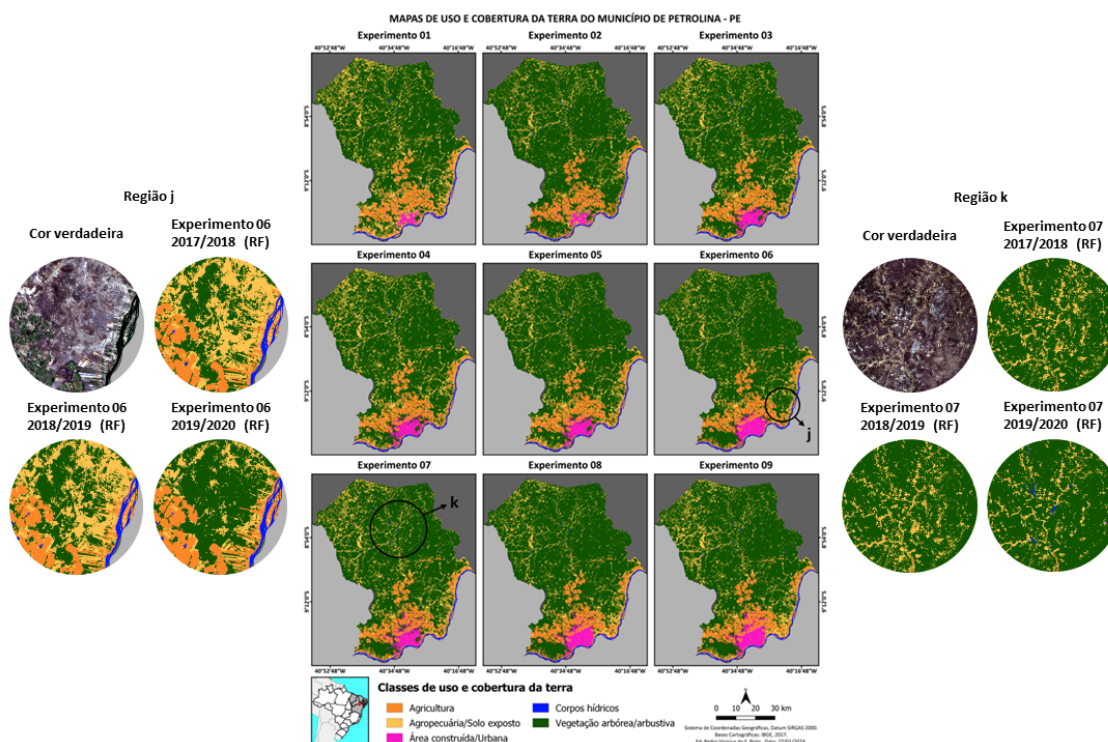
Comparando os dois primeiros experimentos realizados em 2017/2018 (Figura 4.8), nota-se que nos experimentos correspondentes de 2018/2019 (Figura 4.9), houve uma redução na identificação da área urbana da cidade de Petrolina, com valores de 44,77 e 36,69  $km^2$ , respectivamente (região c). Uma das possíveis causas é a maior presença de nuvens nas imagens de satélite em 2018/2019 na região (c). Apesar da remoção dos pixels com nuvens das imagens de satélite e da aplicação da interpolação linear nas séries temporais, a qualidade dos dados interpolados pode ser inferior à dos dados originais. Isso ocorre porque as informações sobre as características do terreno não podem ser completamente recuperadas e a interpolação linear tende a suavizar a série temporal, reduzindo a variabilidade e dificultando a identificação de mudanças no estado da superfície terrestre (MORENO-MARTÍNEZ et al., 2020).

Em 2018/2019 (Figura 4.9) o experimento 03 (bandas, índices e T ar) apresentou um reconhecimento mais próximo da realidade da área urbana da cidade de Petrolina, com 94,78  $km^2$  (região c). Por outro lado, os experimentos 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) superestimaram significativamente a região, com 105,12; 106,17; 109,21; 108,75; 133,68 e 142,57  $km^2$ , respectivamente.

A seguir tem-se as classificações de uso e cobertura da Terra para o período de 2019/2020 usando o modelo RF. Percebe-se que na Figura 4.10, assim como nos experimentos dos anos anteriores (2017/2018 e 2018/2019, Figura 4.8 e 4.9, respectivamente) mais uma vez representaram bem visivelmente as classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco. Além disso, nesse período foram identificadas áreas mais extensas de Vegetação arbórea/arbustiva em todos os experimentos em comparação com os períodos de 2017/2018 e 2018/2019, principalmente nas regiões (f) e (j).

Diferentemente do que ocorreu nos dois primeiros experimentos de 2017/2018 (Fi-

Figura 4.10 - Classificações de uso e cobertura da Terra para 2019/2020 via RF para Petrolina-PE.



Fonte: Próprio autor.

gura 4.8) e 2018/2019 (Figura 4.9) nas regiões a e b, em 2019/2020 não foram observadas áreas de Vegetação arbustiva esparsa classificadas como Área construída/urbana (Figura 4.10). Além disso, observa-se uma maior identificação de corpos hídricos em todos os experimentos na região (k). Por outro lado, no primeiro experimento foram identificadas algumas áreas pequenas de reservatórios e suas margens (região a) classificadas como Área construída/urbana. Na região (b), as áreas de mineração foram melhor identificadas, porém ainda classificadas como Área construída/urbana, como ocorreu em 2017/2018 e 2018/2019.

Nos dois primeiros experimentos de 2019/2020 (Figura 4.10), é evidente um aumento na identificação da área urbana da cidade de Petrolina (região c) em comparação com os dois primeiros experimentos de 2018/2019 (Figura 4.9, respectivamente), com uma diferença de 9,66 e 22,29  $km^2$ , respectivamente. Além da extensão da área urbana, esse incremento na identificação pode ser atribuído à reduzida ocorrência ou ausência de nuvens na região como mencionado anteriormente (MORENO-MARTÍNEZ et al.,

2020).

Assim como nos anos de 2017/2018 e 2018/2019 (Figuras 4.8 e 4.9, respectivamente), a área urbana em 2019/2020 foi superestimada quando variáveis meteorológicas foram adicionadas (Figura 4.10), principalmente nos experimentos 06 (bandas, índices, T ar, UR ar, P e E solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), com 121,76; 132,22 e 139,72  $km^2$ , respectivamente (região c). Em comparação com outros experimentos no mesmo ano, os experimentos 03 (bandas, índices e T ar) e 05 (bandas, índices, T ar, UR ar e P) com 92,08 e 96,17  $km^2$ , respectivamente, apresentaram resultados próximos a realidade da área urbana da cidade de Petrolina.

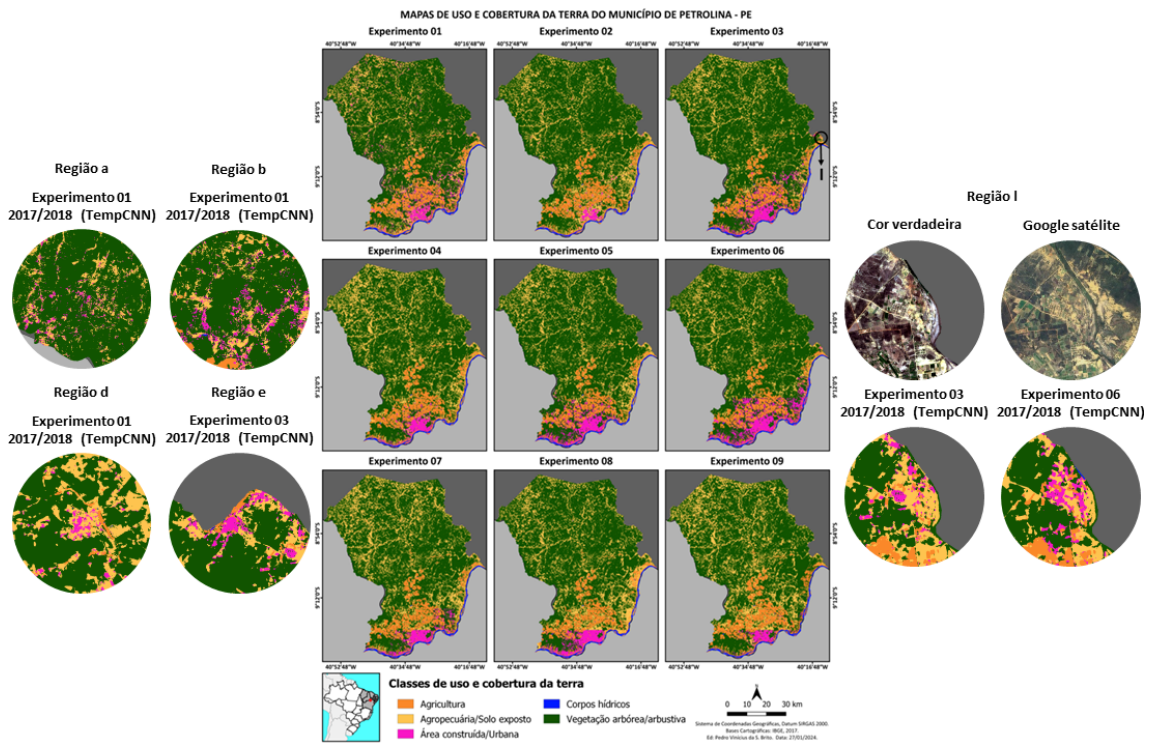
O distrito de Rajada (região d) e o povoado de Izacolândia (região e) novamente não foram identificados nos experimentos 03 (bandas, índices e T ar) a 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) (Figura 4.10). No entanto, o uso de índices permitiu identificar mais áreas desses distritos em 2019/2020 (regiões d e e) em comparação com o uso de bandas. Além disso, em 2019/2020 constata-se que a incorporação de variáveis meteorológicas contribuiu para a identificação da classe Agropecuária/Solo exposto (Figura 4.10).

Na Figura 4.11, são apresentados os mapas de uso e cobertura da Terra de 2017/2018 classificados pelo modelo TempCNN. Observa-se que, semelhante aos resultados obtidos com o modelo RF (Figuras 4.8), esses também evidenciaram boa identificação das classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco. No entanto, no primeiro experimento de 2017/2018 com o TempCNN, notou-se maior ocorrência de áreas de Agropecuária/Solo exposto e Vegetação arbustiva esparsa erroneamente classificadas como Área construída/urbana nas regiões (a e b), em comparação com o modelo RF no mesmo período.

As confusões entre as classes mencionadas anteriormente (regiões a e b) ao utilizar o TempCNN, podem ser atribuídas à capacidade do modelo em incorporar informações temporais (Figura 4.11). O modelo TempCNN pode encontrar dificuldades em distinguir entre classes com padrões temporais semelhantes, como áreas de Agropecuária/Solo exposto, Vegetação arbustiva esparsa e áreas Construídas/Urbanas em determinados momentos em função das mudanças sazonais na região de estudo (PELLETIER et al., 2019; ALBA et al., 2022; BRITO et al., 2023).

No segundo experimento de 2017/2018 utilizando índices espectrais e modelo TempCNN (Figura 4.11), observa-se uma redução nas confusões entre as classes

Figura 4.11 - Classificações de uso e cobertura da Terra para 2017/2018 via TempCNN para Petrolina-PE.



Fonte: Próprio autor.

de Agropecuária/Solo exposto e Vegetação arbustiva esparsa com a Área construída/urbana nas regiões (a e b), em comparação com o experimento 01 (bandas). Esse resultado é semelhante ao obtido com o modelo RF para o mesmo ano e experimento (Figuras 4.8). No entanto, com o modelo RF, tais confusões foram ainda menos frequentes. As áreas de mineração (região b) que foram identificadas e classificadas como Área construída/urbana pelo modelo RF, também foram classificadas dessa forma com o modelo TempCNN, nos experimentos 01, 02 (índices), 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar) e 05 (bandas, índices, T ar, UR ar e P) de 2017/2018.

No primeiro experimento de 2017/2018, observa-se que a área urbana da cidade de Petrolina apresentou maior expansão usando TempCNN comparado ao RF (Figuras 4.11 e 4.8, respectivamente), com 219,93 e 55,37  $km^2$  nos experimentos 01 e 02, respectivamente (região c). No entanto, ao utilizar índices espectrais não houve ganho na identificação da área urbana de Petrolina com o TempCNN em comparação

ao RF em 2017/2018. Assim como nos resultados obtidos com o RF, verifica-se que a adição de variáveis meteorológicas resultou em uma superestimativa da área urbana nos experimentos de 2017/2018 com o TempCNN, principalmente nos experimentos 03 (bandas, índices e T ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) com 198,40; 182,11; 268,89 e 128,55  $km^2$ , respectivamente.

Além disso, a partir do experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) os detalhes finos da área urbana foram perdidos (região c) (Figuras 4.11). A principal causa pode ser atribuída à resolução espacial baixa dos produtos meteorológicos conforme mencionado anteriormente. Por outro lado, o experimento 04 (bandas, índices, T ar e UR ar) apresentou resultados mais condizentes com a realidade da área urbana da região, com 95,22  $km^2$ .

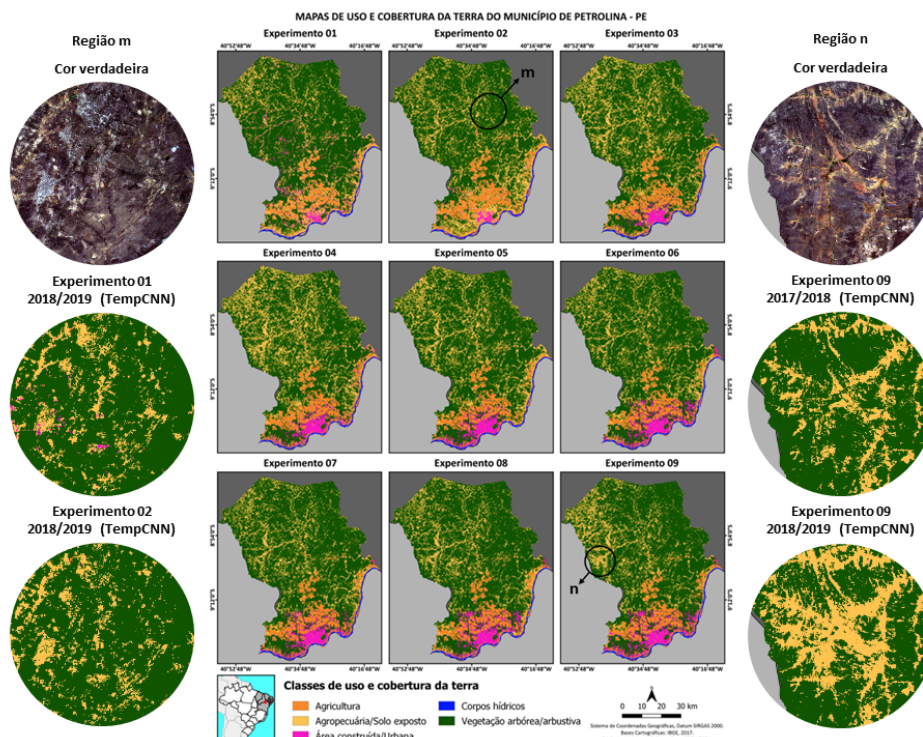
Ambos o distrito de Rajada (região d) e o povoado de Izacolândia (região e) foram identificados nos dois primeiros experimentos de 2017/2018 com o modelo TempCNN (Figuras 4.11), assim como ocorreu com o RF para o mesmo ano (Figura 4.8). No entanto, em comparação com o RF, o TempCNN identificou mais áreas do distrito de Rajada (região d) no primeiro experimento. Semelhante aos resultados obtidos com o modelo RF, a partir do terceiro experimento, o distrito de Rajada (região d) não foi mais identificado com o modelo TempCNN. Entretanto, o povoado de Izacolândia (região e) foi identificado no terceiro experimento com o TempCNN em 2017/2018. Além disso, o modelo TempCNN conseguiu identificar mais áreas de distrito. Percebe-se pela primeira vez em todos os anos e experimentos apresentados anteriormente, no terceiro experimento, a Agrovila São Francisco (região l). Ela também foi reconhecida no experimento 06 (bandas, índices, T ar, UR ar, P e E solo).

Observa-se na região (f), que nos experimentos com o modelo TempCNN em 2017/2018 (Figuras 4.11) houve áreas de Vegetação arbustiva e solo exposto classificadas erroneamente como Área construída/urbana, especialmente nos experimentos 03 (bandas, índices e T ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total). No entanto, os experimentos com o modelo TempCNN conseguiram identificar mais áreas de Vegetação arbustiva do que o RF em 2017/2018 (Figura 4.8), principalmente nos experimentos 03, 04 (bandas, índices, T ar e UR ar), 05, 07, 08 e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total, e U solo). Esse resultado sugere que o uso de

variáveis meteorológicas com o modelo TempCNN pode contribuir para distinguir a classe de Vegetação arbustiva de solo exposto, uma das principais dificuldades na região (ALBA et al., 2022; BRITO et al., 2023).

Observando os mapas de uso e cobertura da Terra para 2018/2019 classificados com o modelo TempCNN (Figura 4.12), verifica-se que as classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco foram identificadas em todos os experimentos como nos casos anteriores. No entanto, assim como no primeiro experimento de 2017/2018 com o modelo TempCNN (Figuras 4.11), em 2018/2019 utilizando o mesmo modelo, observou-se áreas de Agropecuária/Solo exposto e Vegetação arbustiva esparsa classificadas erroneamente como Área construída/urbana (regiões a e b). Observa-se também que, no segundo experimento de 2018/2019 as confusões com o modelo TempCNN foram maiores do que com o RF no mesmo ano nas regiões a e b (4.9).

Figura 4.12 - Classificações de uso e cobertura da Terra para 2018/2019 via TempCNN para Petrolina-PE.



Fonte: Próprio autor.

Em 2018/2019 utilizando o modelo TempCNN (Figura 4.12), observa-se nos experimentos 02 (índices) (regiões f e m), 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo) e 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) (região m), e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) (região n), redução da área da classe Vegetação arbórea/arbustiva e aumento nas áreas da classe de Agropecuária/Solo exposto em comparação com os mesmos experimentos realizados em 2017/2018 (Figuras 4.11).

Nos dois primeiros experimentos de 2018/2019 (Figura 4.12), verifica-se que a maioria das áreas de corpos hídricos identificadas pelo RF não foram reconhecidas pelo TempCNN (região h) (Figura 4.9). Assim como ocorreu em 2017/2018 com o TempCNN (Figuras 4.11), em 2018/2019 áreas consideradas de mineração foram identificadas e classificadas erroneamente como Área construída/urbana nos experimentos 01 (bandas), 02 (índices), 03 (bandas, índices e T ar), 04 (bandas, índices, T ar e UR ar), 05 (bandas, índices, T ar, UR ar e P) e 07 (bandas, índices, T ar, UR ar, P, E solo e T solo).

Similar ao que ocorreu com o modelo RF no primeiro experimento (Figura 4.9), com o TempCNN em 2018/2019 (Figura 4.12) a área urbana da cidade de Petrolina foi menor comparada a de 2017/2018 de acordo com o resultado citado anteriormente do IbGE, com  $168,72 \text{ km}^2$  (região c). Nota-se que o experimentos 03 (bandas, índices e T ar) em 2018/2019 usando o TempCNN teve uma classificação mais próxima da realidade da área urbana da cidade. Por outro lado, nos experimentos 05 (bandas, índices, T ar, UR ar e P), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), a região foi bastante superestimada, assim como ocorreu com o RF no mesmo ano.

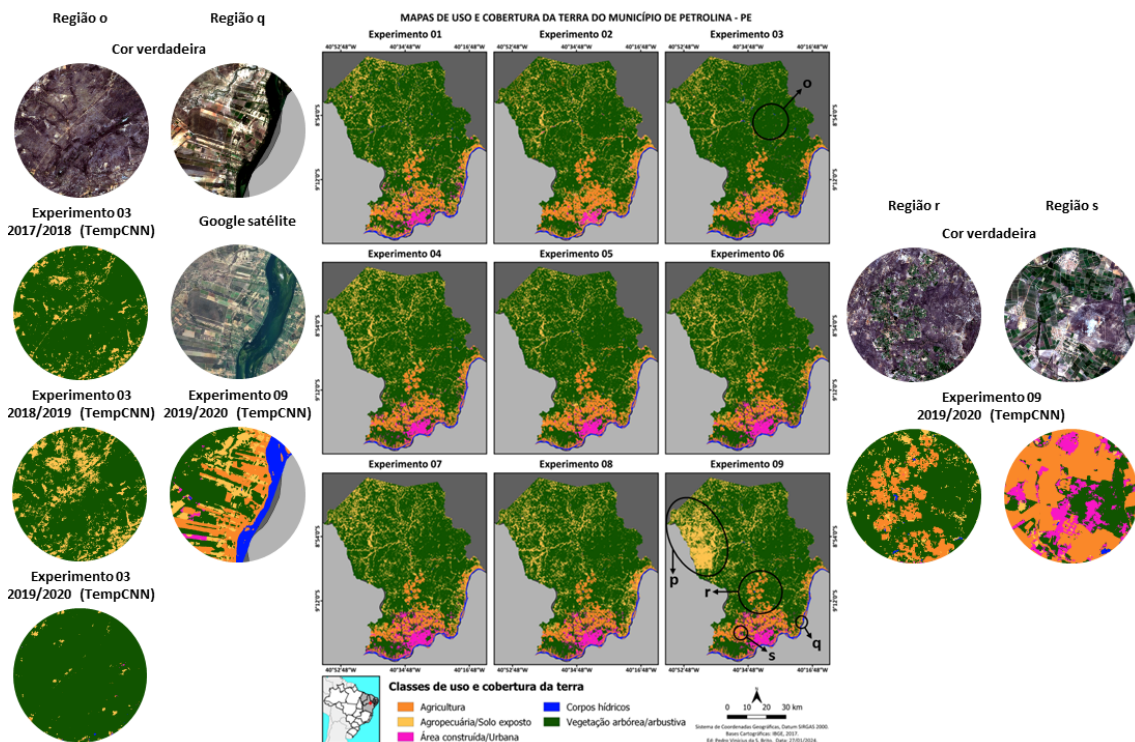
Como também visto em 2017/2018 (Figuras 4.11), áreas do distrito de Rajada (região d) foram identificadas nos dois primeiros experimentos e do povoado de Izacolândia (região e) nos quatro primeiros, com uma identificação mais precisa do distrito de Izacolândia no terceiro experimento. Como observado em outros experimentos apresentados acima, ao sul do município (região f) foram identificadas áreas de Vegetação arbustiva esparsa classificadas erroneamente como Área construída/urbana, especialmente nos seis últimos experimentos.

Os mapas de uso e cobertura da Terra de 2019/2020 classificados com o modelo



TempCNN, são apresentados na Figura 4.13. Nos experimentos, nota-se uma melhoria no reconhecimento das classes de Agricultura, Vegetação arbórea/arbustiva e passagem de água do rio São Francisco em comparação com os anos anteriores (2017/2018 e 2018/2019, Figuras 4.11 e 4.12, respectivamente). Observa-se que, assim como aconteceu com o modelo RF (Figura 4.10), todos os experimentos de 2019/2020 com o TempCNN apresentaram maior identificação de áreas de Vegetação arbórea/arbustiva do que em 2017/2018 e 2018/2019, especialmente nas regiões (f), (j) e (o). O primeiro experimento de 2019/2020 usando o modelo TempCNN foi o que apresentou menor confusão entre as classes Agropecuária/Solo exposto e Vegetação arbustiva esparsa em relação à Área construída/urbana, comparado com o mesmo experimento em 2017/2018 e 2018/2019 utilizando o mesmo modelo, como evidenciado na região (a).

Figura 4.13 - Classificações de uso e cobertura da Terra para 2019/2020 via TempCNN para Petrolina-PE.



Em 2019/2020 utilizando o modelo TempCNN (Figura 4.13), observa-se que, em comparação com os anos anteriores de 2017/2018 e 2018/2019 usando o mesmo

modelo (Figuras 4.11 e 4.12, respectivamente), as áreas de mineração na região (b) foram identificadas com mais detalhes e precisão, embora classificadas erroneamente como Área construída/urbana. Semelhantemente ao que foi observado com o modelo RF (Figura 4.10), no segundo experimento de 2019/2020 com o TempCNN, identificou-se uma área urbana maior da cidade de Petrolina ( $60,23 \text{ km}^2$ ) em comparação com os mesmos experimentos e modelo em 2017/2018 e 2018/2019 (região c). Além disso, em comparação com o modelo RF, o TempCNN identificou uma área urbana maior nos dois primeiros experimentos, com diferença de  $144,87$  e  $1,25 \text{ km}^2$ , respectivamente.

De todos os experimentos que utilizaram produtos meteorológicos em 2019/2020 (Figura 4.13), os experimentos 03 (bandas, índices e T ar) e 05 (bandas, índices, T ar, UR ar e P) foram os que tiveram área urbana mais próxima da realidade do município de Petrolina (região c), com  $102,40$  e  $103,36 \text{ km}^2$ , respectivamente. Por outro lado, os experimentos 04 (bandas, índices, T ar e UR ar), 06 (bandas, índices, T ar, UR ar, P e E solo), 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) superestimaram significativamente, com  $140,67$ ;  $150,10$ ;  $186,02$ ;  $148,19$  e  $138,21 \text{ km}^2$ , respectivamente.

Entre todos os experimentos apresentados, tanto usando o RF quanto o TempCNN, o primeiro experimento com o TempCNN em 2019/2020 foi o que identificou mais áreas do distrito de Rajada (região d). Além disso, similarmente a 2017/2018 e 2018/2019 (Figuras 4.11 e 4.12, respectivamente), áreas do distrito de Rajada foram identificadas nos dois primeiros experimentos de 2019/2020. Contudo, o povoado de Izacolândia (região e) teve áreas identificadas nos experimentos 01 (bandas) a 07 (bandas, índices, T ar, UR ar, P, E solo e T solo) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo).

Outras áreas da Agrovila São Francisco (região l) são identificadas nos experimentos 01 (bandas) a 06 (bandas, índices, T ar, UR ar, P e E solo) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), embora pequenas nos experimentos 02 (índices) e 06 (Figura 4.13). Ao sul do município (região f) foram encontradas áreas de Vegetação arbustiva classificadas erroneamente como Área construída/urbana, conforme visto em outros experimentos apresentados acima. De forma semelhante ao observado usando o modelo RF (Figura 4.10), em 2019/2020 com o TempCNN, em todos os experimentos foram identificadas mais áreas de corpos hídricos do que nos anos de 2017/2018 e 2018/2019 com o mesmo modelo (Figuras 4.11 e 4.12, respec-

tivamente). No entanto, destaca-se uma superestimativa notável da classe Agropecuária/Solo exposto no último experimento de 2019/2020 com o modelo TempCNN (região p).

Observa-se que em todos os anos e experimentos com os modelos RF e TempCNN, foram identificadas algumas áreas de Vegetação arbórea erroneamente classificadas como Agricultura, especialmente aquelas localizadas próximo ao rio São Francisco ou outros reservatórios (região q) (4.13). Por outro lado, todos os casos compartilham a identificação de reservatórios de água próximos às áreas de Agricultura (região r). Além disso, diferentes regiões de Área construída próximas à área urbana da cidade de Petrolina também foram identificadas, como por exemplo o distrito de Curral Queimado e Vila Nova (região s).

O distrito de Curral Queimado foi identificado nos seis primeiros experimentos de 2017/2018 e todos os experimentos de 2018/2019 e 2019/2020 pelo modelo TempCNN (Figuras 4.11, 4.12 e 4.13, respectivamente), enquanto usando o RF, o distrito de Curral Queimado foi identificado apenas nos dois primeiros experimentos em todos os anos (Figuras 4.8, 4.9 e 4.10, respectivamente). Já o distrito de Vila Nova foi identificado usando o TempCNN em todos os anos e experimentos, e pelo RF apenas em 2017/2018 nos experimentos 01 (bandas), 02 (índices) e 06 (bandas, índices, T ar, UR ar, P e E solo), 2018/2019 nos experimentos 01, 02, 06, 08 (bandas, índices, T ar, UR ar, P, E solo, T solo e E total) e 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), e 2019/2020 nos experimentos 01, 02, 04 (bandas, índices, T ar e UR ar), 06, 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), 08 e 09. A identificação desses reservatórios e Áreas construídas demonstram a capacidade dos modelos de identificar essas classes e alvos, embora alguns tenham alcançado uma precisão maior.

Embora haja uma semelhança visual aparente nos resultados das classificações usando os modelos RF e TempCNN, são evidentes diferenças pontuais nas delimitações das classes de uso e cobertura do solo, bem como nos alvos específicos, como distritos e tipos de vegetação, resultados semelhantes foram encontrados no estudo realizado por [Pelletier et al. \(2019\)](#). Em todos os experimentos, o modelo TempCNN demonstrou uma identificação satisfatória das classes de Agricultura, Agropecuária/Solo exposto e Corpos hídricos, assim como também foi observado usando o RF. No entanto, o TempCNN apresentou maiores confusões entre as classes de Solo exposto e vegetação arbustiva esparsa com a área urbana em comparação com o RF.

Estes resultados foram diferentes quando comparados com os mapas gerados por [Alba et al. \(2022\)](#) para o mesmo município deste estudo utilizando algoritmos de aprendizado de máquina kNN, RF, SVM e ANN e imagens OLI/Landsat 8 datada de 24/07/2020. Nota-se que as classificações com os modelos tiveram áreas de Vegetação arbórea (vegetação arbórea e arbustiva), classificadas como solo exposto (afloramento rochoso e/ou solo sendo preparado para um novo ciclo de cultivo agrícola). Nos mapas gerados nesta pesquisa, observa-se menores confusões entre essas classes, além de uma melhor identificação de áreas de corpos hídricos, como passagens de água e reservatórios. Uma possível melhora na classificação dos mapas deste autor poderia ocorrer utilizando séries temporais das imagens OLI/Landsat 8 e alguns índices espectrais e variáveis meteorológicas utilizadas nessa metodologia.

Os resultados desta pesquisa apresentam uma divergência em relação aos achados de [Jardim et al. \(2022\)](#) utilizando produtos do MapBiomas, para o município de Petrolina. Por exemplo, quando se refere a classe Área urbana, que apresenta uma estimativa de 58,29 e 58,25  $km^2$ , referente aos anos 2017 e 2019 respectivamente, quando comparado aos mapas 2017/2018 e 2018/2019 gerados neste trabalho. Por outro lado, as classes de Agricultura, passagem de água do rio São Francisco e Vegetação nativa, demonstraram maior similaridade.

De modo geral, de acordo com os resultados desta pesquisa, a integração de bandas e índices espectrais e produtos meteorológicos em um cubo de dados de imagens traz ganhos significativos para a análise de uso e cobertura da terra. Esse método permite a organização e o processamento eficiente de grandes volumes de dados multidimensionais, facilitando a extração de informações complexas e detalhadas sobre a dinâmica das paisagens ao longo do tempo ([APPEL; PEBESMA, 2019](#)).

As séries temporais obtidas a partir desses cubos, são fundamentais para o monitoramento contínuo das mudanças na superfície terrestre ([FERREIRA et al., 2020](#)). Elas permitem detectar padrões de mudança, analisar tendências e identificar anomalias, como desmatamento e eventos extremos ([LAMBIN et al., 2003](#); [DASTIGERDI et al., 2024](#); [HAMUNYELA et al., 2016](#)). Tornar essas séries temporais mais densas, ao adicionar uma maior variedade de bandas e índices espectrais e dados meteorológicos, pode melhorar significativamente a resolução temporal e a qualidade dos modelos de classificação. Essa abordagem proporciona uma análise mais detalhada e precisa, capturando nuances e variações sutis ao longo do tempo.

#### 4.6 Avaliação dos mapas

Para determinar o grau de confiabilidade e precisão dos mapas classificados, foram calculadas as acurácias do produtor, usuário e global utilizando a técnica de área ponderada. No experimento 01 (Tabela 4.4), onde foram utilizadas apenas bandas espectrais, observa-se uma acurácia global de 99% para o RF nos anos de 2017/2018 e 2018/2019. Já o modelo TempCNN apresentou uma precisão menor nesses mesmos anos, com 98% e 94% de acurácia global, respectivamente. Nota-se que, tanto o modelo RF quanto o TempCNN alcançaram 100% de acurácia global em 2019/2020.

Tabela 4.4 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 01.

<b>Experimento 01</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	99	98
<b>Agropecuária/Solo exposto</b>	100	96	96	90		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	67	72	96	87		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	94	100	100	99	94
<b>Agropecuária/Solo exposto</b>	94	88	100	80		
<b>Vegetação arbórea/arbustiva</b>	100	100	98	98		
<b>Área construída/Urbana</b>	100	48	100	84		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	97	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	100	88		
<b>Corpos hídricos</b>	100	100	100	98		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

Os valores mais baixos de acurácia foram observados no produtor registrando 48% (2018/2019 usando o TempCNN), 67% (2017/2018 usando o RF) e 72% (2017/2018 usando o TempCNN) para a classe Área construída/urbana (Tabela 4.4). Por outro lado, nota-se que o modelo RF alcançou 100% de acurácia do produtor e usuário para todas as classes em 2019/2020. Além disso, observa-se que as classes de Agricultura e Vegetação arbórea/arbustiva atingiram 100% de acurácia do produtor e usuário em 2017/2018 e 2019/2020, tanto com o modelo RF quanto com o TempCNN, enquanto a classe de Corpos hídricos alcançou 100% de acurácia do produtor e usuário em 2018/2019 para ambos os modelos RF e TempCNN.

Da mesma forma, na Tabela 4.5, são apresentados os resultados das acurácias do produtor, usuário e global para o experimento 02 (bandas). Nota-se que tanto o modelo RF quanto o TempCNN alcançaram precisão máxima em 2017/2018, usando o TempCNN também mantendo esse percentual em 2019/2020. Além disso, observa-se que ambos os modelos RF e TempCNN alcançaram uma acurácia global de 98%, embora em anos diferentes (2019/2020 e 2018/2019, respectivamente). Esse foi o valor mais baixo de acurácia global encontrado no segundo experimento.

Semelhante ao primeiro experimento (Tabela 4.4), nesta os valores mais baixos foram observados na acurácia do produtor para a classe Área construída/urbana, alcançando 44% (2018/2019 usando o TempCNN), 51% (2018/2019 com o RF) e 74% (2017/2018 usando TempCNN) (Tabela 4.5). Por outro lado, nota-se que as classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, tanto com o modelo RF quanto com o TempCNN, enquanto a classe de Vegetação arbórea/arbustiva atingiu esse mesmo valor para a acurácia do produtor e usuário em 2017/2018 e 2018/2019 para ambos os modelos RF e TempCNN.

Os resultados das acurácias do produtor, usuário e global para o experimento 03 (bandas, índices e T ar) estão apresentados na Tabela 4.6. Observa-se que, com exceção de 2019/2020 usando o modelo TempCNN, os percentuais de acurácia global no experimento 03 foram iguais ou maiores do que os dos experimentos 01 (bandas) e 02 (índices), ambos com 100% de precisão. Além disso, o menor percentual de acurácia global no experimento 03 ocorreu usando o modelo TempCNN em 2019/2020, atingindo 98% de precisão.

Contrário aos dois primeiros experimentos (Tabelas 4.4 e 4.5), em que o menor valor de acurácia do produtor foi registrado na classe de Área construída/urbana em 2018/2019, no experimento 03 (bandas, índices e T ar), o menor valor de acurácia

Tabela 4.5 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 02.

<b>Experimento 02</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	100	98	98		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	80	74	96	94		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	99	98
<b>Agropecuária/Solo exposto</b>	100	100	96	94		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	51	44	94	92		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	98	100
<b>Agropecuária/Solo exposto</b>	89	100	100	98		
<b>Vegetação arbórea/arbustiva</b>	100	100	98	100		
<b>Área construída/Urbana</b>	100	83	98	96		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

do produtor foi encontrado na classe de Agropecuária/Solo exposto em 2019/2020 (Tabela 4.6). Embora os dois menores valores sejam observados na classe de Área construída/urbana, como nos dois primeiros experimentos, o percentual de 84% ocorreu na acurácia do produtor em 2018/2019 e 89% na acurácia do usuário em 2017/2018. Por outro lado, semelhante ao primeiro experimento, o modelo RF obteve 100% na acurácia do produtor em todas as classes em 2018/2019. Além disso, nota-se que, similarmente ao segundo experimento, no terceiro experimento, as classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, tanto com o modelo RF quanto com o TempCNN, enquanto a classe de Vegetação arbórea/arbustiva atingiu esse mesmo valor para a acurácia do

Tabela 4.6 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 03.

<b>Experimento 03</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	99	97	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	93	89		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	100	100	98		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	84	96	98		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	98
<b>Agropecuária/Solo exposto</b>	99	83	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	98		
<b>Área construída/Urbana</b>	100	100	96	94		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

produtor e usuário apenas nos dois primeiros anos para ambos os modelos RF e TempCNN.

Os resultados das acurácias do produtor, usuário e global para o experimento 04 (bandas, índices, T ar e UR ar) são apresentados na Tabela 4.7. Tanto o modelo RF quanto o TempCNN alcançaram uma acurácia global de 100% em todos os anos no experimento 04. Diferentemente dos experimentos anteriores, não houve valores de acurácia inferiores a 94% (acurácia do usuário usando o modelo TempCNN em 2019/2020) no experimento 04. Além disso, observa-se que os modelos RF e TempCNN alcançaram uma acurácia do produtor de 100% em todas as classes em



2017/2018. O modelo TempCNN também obteve uma acurácia do usuário de 100% em todas as classes em 2017/2018.

Tabela 4.7 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 04.

<b>Experimento 04</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	100	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	100		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	99	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	96		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	99	99	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	94		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

Ainda na Tabela 4.7, destaca-se que, em contraste com os três experimentos anteriores, no experimento 04 (bandas, índices, T ar e UR ar), as classes de Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva alcançaram uma acurácia do produtor e do usuário de 100% em todos os anos, tanto com o modelo RF quanto com o TempCNN. Além disso, observa-se que a classe de Agropecuária/Solo exposto também atingiu uma acurácia do produtor e do usuário de 100% em 2017/2018 para

ambos os modelos RF e TempCNN.

No experimento 05, conforme apresentado na Tabela 4.8, utilizando bandas, índices, temperatura do ar, umidade relativa do ar e precipitação, observa-se que ambos os modelos RF e TempCNN alcançaram uma acurácia global de 100% em 2017/2018, para o RF em 2019/2020 e para o TempCNN em 2018/2019. No entanto, o menor valor de acurácia global foi registrado no modelo TempCNN, alcançando 98% em 2019/2020, com uma diferença de apenas um ponto percentual em relação ao segundo menor valor (99% em 2018/2019), o qual foi alcançado pelo RF.

Tabela 4.8 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 05.

<b>Experimento 05</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	98	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	91		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	99	100
<b>Agropecuária/Solo exposto</b>	94	99	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	98	100		
<b>Área construída/Urbana</b>	100	100	96	94		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	98
<b>Agropecuária/Solo exposto</b>	99	87	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	98		
<b>Área construída/Urbana</b>	100	100	96	96		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

Conforme evidenciado na Tabela 4.6, semelhante ao experimento 03 (bandas, índices e T ar), o menor valor de acurácia do produtor também foi identificado na classe de Agropecuária/Solo exposto em 2019/2020 no experimento 05 (bandas, índices, T ar, UR ar e P), utilizando o modelo TempCNN (Tabela 4.8). Esse valor também representa o percentual mais baixo entre todos os resultados obtidos. Observa-se que o modelo TempCNN também registrou os dois valores mais baixos de acurácia do usuário (91% e 94% em 2017/2018 e 2018/2019, respectivamente), ambos na classe de Área construída/urbana. Semelhante ao modelo TempCNN, o RF também alcançou 94% em 2018/2019, mas na acurácia do produtor. Por outro lado, nota-se que, assim como no experimento 04 (bandas, índices, T ar e UR ar) em 2017/2018 (Tabela 4.7), o modelo RF atingiu 100% de acurácia do produtor em todas as classes no mesmo ano. Além disso, conforme observado nos experimentos 02 utilizando índices, 03 e 04, no experimento 05, as classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, tanto com o modelo RF quanto com o TempCNN. Entretanto, é a primeira vez que a classe de Vegetação arbórea/arbustiva alcançou esse mesmo valor para a acurácia do produtor e usuário, apenas em 2017/2018 para ambos os modelos RF e TempCNN.

Os resultados das acurácias do produtor, usuário e global para o experimento 06 (bandas, índices, T ar, UR ar, P e E solo) estão apresentados na Tabela 4.9. Observa-se que tanto os modelos RF quanto o TempCNN registraram os mesmos valores de acurácia global de 99% e 100% em 2017/2018 e 2019/2020, respectivamente. Além disso, o modelo RF alcançou uma acurácia global de 100% em 2018/2019, enquanto o TempCNN obteve 99% no mesmo período, representando uma diferença de apenas 1% entre os dois modelos.

No experimento 06 (bandas, índices, T ar, UR ar, P e E solo), os três menores valores de acurácia também foram registrados na classe de Área construída/urbana em 2017/2018 (85% na acurácia do produtor com o RF e 86% na acurácia do usuário com o TempCNN) e 2018/2019 (88% na acurácia do usuário usando TempCNN) (Tabela 4.9). No entanto, ao contrário dos experimentos 01 (bandas) e 02 (índices), no experimento 06, os menores valores não se limitaram à acurácia do produtor. Por outro lado, semelhante aos experimentos 03 (bandas, índices e T ar) e 04 (bandas, índices, T ar e UR ar) em 2018/2019 (Tabelas 4.6 e 4.7, respectivamente), no experimento 06, o modelo RF alcançou 100% de acurácia do produtor em todas as classes em 2018/2019. Além disso, nota-se que, assim como no experimento 04, as classes de Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva obtiveram 100% de acurácia do produtor e usuário em todos os anos, usando o modelo RF e o

Tabela 4.9 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 06.

<b>Experimento 06</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	99	99
<b>Agropecuária/Solo exposto</b>	99	95	98	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	85	100	94	86		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	100	99
<b>Agropecuária/Solo exposto</b>	100	97	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	88		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	99	99	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	94		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

TempCNN.

Na Tabela 4.10, estão apresentados os resultados das acurácias do produtor, usuário e global para o experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo). Observa-se que tanto o modelo RF quanto o TempCNN alcançaram 100% de acurácia global em 2018/2019 e 2019/2020. Além disso, o modelo RF também atingiu esse mesmo percentual em 2017/2018. Entretanto, o menor valor de acurácia global foi registrado no modelo TempCNN, com 99% em 2017/2018, uma diferença de apenas um ponto percentual em relação ao RF no mesmo ano.

Tabela 4.10 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 07.

<b>Experimento 07</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	100	99
<b>Agropecuária/Solo exposto</b>	100	99	100	98		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	84	98	94		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	100	97	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	89		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	99	98	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	94		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

Ainda na Tabela 4.10, experimento 07 (bandas, índices, T ar, UR ar, P, E solo e T solo), observa-se os menores percentuais de acurácia registrados com o modelo TempCNN na classe de Área construída/urbana, com 84% (na acurácia do produtor em 2017/2018), 89% (na acurácia do usuário em 2018/2019) e 94% (na acurácia do usuário em 2017/2018 e 2019/2020). Como ocorreu no experimento 04 (bandas, índices, T ar, UR ar) (Tabela 4.7), no experimento 07, o modelo RF alcançou uma acurácia do produtor de 100% em todas as classes durante os períodos de 2017/2018 e 2018/2019. Além disso, semelhante aos resultados dos experimentos 04 e 06 (bandas, índices, T ar, UR ar, P e E solo), anteriormente apresentados, as classes de Agricultura, Corpos hídricos e Vegetação arbórea/arbustiva alcançaram 100% de acurácia

do produtor e usuário em todos os anos, usando o modelo RF e o TempCNN.

Os mesmos resultados das acurácias são apresentados para o experimento 08 (bandas, índices, T ar, UR ar, P e E solo) são apresentados na Tabela 4.11. Assim como observado no experimento 06 (bandas, índices, T ar, UR ar, P e E solo) (Tabela 4.9), no experimento 08, tanto os modelos RF quanto o TempCNN apresentaram os mesmos valores de acurácia global, alcançando 99% e 100% em 2017/2018 e 2019/2020, respectivamente. Porém, ao contrário do experimento 06, foi o modelo TempCNN que atingiu 100% de acurácia global em 2018/2019, enquanto o RF alcançou 99%.

Tabela 4.11 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 08.

<b>Experimento 08</b>						
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>	
<b>2017/2018</b>						
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>
<b>Agricultura</b>	100	100	100	100	99	99
<b>Agropecuária/Solo exposto</b>	100	99	98	98		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	78	86	96	92		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2018/2019</b>						
<b>Agricultura</b>	100	100	100	100	99	100
<b>Agropecuária/Solo exposto</b>	93	97	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	98	100		
<b>Área construída/Urbana</b>	100	100	96	89		
<b>Corpos hídricos</b>	100	100	100	100		
<b>2019/2020</b>						
<b>Agricultura</b>	100	100	100	100	100	100
<b>Agropecuária/Solo exposto</b>	99	98	100	100		
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100		
<b>Área construída/Urbana</b>	100	100	96	93		
<b>Corpos hídricos</b>	100	100	100	100		

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.

Na Tabela 4.11, nota-se que, semelhante aos experimentos 01 (bandas), 02 (índices) e 06 (bandas, índices, T ar, UR ar, P, E solo) (Tabelas 4.4, 4.5 e 4.9, respectivamente), os menores percentuais de acurácia foram registrados na classe de Área construída/urbana em 2017/2018 (78% e 86% de acurácia do produtor nos modelos RF e TempCNN, respectivamente) e em 2018/2019 (89% de acurácia do usuário no modelo TempCNN). Por outro lado, observa-se que as classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, usando o modelo RF e TempCNN. Já a classe de Vegetação arbórea/arbustiva atingiu esse mesmo nível de acurácia do produtor e usuário somente em 2017/2018 e 2019/2020, para ambos os modelos RF e TempCNN.

Os resultados das acurácias do produtor, usuário e global para o experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo) estão apresentados na Tabela 4.12. Nota-se que apenas o modelo RF atingiu uma acurácia global de 100%, evidenciada nos anos de 2017/2018 e 2019/2020. Por outro lado, os valores mais baixos de acurácia global foram registrados pelo modelo TempCNN, obtendo 98% nos anos de 2017/2018 e 2018/2019. Além disso, destaca-se que tanto o modelo RF quanto o TempCNN apresentaram uma acurácia global de 99%, o RF em 2018/2019 e o TempCNN em 2019/2020.

Na Tabela 4.12, referente ao experimento 09 (bandas, índices, T ar, UR ar, P, E solo, T solo, E total e U solo), observa-se que, assim como nos dois primeiros experimentos, os menores valores foram encontrados na acurácia do produtor na classe de Área construída/urbana, com 74% (TempCNN em 2017/2018) e 80% (RF e TempCNN em 2017/2018 e 2019/2020, respectivamente). Por outro lado, nota-se mais uma vez que, as classes de Agricultura e Corpos hídricos alcançaram 100% de acurácia do produtor e usuário em todos os anos, tanto com o modelo RF quanto com o TempCNN. Entretanto, a classe de Vegetação arbórea/arbustiva atingiu esse mesmo valor para a acurácia do produtor e usuário apenas em 2019/2020 para ambos os modelos RF e TempCNN.

Apesar de pesquisas na literatura indicarem maiores precisões para o modelo TempCNN em comparação com o RF. De acordo com resultados semelhante encontrados em Pelletier et al. (2019) e Moncrieff (2022), esta pesquisa constatou que o modelo RF demonstrou maior precisão na maioria dos experimentos realizados, com base na acurácia do produtor, do usuário e global.

Tabela 4.12 - Valor de acurácia das imagens classificadas a partir da técnica de área ponderada para o experimento 09.

<b>Experimento 09</b>							
	<b>AP(%)</b>		<b>AU(%)</b>		<b>AG(%)</b>		
<b>2017/2018</b>							
<b>Classes</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	<b>RF</b>	<b>TC</b>	
<b>Agricultura</b>	100	100	100	100	100	98	
<b>Agropecuária/Solo exposto</b>	100	94	98	98			
<b>Vegetação arbórea/arbustiva</b>	100	100	100	98			
<b>Área construída/Urbana</b>	80	74	98	96			
<b>Corpos hídricos</b>	100	100	100	100			
<b>2018/2019</b>							
<b>Agricultura</b>	100	100	100	100	99	98	
<b>Agropecuária/Solo exposto</b>	92	90	100	100			
<b>Vegetação arbórea/arbustiva</b>	100	100	98	98			
<b>Área construída/Urbana</b>	100	100	96	88			
<b>Corpos hídricos</b>	100	100	100	100			
<b>2019/2020</b>							
<b>Agricultura</b>	100	100	100	100	100	99	
<b>Agropecuária/Solo exposto</b>	99	99	98	96			
<b>Vegetação arbórea/arbustiva</b>	100	100	100	100			
<b>Área construída/Urbana</b>	100	80	96	94			
<b>Corpos hídricos</b>	100	100	100	100			

RF = *Random forest*, TC = TempCNN, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia global.

Fonte: Próprio autor.



## 5 CONCLUSÕES

Nesta pesquisa, foi possível combinar bandas de satélites, índices espectrais e produtos meteorológicos para aprimorar a classificação dos mapas de uso e cobertura da Terra por meio de SITS e algoritmos de ML e DL. A metodologia desenvolvida para a integração de cubos de dados de observação da Terra foi composta pelas seguintes etapas: Coleta de amostras de uso e cobertura Terra, obtenção dos dados espectrais e meteorológicos, aplicação da técnica de fusão, análise das amostras, ajuste dos modelos de classificação, e por fim, classificação e avaliação dos mapas.

Foram elaborados cinco objetivos específicos, aplicando métodos de ML e DL para responder o questionamento da pesquisa. No primeiro objetivo, observou-se o ganho em utilizar séries temporais de imagens de satélite por meio de duas abordagens: (1) classificação de séries temporais de imagens de satélite e (2) classificação de uma imagem por ano de forma individual. O que evidenciou que o uso de séries temporais de imagens NDVI aprimorou a identificação das classes de uso e cobertura da terra em comparação com a abordagem de classificação baseada em uma única imagem anual, além de facilitar a detecção de mudanças na área de estudo.

No segundo objetivo, foi proposta uma metodologia para representar dados meteorológicos em forma de cubo de dados. Para isso, foi desenvolvido um código que está disponível no repositório do GitHub. Isso permitiu a integração de dados meteorológicos com dados espectrais, possibilitando sua utilização em diversas situações.

No terceiro objetivo, foi integrado os cubos de dados de observação da Terra e produtos meteorológicos na classificação de séries temporais de imagens de satélite. Essa abordagem possibilitou extrair diferentes informações espectrais e meteorológicas das classes: Agricultura, Agropecuária/Solo exposto, Área construída/Urbana, Corpos hídricos e Vegetação arbórea/arbustiva.

Nos dois últimos objetivos, foram gerados cinquenta e quatro mapas de uso e cobertura da Terra para o município de Petrolina, Pernambuco, Brasil, a partir das diferentes combinações de bandas, índices espectrais, dados meteorológicos e modelos RF e TempCNN. Por fim, avaliou-se os mapas classificados por meio das melhores práticas de [Olofsson et al. \(2014\)](#).

Neste contexto, com base nos resultados desta dissertação, conclui-se que os experimentos 03 (bandas, índices e T ar) e 04 (bandas, índices, T ar e UR ar) apresentaram os melhores resultados de classificação, em especial utilizando o modelo RF.

Aplicando a técnica de área ponderada, observou-se que as combinações de dados espectrais com meteorológicos são favoráveis para aumentar o desempenho das classificações de uso e cobertura da Terra no município de Petrolina, Pernambuco, Brasil. Em particular, as classes de Agropecuária/Solo exposto e Área construída/Urbana, historicamente desafiadoras em termos de separabilidade na literatura, demonstraram acurácias igual ou superior a 83% e 74% (na acurácia do produtor em 2019/2020 e 2017/2018, respectivamente, utilizando o modelo TempCNN), em todos os experimentos realizados. Além disso, todas as classes analisadas apresentaram acurácias satisfatórias com a inclusão de dados meteorológicos em todos os experimentos.

### 5.1 Trabalhos futuros

Devido à variedade de classes de vegetação envolvidas na dinâmica de conversão de uso da Terra no município de Petrolina, recomenda-se aplicar a metodologia desenvolvida nesta pesquisa em diferentes momentos ao longo do ano e em intervalos temporais alternativos. Isso possibilitará identificar mudanças associadas à sazonalidade climática e à dinâmica da vegetação, sendo relevante também para áreas de estudo com características semelhantes. Considerando a escassez de estudos relacionados a fusão de SITS ópticas e dados meteorológicos utilizando técnicas de ML e DL, torna-se necessário conduzir trabalhos futuros para investigar as contribuições de diferentes técnicas de fusão, como a *late fusion*, na melhoria da precisão dos mapas classificados com base nesses dados. Além disso, integrar outros dados à metodologia, por exemplo, Modelo Digital de Elevação (MDE).

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABDULHAFEDH, A. Comparison between common statistical modeling techniques used in research, including: discriminant analysis vs logistic regression, ridge regression vs lasso, and decision tree vs random forest. **Open Access Library Journal**, v. 9, n. 2, p. 1–19, 2022. 18, 19, 20
- ABRAHAM, S.; HUYNH, C.; VU, H. Classification of soils into hydrologic groups using machine learning. **Data**, v. 5, n. 1, p. 2, 2019. 17
- AIK, D. H. J.; ISMAIL, M. H.; MUHARAM, F. M. Land use/land cover changes and the relationship with land surface temperature using landsat and modis imageries in cameron highlands, Malaysia. **Land**, v. 9, n. 10, p. 372, 2020. 3
- ALBA, E.; ALEXANDRE, M. L. de S.; MARCHESAN, J.; SOUZA, L. S. B. de; BEZERRA, A. C.; SILVA, E. A. Comparação entre algoritmos de aprendizado de máquina para a identificação de floresta tropical sazonalmente seca. **Anuário do Instituto de Geociências**, v. 45, p. 1–10, 2022. 39, 40, 55, 61, 64, 68, 71, 76
- ALEXANDER, C. Normalised difference spectral indices and urban land cover as indicators of land surface temperature (lst). **International Journal of Applied Earth Observation and Geoinformation**, v. 86, p. 102013, 2020. 16
- ALI, K.; JOHNSON, B. A. Land-use and land-cover classification in semi-arid areas from medium-resolution remote-sensing imagery: a deep learning approach. **Sensors**, v. 22, n. 22, p. 8750, 2022. 44, 53
- ALQAHTANI, A.; ALI, M.; XIE, X.; JONES, M. W. Deep time-series clustering: a review. **Electronics**, v. 10, n. 23, p. 3001, 2021. 12
- ALTHOFF, T. D.; MENEZES, R. S. C.; PINTO, A. de S.; PAREYN, F. G. C.; CARVALHO, A. L. de; MARTINS, J. C. R.; CARVALHO, E. X. de; SILVA, A. S. A. da; DUTRA, E. D.; 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. 4, 8
- ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013. 26
- ALVES, J. J. A.; ARAÚJO, M. A. de; NASCIMENTO, S. S. do. Degradação da caatinga: uma investigação ecogeográfica. **Revista Caatinga**, v. 22, n. 3, p. 126–135, 2009. 7, 35
- ANJALI, K.; ROSHNI, T. Linking satellite-based forest cover change with rainfall and land surface temperature in Kerala, India. **Environment, Development and Sustainability**, p. 1–19, 2022. 16
- APPEL, M.; PEBESMA, E. On-demand processing of data cubes from satellite image collections with the gdalcubes library. **Data**, v. 4, n. 3, p. 92, 2019. 11, 76

ARORA, V.; YIN-KWEE NG, E.; SINGH, A. Machine learning and its applications. In: SEHGAL, R.; GUPTA, N.; TOMAR, A.; SHARMA, M. D.; KUMARAN, V. (Ed.). **Smart electrical and mechanical systems**. Academic Press, 2022. p. 1–37. ISBN 978-0-323-90789-7. Disponível em:

<<<https://www.sciencedirect.com/science/article/pii/B9780323907897000026>>>. 21

ATKINSON, P. M.; TATNALL, A. R. Introduction neural networks in remote sensing. **International Journal of Remote Sensing**, v. 18, n. 4, p. 699–709, 1997. 18

AUDEBERT, N.; SAUX, B. L.; LEFÈVRE, S. Segment-before-detect: vehicle detection and classification through semantic segmentation of aerial images. **Remote Sensing**, v. 9, n. 4, p. 368, 2017. 23

BARBOSA, H. A.; KUMAR, T. L.; PAREDES, F.; ELLIOTT, S.; AYUGA, J. 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. 9, 42, 65

BARBOZA, A.; OLIVEIRA, A.; ALMEIDA, I.; SANTOS, C. Impactos ambientais na região do médio curso do rio são francisco. **Geoambiente on-line**, p. 82–100, 08 2018. 9

BARTOLUCCI, F.; SCRUCICA, L. Point estimation methods with applications to item response theory models. In: PETERSON, P.; BAKER, E.; MCGAW, B. (Ed.). **International encyclopedia of education**. 3. ed. Oxford: [s.n.], 2010. p. 366–373. ISBN 978-0-08-044894-7. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780080448947013762>>>. 18

BAUMANN, P. The datacube manifesto. 2017. Disponível em: <<<http://earthserver.eu/tech/datacube-manifesto>>>. 11

BECERRA, J. A. B.; CARVALHO, S. d.; OMETTO, J.; JPH, B. Relação das sazonalidades da precipitação e da vegetação no bioma caatinga: abordagem multitemporal. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 25., 2015, João Pessoa, Brasil. **Anais. . .** São José dos Campos: SBSR, 2015. v. 25, p. 6668–6674. 45

BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016. 2, 18

BELYADI, H.; HAGHIGHAT, A. Neural networks and deep learning. In: BELYADI, H.; HAGHIGHAT, A. (Ed.). **Machine learning guide for oil and gas using python**. Gulf Professional, 2021. p. 297–347. ISBN 978-0-12-821929-4. Disponível em:

<<<https://www.sciencedirect.com/science/article/pii/B9780128219294000081>>>. 22

- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: a review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 8, p. 1798–1828, 2013. 2
- BEZERRA, J. M.; MOURA, G. B. d. A.; SILVA, B. B. d.; LOPES, P. M.; SILVA, Ê. F. d. F. Parâmetros biofísicos obtidos por sensoriamento remoto em região semiárida do estado do Rio Grande do Norte, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, p. 73–84, 2014. 4
- BOUCHEFRY, K. E.; SOUZA, R. S. de. Learning in big data: introduction to machine learning. In: ŠKODA, P.; ADAM, F. (Ed.). **Knowledge discovery in big data from astronomy and earth observation**. [S.l.: s.n.], 2020. p. 225–249. 17
- BOYAJ, A.; DASARI, H. P.; HOTEIT, I.; ASHOK, K. Increasing heavy rainfall events in south India due to changing land use and land cover. **Quarterly Journal of the Royal Meteorological Society**, v. 146, n. 732, p. 3064–3085, 2020. 16
- BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, p. 123–140, 1996. 19
- \_\_\_\_\_. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001. 2, 18
- BRITO, P.; CHAVES, M.; CARVALHO, H.; SOUZA, F.; SILVA, B.; FERREIRA, K.; SANTOS, R.; QUEIROZ, G. R. Uso de séries temporais para classificações de uso e cobertura da terra em Petrolina, Pernambuco. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 20., 2023, Florianópolis, Brazil. **Anais. . . .** São José dos Campos: SBSR, 2023. p. 2530–2533. 5, 39, 41, 44, 53, 55, 59, 61, 62, 63, 68, 71
- BRITO, P. V. d. S. Integração de índices espectrais e produtos meteorológicos na classificação de uso e cobertura da terra usando séries temporais de imagens e machine learning. In: WORKSHOP DOS CURSOS DE COMPUTAÇÃO APLICADA DO INPE, 22., 2023, São José dos Campos, Brasil. **Anais. . . .** São José dos Campos: INPE, 2023. 5
- BRITO, P. V. da S.; MORAIS, Y. C. B.; FERREIRA, H. dos S.; SILVA, J. F. da; GALVÍNCIO, J. D. Análise comparativa da umidade da vegetação de áreas de caatinga preservada, agricultura irrigada e sequeiro. **Journal of Environmental Analysis and Progress**, p. 493–498, 2017. 10
- BRITO, P. V. da S.; SANTOS, T. O. dos; MORAIS, Y. C. B. Análise da sazonalidade da vegetação de caatinga do município de petrolina (pernambuco) a partir de imagens do satélite landsat 8–oli. **Revista Brasileira de Sensoriamento Remoto**, v. 1, n. 1, 2020. 10
- CAMARA, G.; ASSIS, L. F.; RIBEIRO, G.; FERREIRA, K. R.; LLAPA, E.; VINHAS, L. Big earth observation data analytics: matching requirements to system architectures. In: ACM SIGSPATIAL INTERNATIONAL WORKSHOP

ON ANALYTICS FOR BIG GEOSPATIAL DATA, 5., 2016, Califórnia, USA. **Proceedings...** San Francisco, 2016. p. 1–6. [2](#), [12](#), [13](#)

CAO, X.; WEI, C.; HAN, J.; JIAO, L. Hyperspectral band selection using improved classification map. **IEEE Geoscience and Remote Sensing Letters**, v. 14, n. 11, p. 2147–2151, 2017. [37](#), [39](#), [40](#), [41](#), [42](#), [48](#), [54](#)

CÁRDENAS, D. A. G.; VALENCIA, J. A. R.; VELÁSQUEZ, D. F. A.; GONZALEZ, J. R. P. Dynamics of the indices ndvi and gndvi in a rice growing in its reproduction phase from multi-spectral aerial images taken by drones. In: SPRINGER. **International Conference of ICT for Adapting Agriculture to Climate Change**. [S.l.], 2018. p. 106–119. [15](#)

CARVALHO, K. da S.; CRUZ, R. C. D. da; SOUZA, I. A. de. Plant species from brazilian caatinga: a control alternative for aedes aegypti. **Journal of Asia-Pacific Entomology**, p. 102051, 2023. [7](#)

CASTELLETTI, C. H. M.; SILVA, J. M. C. d.; TABARELLI, M.; SANTOS, A. M. M. Quanto ainda resta da caatinga? uma estimativa preliminar. In: LEAL, I. R.; TABARELLI, M.; SILVA, J. M. C. (Ed.). **Ecologia e conservação da caatinga**. Recife: UFPE, 2003. p. 777–796. [8](#)

CHANG, V.; BHAVANI, V. R.; XU, A. Q.; HOSSAIN, M. An artificial intelligence model for heart disease detection using machine learning algorithms. **Healthcare Analytics**, v. 2, p. 100016, 2022. [18](#)

CHAVES, M. E.; PICOLI, M. C.; SANCHES, I. D. Recent applications of landsat 8/oli and sentinel-2/msi for land use and land cover mapping: a systematic review. **Remote Sensing**, v. 12, n. 18, p. 3062, 2020. [15](#)

CHAVES, M. E.; SOARES, A. R.; SANCHES, I. D.; FRONZA, J. G. Cbers data cubes for land use and land cover mapping in the brazilian cerrado agricultural belt. **International Journal of Remote Sensing**, v. 42, n. 21, p. 8398–8432, 2021. [13](#), [14](#), [15](#), [17](#)

CHEN, K.; ZHAI, X.; WANG, S.; LI, X.; LU, Z.; XIA, D.; LI, M. Emerging trends and research foci of deep learning in spine: bibliometric and visualization study. **Neurosurgical Review**, v. 46, n. 1, p. 81, 2023. [21](#)

CHEN, Y.; LIN, Z.; ZHAO, X.; WANG, G.; GU, Y. Deep learning-based classification of hyperspectral data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 6, p. 2094–2107, 2014. [22](#)

CHENG, G.; ZHOU, P.; HAN, J. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 54, n. 12, p. 7405–7415, 2016. [22](#)

CHENG, X.; HE, X.; QIAO, M.; LI, P.; HU, S.; CHANG, P.; TIAN, Z. Enhanced contextual representation with deep neural networks for land cover classification based on remote sensing images. **International Journal of Applied Earth Observation and Geoinformation**, v. 107, p. 102706, 2022. 22

CORTI, M.; CAVALLI, D.; CABASSI, G.; GALLINA, P. M.; BECHINI, L. Does remote and proximal optical sensing successfully estimate maize variables? a review. **European Journal of Agronomy**, v. 99, p. 37–50, 2018. 14, 15

COUTRIM, R. L.; SOUZA, L. H. identificação de árvores de potencial medicinal nativas dos biomas caatinga e cerrado na Bahia. **Geopauta**, v. 2, n. 2, p. 38–45, 2018. 8

DAHAM, A.; HAN, D.; RICO-RAMIREZ, M.; MARSH, A. Analysis of NVDI variability in response to precipitation and air temperature in different regions of iraq, using modis vegetation indices. **Environmental Earth Sciences**, v. 77, p. 1–24, 2018. 16

DAHANI, S.; RAJI, M.; HAKDAOUI, M.; LHISSOU, R. Land cover mapping using sentinel-1 time-series data and machine-learning classifiers in agricultural sub-saharan landscape. **Remote Sensing**, v. 15, n. 1, p. 65, 2022. 37, 39, 40, 41, 42, 48, 54

DASTIGERDI, M.; NADI, M.; SARJAZ, M. R.; KIAPASHA, K. Trend analysis of modis ndvi time series and its relationship to temperature and precipitation in northeastern of iran. **Environmental Monitoring and Assessment**, Springer, v. 196, n. 4, p. 346, 2024. 76

DENG, Y.; WU, C.; LI, M.; CHEN, R. RNSDI: a ratio normalized difference soil index for remote sensing of urban/suburban environments. **International Journal of Applied Earth Observation and Geoinformation**, v. 39, p. 40–48, 2015. 14

DEVKOTA, P.; DHAKAL, S.; SHRESTHA, S.; SHRESTHA, U. B. Land use land cover changes in the major cities of nepal from 1990 to 2020. **Environmental and Sustainability Indicators**, p. 100227, 2023. 15

DIEK, S.; FORNALLAZ, F.; SCHAEPMAN, M. E.; JONG, R. D. Barest pixel composite for agricultural areas using landsat time series. **Remote Sensing**, v. 9, n. 12, p. 1245, 2017. 15

DIGRA, M.; DHIR, R.; SHARMA, N. Land use land cover classification of remote sensing images based on the deep learning approaches: a statistical analysis and review. **Arabian Journal of Geosciences**, v. 15, n. 10, p. 1003, 2022. 22

DISTRITO DE IRRIGAÇÃO SENADOR NILO COELHO (DINC). **DINC**. 2022. Disponível em: <<<https://www.dinc.org.br/>>>. Acesso em: 28 mar. 2024. 10

EITEN, G. The cerrado vegetation of brazil. **The Botanical Review**, Springer, v. 38, p. 201–341, 1972. 7

- ESTEVA, A.; ROBICQUET, A.; RAMSUNDAR, B.; KULESHOV, V.; DEPRISTO, M.; CHOU, K.; CUI, C.; CORRADO, G.; THRUN, S.; DEAN, J. A guide to deep learning in healthcare. **Nature Medicine**, v. 25, n. 1, p. 24–29, 2019. 21
- FADL, M. E.; ABDELRAHMAN, M. A.; EL-DESOKY, A. I.; SAYED, Y. A. Assessing soil productivity potential in arid region using remote sensing vegetation indices. **Journal of Arid Environments**, v. 222, p. 105166, 2024. 15
- FARIDATUL, M. I.; WU, B. Multi-temporal urban land cover mapping using spectral indices. **International Journal of Mechanical and Mechatronics Engineering**, v. 13, n. 2, p. 147–153, 2019. 28
- FERREIRA, K. R. et al. Earth observation data cubes for Brazil: requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, p. 4033, 2020. 1, 2, 12, 27, 76
- FILHO, W. L. F. C. et al. Rainfall variability in the brazilian northeast biomes and their interactions with meteorological systems and enso via chelsa product. **Big Earth Data**, v. 3, n. 4, p. 315–337, 2019. 9
- FREITAS, L. O. de; ARANHA, S. de A. Produção de livro digital: fragmentos da caatinga macauense. **Revista Brasileira de Educação Ambiental (RevBEA)**, v. 19, n. 2, p. 181–197, 2024. 8
- FREUND, Y. et al. Experiments with a new boosting algorithm. In: CITESSEER. **icml**. [S.l.], 1996. v. 96, p. 148–156. 19
- FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. **Remote Sensing of Environment**, v. 61, n. 3, p. 399–409, 1997. 2, 18
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). **The Annals of Statistics**, v. 28, n. 2, p. 337–407, 2000. 19
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of Statistics**, p. 1189–1232, 2001. 19
- FRIMPONG, B. F.; MOLKENTHIN, F. Tracking urban expansion using random forests for the classification of landsat imagery (1986–2015) and predicting urban/built-up areas for 2025: a study of the kumasi metropolis, ghana. **Land**, v. 10, n. 1, p. 44, 2021. 64
- FU, P.; WENG, Q. A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with landsat imagery. **Remote Sensing of Environment**, v. 175, p. 205–214, 2016. 16
- GANJIRAD, M.; BAGHERI, H. Google earth engine-based mapping of land use and land cover for weather forecast models using landsat 8 imagery. **Ecological Informatics**, v. 80, p. 102498, 2024. 3, 15, 16, 62



GEOLÓGICO, S.; BRASIL-CPRM, D. Projeto cadastro de fontes de abastecimento por água subterrânea. **Diagnóstico do município de Petrolina, estado de Pernambuco/Organizado [por] João de Castro Mascarenhas, Breno Augusto Beltrão, Luiz Carlos de Souza Junior, Manoel Júlio da Trindade G. Galvão, Simone Neri Pereira, Jorge Luiz Fortunato de Miranda.** Recife: CPRM/PRODEEM, 2005. 9

GHANBARI, R.; HEIDARIMOZAFFAR, M.; SOLTANI, A.; AREFI, H. Land surface temperature analysis in densely populated zones from the perspective of spectral indices and urban morphology. **International Journal of Environmental Science and Technology**, p. 1–20, 2023. 16

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from eos-modis. **Remote Sensing of Environment**, v. 58, n. 3, p. 289–298, 1996. 15, 28

GIULIANI, G.; CHATENOUX, B.; BONO, A. D.; RODILA, D.; RICHARD, J.-P.; ALLENBACH, K.; DAO, H.; PEDUZZI, P. Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). **Big Earth Data**, v. 1, n. 1-2, p. 100–117, 2017. 1, 2, 11

GOMES, V. C.; QUEIROZ, G. R.; FERREIRA, K. R. An overview of platforms for big earth observation data management and analysis. **Remote Sensing**, v. 12, n. 8, p. 1253, 2020. 12

GÓMEZ, C.; WHITE, J. C.; WULDER, M. A. Optical remotely sensed time series data for land cover classification: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 116, p. 55–72, 2016. 2

GOVAERTS, Y. M.; VERSTRAETE, M. M.; PINTY, B.; GOBRON, N. Designing optimal spectral indices: a feasibility and proof of concept study. **International Journal of Remote Sensing**, v. 20, n. 9, p. 1853–1873, 1999. 14

GUO, Q.; ZHANG, J.; GUO, S.; YE, Z.; DENG, H.; HOU, X.; ZHANG, H. Urban tree classification based on object-oriented approach and random forest algorithm using unmanned aerial vehicle (uav) multispectral imagery. **Remote Sensing**, v. 14, n. 16, p. 3885, 2022. 18

HAMIDA, A. B.; BENOIT, A.; LAMBERT, P.; AMAR, C. B. 3-d deep learning approach for remote sensing image classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 56, n. 8, p. 4420–4434, 2018. 24

HAMUNYELA, E.; VERBESSELT, J.; BRUIN, S. D.; HEROLD, M. Monitoring deforestation at sub-annual scales as extreme events in landsat data cubes. **Remote sensing**, MDPI, v. 8, n. 8, p. 651, 2016. 76

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H.; FRIEDMAN, J. H. **The elements of statistical learning: data mining, inference, and prediction.** [S.l.]: Springer, 2009. 55

- HONG, T.; WANG, Z.; LUO, X.; ZHANG, W. State-of-the-art on research and applications of machine learning in the building life cycle. **Energy and Buildings**, v. 212, p. 109831, 2020. 17
- HU, F.; XIA, G.-S.; HU, J.; ZHANG, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. **Remote Sensing**, v. 7, n. 11, p. 14680–14707, 2015. 23
- HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195–213, 2002. 15
- HUETE, A.; JUSTICE, C.; LEEUWEN, W. V. Modis vegetation index (mod13). **Algorithm Theoretical Basis Document**, v. 3, n. 213, p. 295–309, 1999. 16, 28
- HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295–309, 1988. 15, 28
- HUNG, C.-L. Deep learning in biomedical informatics. In: ZHENG, Y.; WU, Z. (Ed.). **Intelligent nanotechnology**. [s.n.], 2023, (Materials Today). p. 307–329. ISBN 978-0-323-85796-3. Disponível em:  
<<<https://www.sciencedirect.com/science/article/pii/B9780323857963000111>>>. 22, 23
- HUSSAIN, S.; KARUPPANNAN, S. Land use/land cover changes and their impact on land surface temperature using remote sensing technique in district khanewal, Punjab Pakistan. **Geology, Ecology, and Landscapes**, v. 7, n. 1, p. 46–58, 2023. 16
- IENCO, D.; INTERDONATO, R.; GAETANO, R.; MINH, D. H. T. Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 158, p. 11–22, 2019. 13, 17
- IMMITZER, M.; VUOLO, F.; ATZBERGER, C. First experience with sentinel-2 data for crop and tree species classifications in central Europe. **Remote Sensing**, v. 8, n. 3, p. 166, 2016. 13
- INGLADA, J. et al. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. **Remote Sensing**, v. 7, n. 9, p. 12356–12379, 2015. 2, 18
- INGLADA, J.; VINCENT, A.; ARIAS, M.; TARDY, B.; MORIN, D.; RODES, I. Operational high resolution land cover map production at the country scale using satellite image time series. **Remote Sensing**, v. 9, n. 1, p. 95, 2017. 13
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Petrolina panorama**. 2019. Disponível em:  
<<<https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>>>. Acesso em: 02 mar. 2024. 62

\_\_\_\_\_. \_\_\_\_\_. 2022. Disponível em:  
<<<https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>>>. Acesso em: 02  
mar. 2024. 25

\_\_\_\_\_. **Petrolina histórico**. 2024. Disponível em:  
<<<https://cidades.ibge.gov.br/brasil/pe/petrolina/historico>>>. Acesso em: 17  
abr. 2024. 9

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **TerraBrasilis - Monitoramento de Queimadas**. 2023. Disponível em:  
<<<https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>>>. Acesso em: 15  
jan. 2023. 29

JARDIM, A. M. d. R. F. et al. Using remote sensing to quantify the joint effects of climate and land use/land cover changes on the caatinga biome of northeast brazilian. **Remote Sensing**, v. 14, n. 8, p. 1911, 2022. 9, 25, 26, 76

KADAM, S.; GORANTIWAR, S.; DAHIWALKAR, S.; SHINDE, M. et al. Crop evapotranspiration and normalized difference vegetation index relationship for wheat crop. **Agricultural Research Journal**, v. 56, n. 2, p. 336–339, 2019. 16

KHAN, S. H.; HAYAT, M.; BENNAMOUN, M.; SOHEL, F. A.; TOGNERI, R. Cost-sensitive learning of deep feature representations from imbalanced data. **IEEE transactions on neural networks and learning systems**, IEEE, v. 29, n. 8, p. 3573–3587, 2017. 27

KHATAMI, R.; MOUNTRAKIS, G.; STEHMAN, S. V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. **Remote Sensing of Environment**, v. 177, p. 89–100, 2016. 18

KILLOUGH, B. The impact of analysis ready data in the africa regional data cube. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2019. **Proceedings...** [S.l.]: IEEE, 2019. p. 5646–5649. 2, 12

KISNER, H.; DING, Y.; THOMAS, U. Capacitive material detection with machine learning for robotic grasping applications. In: LI, Q.; LUO, S.; CHEN, Z.; YANG, C.; ZHANG, J. (Ed.). **Tactile sensing, skill learning, and robotic dexterous manipulation**. Academic Press, 2022. p. 59–79. ISBN 978-0-323-90445-2. Disponível em:  
<<<https://www.sciencedirect.com/science/article/pii/B9780323904452000118>>>. 22, 23, 24

KOBAYASHI, Y.; HIGA, M.; HIGASHIYAMA, K.; NAKAMURA, F. Drivers of land-use changes in societies with decreasing populations: a comparison of the factors affecting farmland abandonment in a food production area in Japan. **PLoS One**, v. 15, n. 7, p. e0235846, 2020. 1

KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, p. 1464–1480, 1990. 31

LAMBIN, E. F.; GEIST, H. J.; LEPERS, E. Dynamics of land-use and land-cover change in tropical regions. **Annual review of environment and resources**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 28, n. 1, p. 205–241, 2003. 76

LANGE, H.; SIPPEL, S. Machine learning applications in hydrology. **Forest-Water Interactions**, p. 233–257, 2020. 17

LARY, D. J. Machine learning for environmental sensing. In: SHRODER, J. J. F. (Ed.). **Treatise on geomorphology**. 2. ed. Oxford: Academic Press, 2022. p. 260–291. ISBN 978-0-12-818235-2. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780128182345001498>>>. 17

LARY, D. J.; ALAVI, A. H.; GANDOMI, A. H.; WALKER, A. L. Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, v. 7, n. 1, p. 3–10, 2016. 17

LEAL, I. R.; SILVA, J. d.; TABARELLI, M.; JR, T. E. L. Mudando o curso da conservação da biodiversidade na caatinga do nordeste do Brasil. **Megadiversidade**, v. 1, n. 1, p. 139–146, 2005. 7

LEITE, C.; LOPES, P.; NASCIMENTO, C.; MOURA, G.; SILVA, D.; SILVA, J. Núcleo de semiaridez associado a ocupação e umidade do solo em Pernambuco, Brasil. **Journal of Hyperspectral Remote Sensing**, v. 10, n. 1, p. 20–33, 2020. ISSN 2237-2202. Disponível em: <<<https://periodicos.ufpe.br/revistas/jhrs/article/view/244195>>>. 8

LEWIS, A. et al. The australian geoscience data cube—foundations and lessons learned. **Remote Sensing of Environment**, v. 202, p. 276–292, 2017. 1, 2, 11

LI, G.; ZHANG, F.; JING, Y.; LIU, Y.; SUN, G. Response of evapotranspiration to changes in land use and land cover and climate in china during 2001–2013. **Science of the Total Environment**, v. 596, p. 256–265, 2017. 16, 24

LI, J.; ZHENG, X.; ZHANG, C.; CHEN, Y. Impact of land-use and land-cover change on meteorology in the beijing–tianjin–hebei region from 1990 to 2010. **Sustainability**, v. 10, n. 1, p. 176, 2018. 3

LI, Y.; HUANG, Q.; PEI, X.; CHEN, Y.; JIAO, L.; SHANG, R. Cross-layer attention network for small object detection in remote sensing imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 2148–2161, 2020. 22

LIANG, H.; LI, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. **Remote Sensing**, v. 8, n. 2, p. 99, 2016. 24

LIAW, A. et al. Classification and regression by randomforest. **R News**, v. 2, n. 3, p. 18–22, 2002. 20

- LÖW, F.; BIRADAR, C.; DUBOVYK, O.; FLIEMANN, E.; AKRAMKHANOV, A.; VALLEJO, A. N.; WALDNER, F. Regional-scale monitoring of cropland intensity and productivity with multi-source satellite image time series. **GIScience & Remote Sensing**, v. 55, n. 4, p. 539–567, 2018. 13
- LU, M.; APPEL, M.; PEBESMA, E. Multidimensional arrays for analysing geoscientific data. **ISPRS International Journal of Geo-Information**, v. 7, n. 8, p. 313, 2018. 11
- LYNNES, C.; HUANG, T. Future of big earth data analytics. **Big Data Analytics in Earth, Atmospheric, and Ocean Sciences**, p. 293–305, 2022. 13
- MA, L.; LIU, Y.; ZHANG, X.; YE, Y.; YIN, G.; JOHNSON, B. A. Deep learning in remote sensing applications: a meta-analysis and review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 152, p. 166–177, 2019. 21, 22
- MA, S.; ZHOU, Z.; ZHANG, Y.; AN, Y.; YANG, G. Identification of forest disturbance and estimation of forest age in subtropical mountainous areas based on landsat time series data. **Earth Science Informatics**, v. 15, n. 1, p. 321–334, 2022. 13
- MACÊDO, M. S.; MENEZES, B. S.; LEDRU, M. P.; MAS, J.-F.; SILVA, F. K. G.; CARVALHO, C. E.; COSTA, R. C.; ZANDAVALLI, R. B.; SOARES, A. A.; ARAÚJO, F. S. Everything's not lost: Caatinga areas under chronic disturbances still have well-preserved plant communities. **Journal of Arid Environments**, v. 222, p. 105164, 2024. 8
- MAGGIORI, E.; TARABALKA, Y.; CHARPIAT, G.; ALLIEZ, P. Convolutional neural networks for large-scale remote-sensing image classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 2, p. 645–657, 2016. 23
- MALEDE, D. A.; ALAMIREW, T.; KOSGIE, J. R.; ANDUALEM, T. G. Analysis of land use/land cover change trends over birr river watershed, abbay basin, ethiopia. **Environmental and Sustainability Indicators**, v. 17, p. 100222, 2023. 1
- MARENCO, J. A.; TORRES, R. R.; ALVES, L. M. Drought in northeast brazil—past, present, and future. **Theoretical and Applied Climatology**, v. 129, p. 1189–1200, 2017. 26
- MASI, G.; COZZOLINO, D.; VERDOLIVA, L.; SCARPA, G. Pansharpening by convolutional neural networks. **Remote Sensing**, v. 8, n. 7, p. 594, 2016. 23
- MASINI, R. P.; MEDEIROS, M. C.; MENDES, E. F. Machine learning advances for time series forecasting. **Journal of Economic Surveys**, v. 37, n. 1, p. 76–111, 2023. 17
- MASROOR, M.; AVTAR, R.; SAJJAD, H.; CHOUDHARI, P.; KULIMUSHI, L. C.; KHEDHER, K. M.; KOMOLAFE, A. A.; YUNUS, A. P.; SAHU, N.

Assessing the influence of land use/land cover alteration on climate variability: an analysis in the aurangabad district of Maharashtra state, India. **Sustainability**, v. 14, n. 2, p. 642, 2022. 16

MATOSAK, B. M.; FONSECA, L. M. G.; TAQUARY, E. C.; MARETTO, R. V.; BENDINI, H. d. N.; ADAMI, M. Mapping deforestation in cerrado based on hybrid deep learning architecture and medium spatial resolution satellite time series. **Remote Sensing**, v. 14, n. 1, p. 209, 2022. 13

MATTON, N.; CANTO, G. S.; WALDNER, F.; VALERO, S.; MORIN, D.; INGLADA, J.; ARIAS, M.; BONTEMPS, S.; KOETZ, B.; DEFOURNY, P. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. **Remote Sensing**, v. 7, n. 10, p. 13208–13232, 2015. 13

MAURO, N. D.; VERGARI, A.; BASILE, T. M. A.; VENTOLA, F. G.; ESPOSITO, F. End-to-end learning of deep spatio-temporal representations for satellite image time series classification. In: **DC@ PKDD/ECML**. [S.l.: s.n.], 2017. 24

MIRANDA, P. **Introdução à Meteorologia**. Lisboa: [s.n.], 2013. 780 p. IDL, Faculdade de Ciências, Universidade de Lisboa. 16

MO, Y.; ZHONG, R.; CAO, S. Orbita hyperspectral satellite image for land cover classification using random forest classifier. **Journal of Applied Remote Sensing**, v. 15, n. 1, p. 014519–014519, 2021. 18

MOHINDRU, G.; MONDAL, K.; DUTTA, P.; BANKA, H. Mining challenges in large-scale iot data framework—a machine learning perspective. **Advanced Data Mining Tools and Methods for Social Computing**, p. 239–259, 2022. 20

MONCRIEFF, G. R. Continuous land cover change detection in a critically endangered shrubland ecosystem using neural networks. **Remote Sensing**, v. 14, n. 12, p. 2766, 2022. 87

MONTEIRO, E. R. et al. Genetic diversity and structure of populations in *pilosocereus gounellei* (fac weber ex k. schum.)(cactaceae) in the caatinga biome as revealed by heterologous microsatellite primers. **Biochemical Systematics and Ecology**, v. 58, p. 7–12, 2015. 8

MONTERO, D.; AYBAR, C.; MAHECHA, M. D.; MARTINUZZI, F.; SÖCHTING, M.; WIENEKE, S. A standardized catalogue of spectral indices to advance the use of remote sensing in earth system research. **Scientific Data**, v. 10, n. 1, p. 197, 2023. 14, 15

MORAIS, L. F. d.; CAVALCANTE, A. C. R.; AQUINO, D. d. N.; NOGUEIRA, F. H. M.; CÂNDIDO, M. J. D. Spectral responses in rangelands and land cover change by livestock in regions of the caatinga biome, brazil. **Scientific Reports**, v. 11, n. 1, p. 1–10, 2021. 8

- MORENO-MARTÍNEZ, Á.; IZQUIERDO-VERDIGUIER, E.; MANETA, M. P.; CAMPS-VALLS, G.; ROBINSON, N.; MUÑOZ-MARÍ, J.; SEDANO, F.; CLINTON, N.; RUNNING, S. W. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. **Remote Sensing of Environment**, v. 247, p. 111901, 2020. 66, 68
- MOSAFFA, H.; SADEGHI, M.; MALLAKPOUR, I.; JAHROMI, M. N.; POURGHASEMI, H. R. Application of machine learning algorithms in hydrology. In: POURGHASEMI, H. R. (Ed.). **Computers in earth and environmental sciences**. [S.l.: s.n.], 2022. p. 585–591. 17
- MUÑOZ-SABATER, J. Era5-land hourly data from 1950 to present. **Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[Data Set]**, 2019. Disponível em: <<<https://doi.org/10.24381/cds.e2161bac>>>. Acesso em: 15 jan. 2023. 30
- MURO, J.; STRAUCH, A.; HEINEMANN, S.; STEINBACH, S.; THONFELD, F.; WASKE, B.; DIEKKRÜGER, B. Land surface temperature trends as indicator of land use changes in wetlands. **International Journal of Applied Earth Observation and Geoinformation**, v. 70, p. 62–71, 2018. 16
- NATIVI, S.; MAZZETTI, P.; CRAGLIA, M. A view-based model of data-cube to support big earth data systems interoperability. **Big Earth Data**, v. 1, n. 1-2, p. 75–99, 2017. 1
- NIJHAWAN, R.; JOSHI, D.; NARANG, N.; MITTAL, A.; MITTAL, A. A futuristic deep learning framework approach for land use-land cover classification using remote sensing imagery. In: MANDAL, J. K.; BHATTACHARYYA, D.; AULUCK, N. (Ed.). **Advanced computing and communication technologies: proceedings of the 11th ICACCT 2018**. [S.l.]: Springer, 2019. p. 87–96. 22
- OFORI-AMPOFO, S.; PELLETIER, C.; LANG, S. Crop type mapping from optical and radar time series using attention-based deep learning. **Remote Sensing**, v. 13, n. 22, p. 4668, 2021. 30
- OLDELAND, J.; DORIGO, W.; LIECKFELD, L.; LUCIEER, A.; JÜRGENS, N. Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery. **Remote Sensing of Environment**, v. 114, n. 6, p. 1155–1166, 2010. 40
- OLIVEIRA, M. L. de; SANTOS, C. A. D.; OLIVEIRA, G. de; PEREZ-MARIN, A. M.; SANTOS, C. A. Effects of human-induced land degradation on water and carbon fluxes in two different brazilian dryland soil covers. **Science of the Total Environment**, v. 792, p. 148458, 2021. 8, 42
- OLIVEIRA, P. d.; SILVA, C. Santos e; LIMA, K. C. Climatology and trend analysis of extreme precipitation in subregions of northeast brazil. **Theoretical and Applied Climatology**, v. 130, p. 77–90, 2017. 26

OLOFSSON, P.; FOODY, G. M.; HEROLD, M.; STEHMAN, S. V.; WOODCOCK, C. E.; WULDER, M. A. Good practices for estimating area and assessing accuracy of land change. **Remote Sensing of Environment**, v. 148, p. 42–57, 2014. 4, 33, 35, 89

PELLETIER, C.; JI, Z.; HAGOLLE, O.; MORSE-MCNABB, E.; SHEFFIELD, K.; WEBB, G. I.; PETITJEAN, F. Using sentinel-2 image time series to map the state of Victoria, Australia. In: IEEE INTERNATIONAL WORKSHOP ON THE ANALYSIS OF MULTITEMPORAL REMOTE SENSING IMAGES, 2019. **Proceedings...** [S.l.]: IEEE, 2019. p. 1–4. 57, 59

PELLETIER, C.; VALERO, S.; INGLADA, J.; CHAMPION, N.; DEDIEU, G. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. **Remote Sensing of Environment**, v. 187, p. 156–168, 2016. 2, 14, 18, 20

PELLETIER, C.; WEBB, G. I.; PETITJEAN, F. Temporal convolutional neural network for the classification of satellite image time series. **Remote Sensing**, v. 11, n. 5, p. 523, 2019. 1, 2, 3, 13, 17, 22, 23, 24, 57, 59, 68, 75, 87

PICOLI, M. C. A. et al. Big earth observation time series analysis for monitoring brazilian agriculture. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 145, p. 328–339, 2018. 1, 17

PIYOOSH, A. K.; GHOSH, S. K. Analysis of land use land cover change using a new and existing spectral indices and its impact on normalized land surface temperature. **Geocarto International**, v. 37, n. 8, p. 2137–2159, 2022. 38

POSTADJIAN, T.; BRIS, A. L.; SAHBI, H.; MALLETT, C. Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images. **ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 4, 2017. 23

POURGHASEMI, H. R.; POUYAN, S.; BORDBAR, M.; GOLKAR, F.; CLAGUE, J. J. Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination. **Natural Hazards**, p. 1–20, 2023. 17

PRASAD, A.; GANASALA, P.; HERNÁNDEZ-GUZMÁN, R.; FATHIAN, F. Remote sensing satellite data and spectral indices: an initial evaluation for the sustainable development of an urban area. **Sustainable Water Resources Management**, v. 8, n. 1, p. 19, 2022. 14

QI, J.; CHEHBOUNI, A.; HUETE, A. R.; KERR, Y. H.; SOROOSHIAN, S. A modified soil adjusted vegetation index. **Remote Sensing of Environment**, v. 48, n. 2, p. 119–126, 1994. 16, 28

QIU, J.; YANG, J.; WANG, Y.; SU, H. A comparison of ndvi and evi in the distrad model for thermal sub-pixel mapping in densely vegetated areas: A case



study in southern china. **International Journal of Remote Sensing**, Taylor & Francis, v. 39, n. 8, p. 2105–2118, 2018. 15

RASUL, A.; BALZTER, H.; IBRAHIM, G. R. F.; HAMEED, H. M.; WHEELER, J.; ADAMU, B.; IBRAHIM, S.; NAJMADDIN, P. M. Applying built-up and bare-soil indices from landsat 8 to cities in dry climates. **Land**, v. 7, n. 3, p. 81, 2018. 16, 28

RÊGO, L. G. C. do; OLIVEIRA, B. S. N.; GASPAR, L. P.; BRANCO, J. A. C.; MACÊDO, J. A. F. de. **Uso de meta-learning em tarefas de aprendizado profundo**. [S.l.]: Sociedade Brasileira de Computação, 2022. 36

ROUSE, J. W. et al. **Monitoring vegetation systems in the Great Plains with ERTS**. 1. ed. Washington: NASA, 1974. 309 p. 28

ROY, P. S.; RAMACHANDRAN, R. M.; PAUL, O.; THAKUR, P. K.; RAVAN, S.; BEHERA, M. D.; SARANGI, C.; KANAWADE, V. P. Anthropogenic land use and land cover changes—a review on its environmental consequences and climate change. **Journal of the Indian Society of Remote Sensing**, v. 50, n. 8, p. 1615–1640, 2022. 1

SAGI, O.; ROKACH, L. Ensemble learning: a survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 8, n. 4, p. e1249, 2018. 19

SAMSUDIN, S. H.; SHAFRI, H. Z.; HAMEDIANFAR, A. Development of spectral indices for roofing material condition status detection using field spectroscopy and worldview-3 data. **Journal of Applied Remote Sensing**, v. 10, n. 2, p. 025021–025021, 2016. 15

SANTOS, C. A. G.; SILVA, R. M. da; SILVA, A. M.; NETO, R. M. B. Estimation of evapotranspiration for different land covers in a brazilian semi-arid region: a case study of the brígida river basin, Brazil. **Journal of South American Earth Sciences**, v. 74, p. 54–66, 2017. 9

SANTOS, F. et al. Albedo seasonal behavior and vegetation indices in the upper basin Paraíba river. **Revista Brasileira de Geografia Física**. v07, n. 5, p. 1015–1027, 2014. 15, 27, 63

SANTOS, L.; FERREIRA, K. R.; PICOLI, M.; CAMARA, G. Self-organizing maps in earth observation data cubes analysis. In: VELLIDO, A.; GIBERT, K.; ANGULO, C.; GUERRERO, J. D. M. (Ed.). **Advances in self-organizing maps, learning vector quantization, clustering and data visualization: proceedings of the 13th International Workshop, WSOM+ 2019, Barcelona, Spain, June 26-28, 2019** 13. [S.l.]: Springer, 2020. p. 70–79. 38

SANTOS, L. et al. Quality control and class noise reduction of satellite image time series. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 177, p. 75–88, 2021. 31, 35, 36, 38, 44, 53

- SANTOS, N.; PEREIRA, L. Cálculo do escoamento superficial e resposta hidrológica de uma bacia de drenagem. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA, 10., 2016 . . . , 2016. 9
- SCHULTZ, P.; HALPERT, M. Global analysis of the relationships among a vegetation index, precipitation and land surface temperature. **Remote Sensing**, v. 16, n. 15, p. 2755–2777, 1995. 16
- SHRESTHA, M.; MITRA, C.; RAHMAN, M.; MARZEN, L. Mapping and predicting land cover changes of small and medium size cities in Alabama using machine learning techniques. **Remote Sensing**, v. 15, n. 1, p. 106, 2023. 17
- SILVA, D. A. d. O.; LOPES, P. M. O.; MOURA, G. B. d. A.; SILVA, Ê. F. d. F.; SILVA, J. L. B. d.; BEZERRA, A. C. Evolução espaço-temporal do risco de degradação da cobertura vegetal de Petrolina-PE. **Revista Brasileira de Meteorologia**, v. 34, p. 89–99, 2019. 8, 25, 26, 62, 63
- SILVA, D. A. O. et al. Monitoring of degradation in the caatinga biome in the semi-arid northeastern region, Brazil. **Revista Brasileira de Geografia Física**, v. 13, n. 05, p. 2293–2303, 2020. 8, 9
- SILVA, É. E. d. M.; PAIXÃO, V. H. F.; TORQUATO, J. L.; LUNARDI, D. G.; LUNARDI, V. de O. Fruiting phenology and consumption of zoochoric fruits by wild vertebrates in a seasonally dry tropical forest in the brazilian caatinga. **Acta Oecologica**, v. 105, p. 103553, 2020. 9
- SILVA, F. H. B. B. da; SILVA, M. S. L. da; CAVALCANTI, A. C.; CUNHA, T. J. F. **Principais solos do semi-árido do Nordeste do Brasil:"Dia de Campo"**. [S.l.]: In: CURSO [SOBRE] MANEJO E CONSERVAÇÃO DO SOLO E DA ÁGUA, 2., 2006, Juazeiro . . . , 2006. 10
- SILVA, F. P. da; LOPES, R. J. de C.; SANTOS, A. M. dos; OLIVEIRA FILHO, J. C. A.; CAVALCANTE, A. A. Panorama da mineração no semiárido brasileiro. 2023. 10, 62
- SILVA, G. J. F.; MACÊDO, M. L. A.; SILVA, R. M. Caracterização espaço-temporal das tipologias vegetais em ambiente de caatinga degradada: quantificação por meio de variáveis biofísicas. **Revista Ateliê Geográfico**, v. 13, n. 2, 2019. 40, 48, 54
- SILVA, J. L. B. et al. Spatial-temporal dynamics of the caatinga vegetation cover by remote sensing in municipality of the brazilian semi-arid. **Revista Brasileira de Ciências Agrárias**, v. 14, n. 4, p. 1–10, 2019. 41
- \_\_\_\_\_. Environmental degradation of vegetation cover and water bodies in the semiarid region of the brazilian northeast via cloud geoprocessing techniques applied to orbital data. **Journal of South American Earth Sciences**, v. 121, p. 104164, 2023. 41

SILVA, J. L. B. da; REFATI, D. C.; LIMA, R. da C. C.; CARVALHO, A. A. de; FERREIRA, M. B.; PANDORFI, H.; SILVA, M. V. da. Techniques of geoprocessing via cloud in google earth engine applied to vegetation cover and land use and occupation in the brazilian semiarid region. **Geographies**, v. 2, n. 4, p. 593–608, 2022. 41

SILVA, J. M. C.; LACHER, T. E. Caatinga–South america. In: GOLDSTEIN, M. I.; DELLASALA, D. A. (Ed.). **Encyclopedia of the World's Biomes**. Oxford: [s.n.], 2020. p. 554–561. ISBN 978-0-12-816097-8. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780124095489119840>>>. 7

SILVA, J. M. C. d.; BARBOSA, L. C. F.; LEAL, I. R.; TABARELLI, M. The caatinga: understanding the challenges. In: \_\_\_\_\_. **Caatinga: The largest tropical dry forest region in South America**. Cham: Springer International Publishing, 2017. p. 3–19. ISBN 978-3-319-68339-3. Disponível em: <<[https://doi.org/10.1007/978-3-319-68339-3\\_1](https://doi.org/10.1007/978-3-319-68339-3_1)>>. 7

SILVA, L. G. da; GALVÍNCIO, J. D. Análise comparativa da variação nos índices NDVI e SAVI no sítio PELD–22, em Petrolina–PE, na primeira década do século XXI. **Revista Brasileira de Geografia Física**, v. 5, n. 6, p. 1446–1456, 2013. 41

SIMOES, R.; CAMARA, G.; QUEIROZ, G.; SOUZA, F.; ANDRADE, P. R.; SANTOS, L.; CARVALHO, A.; FERREIRA, K. Satellite image time series analysis for big earth observation data. **Remote Sensing**, v. 13, n. 13, p. 2428, 2021. 1, 2, 11, 12, 13, 33, 55

SIMOES, R.; PICOLI, M. C.; CAMARA, G.; MACIEL, A.; SANTOS, L.; ANDRADE, P. R.; SANCHEZ, A.; FERREIRA, K.; CARVALHO, A. Land use and cover maps for Mato Grosso state in Brazil from 2001 to 2017. **Scientific Data**, v. 7, n. 1, p. 34, 2020. 17

SIMS, D. A.; GAMON, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. **Remote Sensing of Environment**, v. 81, n. 2-3, p. 337–354, 2002. 40

SOILLE, P.; BURGER, A.; MARCHI, D. D.; KEMPENEERS, P.; RODRIGUEZ, D.; SYRRIS, V.; VASILEV, V. A versatile data-intensive computing platform for information retrieval from big geospatial data. **Future Generation Computer Systems**, v. 81, p. 30–40, 2018. 1

SOUSA, M. d. C. de; VELOSO, G. V.; GOMES, L. C.; FERNANDES-FILHO, E. I.; OLIVEIRA, T. S. de. Spatio-temporal dynamics of land use changes of an intense anthropized basin in the brazilian semi-arid region. **Remote Sensing Applications: Society and Environment**, v. 24, p. 100646, 2021. 8, 64, 65

SOUZA, S. O.; FOLHARINI, S. de O. Mapeamento do uso e ocupação da terra do município de Petrolina (PE)–médio vale do rio São Francisco através do ndvi de

imagem landsat 8 (oli). **Revista Equador**, v. 8, n. 2, p. 489–502, 2019. 40, 41, 44, 53, 62, 63

SRIVASTAVA, R.; KUMAR, S.; KUMAR, B. 7 - classification model of machine learning for medical data analysis. In: GOSWAMI, T.; SINHA, G. (Ed.). **Statistical modeling in machine learning**. Academic Press, 2023. p. 111–132. ISBN 978-0-323-91776-6. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780323917766000178>>>. 20

SZUSTER, B. W.; CHEN, Q.; BORGER, M. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. **Applied Geography**, v. 31, n. 2, p. 525–532, 2011. 18

TABARELLI, M. et al. Caatinga: legado, trajetória e desafios rumo à sustentabilidade. **Ciência e Cultura**, v. 70, n. 4, p. 25–29, 2018. 8

TANG, X.; MA, Q.; ZHANG, X.; LIU, F.; MA, J.; JIAO, L. Attention consistent network for remote sensing scene classification. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 2030–2045, 2021. 22

TEFFERA, Z. L.; LI, J.; DEBSU, T. M.; MENEGESHA, B. Y. Assessing land use and land cover dynamics using composites of spectral indices and principal component analysis: a case study in middle awash subbasin, ethiopia. **Applied Geography**, v. 96, p. 109–129, 2018. 38

TEIXEIRA, A. de C.; BASTIAANSEN, W. G.; AHMAD, M.-u.-D.; MOURA, M. d.; BOS, M. Analysis of energy fluxes and vegetation-atmosphere parameters in irrigated and natural ecosystems of semi-arid brazil. **Journal of Hydrology**, v. 362, n. 1-2, p. 110–127, 2008. 9

TEMENOS, A.; TEMENOS, N.; KASELIMI, M.; DOULAMIS, A.; DOULAMIS, N. Interpretable deep learning framework for land use and land cover classification in remote sensing using shap. **IEEE Geoscience and Remote Sensing Letters**, v. 20, p. 1–5, 2023. 22

THINLEY, J.; PICKERING, C.; NDEHEDEHE, C. Using vegetation and chlorophyll indices to model above ground biomass over time in an urban arboretum in subtropical queensland. **Remote Sensing Applications: Society and Environment**, v. 34, p. 101202, 2024. 16

TRAN, T. V.; REEF, R.; ZHU, X. A review of spectral indices for mangrove remote sensing. **Remote Sensing**, v. 14, n. 19, p. 4868, 2022. 14

TUSSUPOV, J.; YESSENOVA, M.; ABDIKERIMOVA, G.; AIMBETOV, A.; BAKTYBEKOV, K.; MURZABEKOVA, G.; AITIMOVA, U. Analysis of formal concepts for verification of pests and diseases of crops using machine learning methods. **IEEE Access**, 2024. 15

UMOH, U. A.; EYOH, I. J.; MURUGESAN, V. S.; NYOHO, E. E. Fuzzy-machine learning models for the prediction of fire outbreaks: a comparative analysis. In: PANDEY, R.; KHATRI, S. K.; SINGH, N. K.; VERMA, P. (Ed.). **Artificial intelligence and machine learning for EDGE computing**. [S.l.: s.n.], 2022. p. 207–233. 19, 20

VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: Springer, 1999. 2, 18

VETRIVEL, A.; GERKE, M.; KERLE, N.; NEX, F.; VOSSSELMAN, G. Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 140, p. 45–59, 2018. 22

VITÓRIAS, M. Maria das; HOELZEMANN, J. J.; NETO, E. R. de S.; AGUIAR, A. P. D.; MARCIA, R. Emissões de co2 provenientes do uso e mudanças no uso da terra no bioma caatinga no nordeste brasileiro. **Revista Brasileira de Geografia Física**, v. 8, n. 01, p. 144–155, 2015. 4

VOLPI, M.; TUIA, D. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 2, p. 881–893, 2016. 23

VUOLO, F.; NEUWIRTH, M.; IMMITZER, M.; ATZBERGER, C.; NG, W.-T. How much does multi-temporal sentinel-2 data improve crop type classification? **International Journal of Applied Earth Observation and Geoinformation**, v. 72, p. 122–130, 2018. 13

WANG, L.; YANG, Z.-L. Changes in land use influenced by anthropogenic activity. In: **Oxford research encyclopedia of environmental science**. [S.l.: s.n.], 2020. 1

WOODCOCK, C. E.; LOVELAND, T. R.; HEROLD, M.; BAUER, M. E. Transitioning from change detection to monitoring with remote sensing: a paradigm shift. **Remote Sensing of Environment**, v. 238, p. 111558, 2020. 13

XU, H. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. **International Journal of Remote Sensing**, v. 27, n. 14, p. 3025–3033, 2006. 15, 16, 28

XUE, J.; SU, B. Significant remote sensing vegetation indices: a review of developments and applications. **Journal of Sensors**, v. 2017, 2017. 14

YANG, X.; ZHAO, S.; QIN, X.; ZHAO, N.; LIANG, L. Mapping of urban surface water bodies from sentinel-2 msi imagery at 10 m resolution via ndwi-based image sharpening. **Remote Sensing**, MDPI, v. 9, n. 6, p. 596, 2017. 16

YILMAZ, O. S.; GULGEN, F.; SANLI, F. B.; ATES, A. M. The performance analysis of different water indices and algorithms using sentinel-2 and landsat-8

images in determining water surface: Demirkopru dam case study. **Arabian Journal for Science and Engineering**, Springer, v. 48, n. 6, p. 7883–7903, 2023. 16

YU, X.; WU, X.; LUO, C.; REN, P. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. **GIScience & Remote Sensing**, v. 54, n. 5, p. 741–758, 2017. 22

ZHA, Y.; GAO, J.; NI, S. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. **International Journal of Remote Sensing**, v. 24, n. 3, p. 583–594, 2003. 15, 16, 28

ZHANG, K.; LIU, N.; CHEN, Y.; GAO, S. Comparison of different machine learning method for GPP estimation using remote sensing data. **IOP Conference Series: Materials Science and Engineering**, v. 490, n. 6, p. 062010, 2019. Disponível em: <<<https://dx.doi.org/10.1088/1757-899X/490/6/062010>>>. 17

ZHANG, Q.; YUAN, Q.; ZENG, C.; LI, X.; WEI, Y. Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. **IEEE Transactions on Geoscience and Remote Sensing**, v. 56, n. 8, p. 4274–4288, 2018. 23

ZHONG, L.; HU, L.; ZHOU, H. Deep learning based multi-temporal crop classification. **Remote Sensing of Environment**, v. 221, p. 430–443, 2019. 24

ZHU, W.; MA, Y.; ZHOU, Y.; BENTON, M.; ROMAGNOLI, J. Deep learning based soft sensor and its application on a pyrolysis reactor for compositions predictions of gas phase components. In: EDEN, M. R.; IERAPETRITOU, M. G.; TOWLER, G. P. (Ed.). **13th International Symposium on Process Systems Engineering (PSE 2018)**. [s.n.], 2018, (Computer Aided Chemical Engineering, v. 44). p. 2245–2250. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/B9780444642417503694>>>. 22

ZHU, X. X.; TUIA, D.; MOU, L.; XIA, G.-S.; ZHANG, L.; XU, F.; FRAUNDORFER, F. Deep learning in remote sensing: a comprehensive review and list of resources. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 4, p. 8–36, 2017. 2

ZOU, Q.; NI, L.; ZHANG, T.; WANG, Q. Deep learning based feature selection for remote sensing scene classification. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 11, p. 2321–2325, 2015. 22

## **APÊNDICE A**

### **A.1 Uso de séries temporais para classificações de uso e cobertura da terra em Petrolina, Pernambuco**

A pesquisa realizada no município de Petrolina, Pernambuco, comparando as abordagens de classificação de uso e cobertura da terra utilizando séries temporais de NDVI e uma única imagem NDVI por ano, é descrita neste artigo com mais detalhes.

# USO DE SÉRIES TEMPORAIS PARA CLASSIFICAÇÕES DE USO E COBERTURA DA TERRA EM PETROLINA, PERNAMBUCO

Pedro V. da Silva Brito<sup>1</sup>, Michel E. D. Chaves<sup>1</sup>, Herica F. de Sousa Carvalho<sup>2</sup>, Felipe C. de Souza<sup>1</sup>, Baggio L. de Castro e Silva<sup>1</sup>, Karine R. Ferreira<sup>1</sup>, Rafael D. Coelho dos Santos<sup>1</sup> e Gilberto R. de Queiroz<sup>1</sup>,

<sup>1</sup>Instituto Nacional de Pesquisas Espaciais (INPE), Av. dos Astronautas, 1758 - São José dos Campos-SP, Brasil, {pedro.brito;michel.chaves;felipe.carvalho;baggio.silva;karine.ferreira;rafael.santos;gilberto.queiroz}@inpe.br

<sup>2</sup>Universidade Federal de Pernambuco (UFPE), Av. Prof. Moraes Rego 1235, 50670-901, Cidade Universitária, Recife-PE, Brasil, herica.carvalho@ufpe.br

## RESUMO

A vasta gama de imagens de satélite disponibilizadas gratuitamente e os avanços de métodos computacionais ampliaram o uso de técnicas de sensoriamento remoto para analisar a superfície terrestre, melhorando análises sobre vegetação. Em regiões como a zona semiárida do bioma Caatinga, isso tem sido essencial para avaliar as condições e a sazonalidade característica da vegetação local. O objetivo deste estudo foi gerar classificações de uso e cobertura da terra e avaliar mudanças entre 2018 e 2020 por meio de duas abordagens: (1) classificação de séries temporais de imagens de satélite e (2) classificação de cada imagem de forma individual. Esse estudo conclui que a primeira abordagem usando o método *Randon Forest* e séries temporais NDVI potencializou a capacidade de diferenciar classes de uso e cobertura da terra em Petrolina-PE.

**Palavras-chave** – Random Forest, Cubos de dados, Análise temporal.

## ABSTRACT

The wide range of freely available satellite images and advances in computational methods have promoted the use of remote sensing techniques to analyze the Earth's surface, improving analysis of vegetation. In regions such as the semi-arid zone of the Caatinga biome, this has been essential to assess the conditions and seasonality characteristic of the local vegetation. The objective of this study was to generate land use and land cover classifications and assess changes between 2018 and 2020 through two approaches: (1) classification of satellite image time series and (2) classification of each satellite image individually. This study concludes that the first approach using the *Randon Forest* method and NDVI time series enhanced the differentiation of land use and land cover classes in Petrolina-PE.

**Key words** – *Random Forest*, *Data cubes*, *Temporal analysis*.

## 1. INTRODUÇÃO

Rico em recursos naturais quando comparado a outras regiões semiáridas do mundo, o bioma Caatinga está entre os ecossistemas menos estudados e mais ameaçados no quesito mudanças de uso e cobertura da terra no Brasil [1]. Contudo, a retirada de vegetação nativa em demasia, especialmente por ação antrópica, tem alterado processos físicos, químicos e biológicos em escala local, provocando mudanças no

ciclo hidrológico e no clima local [2]. Neste sentido, melhorar a compreensão sobre condições e sazonalidade da vegetação é uma necessidade premente. Sensoriamento remoto é amplamente utilizado no desenvolvimento de metodologias que possam embasar ações voltadas ao planejamento e monitoramento do uso adequado da terra e, conseqüentemente, da vegetação. Análises de imagens de satélite são úteis para o monitoramento sistemático e contínuo da cobertura vegetal de agroecossistemas [3].

A disseminação gratuita de imagens de satélite possibilitou aumentar o entendimento sobre a superfície terrestre e sua dinâmica [4]. Isso tornou possível o uso de séries temporais de imagens para gerar informações precisas sobre uso e cobertura da terra e monitorar processos naturais e antrópicos. Para análise de séries temporais de imagens, o pacote R SITS (*Satellite Image Time Series Analysis on Earth Observation Data Cubes*) fornece um conjunto de métodos, incluindo aprendizado de máquina e aprendizado profundo, para gerar classificações de uso e cobertura da terra [5].

O potencial de ferramentas como esta, tem tornado possível melhorar análises acerca da sazonalidade da vegetação em áreas de constante mudanças no uso da terra ao longo de determinados períodos. A zona semiárida da Caatinga tem sofrido intensa degradação nos últimos 400 anos devido ao uso desordenado e predatório [6]. Um exemplo está no município de Petrolina, localizado em Pernambuco, que é responsável pelo desenvolvimento de grandes projetos agrícolas irrigados ao longo de sua extensão, promovendo e intensificando mudanças do uso da terra. Assim, o objetivo deste estudo foi gerar classificações de uso e cobertura da terra em Petrolina a partir de imagens Sentinel-2/MultiSpectral Instrument (S2/MSI) e avaliar mudanças entre 2018 e 2020 por meio de duas abordagens: (1) classificação de séries temporais de imagens e (2) de uma imagem por ano de forma individual.

## 2. MATERIAL E MÉTODOS

Maior polo de fruticultura do Semiárido brasileiro, Petrolina tem área de 4.561,870  $km^2$  [7]. O clima é Semiárido quente e seco, com períodos definidos de chuva, entre novembro e abril, e estiagem, entre maio e outubro. A temperatura média é acima de 25°C, e a pluviosidade de cerca de 500 mm/ano [8]. As etapas envolveram a classificação de séries temporais e de imagens individuais anuais (Figura 1). Foram coletadas 3000 amostras de uso e cobertura da terra, 1000/ano, 200 de cada uma das classes: Vegetação de caatinga, Agricultura irrigada, Corpos hídricos, Área urbana e Solo exposto.



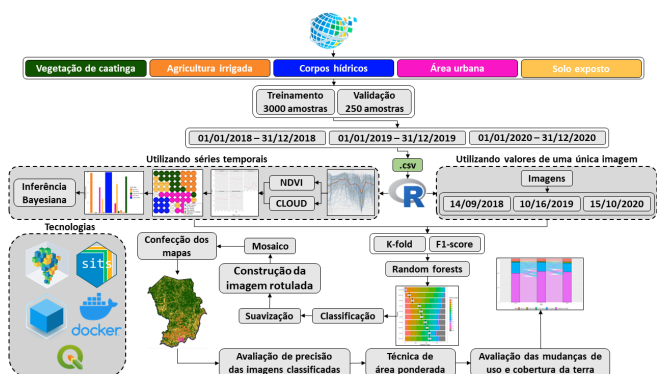


Figura 1: Fluxograma das etapas metodológicas do trabalho.

Para validar, foram coletadas 750 amostras (250/ano, 50/classe) via interpretação visual de imagens S2/MSI, com auxílio de séries temporais NDVI, usando a plataforma *TerraCollect*, desenvolvida no escopo do projeto *Brazil Data Cube* (BDC) [3]. Na classificação de séries temporais, foram usados cubos de dados S2/MSI disponibilizados pelo projeto BDC (S2-SEN2COR\_10\_16D\_STK-1) e séries temporais NDVI extraídas desses cubos [3]. Eles possuem imagens com resolução espacial de 10 metros, projetadas e recortadas na grade BDC, com função de composição temporal de 16 dias usando a abordagem de melhor *pixel*. Também foi usada a banda SCL para mitigar efeitos de nuvens/sombras e aplicada interpolação linear para substituir valores removidos. Padrões espectro-temporais NDVI foram avaliados via *Generalized Additive Model* (GAM). As amostras foram filtradas pelo método proposto por Santos et al. (2021) [9], que usa mapas auto-organizáveis (*Self-Organizing Maps*, SOM) e inferência Bayesiana. Foram gerados 5 experimentos, variando tamanhos da grade de neurônios das amostras, adotando parâmetros propostos [9], que foram  $5\sqrt{\frac{N}{2}}$  neurônios, em que,  $N$  é o número de amostras, taxa de aprendizado decrescente, de 0,50 a 0,01, e métrica euclidiana. Os tamanhos de grades foram 8x8, 9x9, 10x10, 11x11 e 12x12, com taxa de aprendizado de 0,50 a 0,01, 100 interações e distância euclidiana. Para classificar imagens individuais, foi escolhida uma imagem por ano (com menos nuvens/sombras, datadas de 14/09/2018, 16/10/2019 e 15/10/2020).

Nas duas etapas, foi aplicada a validação cruzada *k-fold* ( $k = 5$ ) para selecionar os melhores modelos de classificação. Os modelos *Random Forest* (RF) avaliados diferenciam-se pelo número de árvores (100, 500, 1000, 1500 e 2000). Após a seleção do modelo mais preciso por ano, foi aplicada a avaliação *F1-score*. Os passos seguintes envolveram treinamento dos modelos, classificação e suavização bayesiana, para refiná-las. A precisão das classificações foi avaliada via técnica de área ponderada [10]. Mudanças de uso e cobertura da terra foram avaliadas via diagrama *alluvial*, analisando 3000 amostras aleatórias.

### 3. RESULTADOS E DISCUSSÃO

Os padrões espectro-temporais NDVI (Figura 2) mostram como as classes possuem diferenças. Apenas Área urbana e Solo exposto apresentaram perfis semelhantes. Em relação aos agrupamentos das séries temporais via SOM (Figura 3),

em 2019 e 2020, houve apenas um neurônio *outlier*, com classes divergentes das predominantes na vizinhança. O agrupamento de 2018 não apresentou *outliers*.



Figura 2: Padrões espectro-temporais NDVI de cada classe para o NDVI obtido via GAM para os anos de 2018, 2019 e 2020.

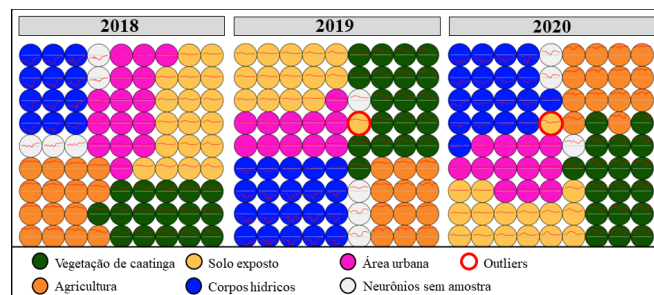


Figura 3: Agrupamentos das séries temporais de NDVI na grade 9x9 para os anos 2018, 2019 e 2020.

A aplicação do SOM mostra, em todo o período, agrupamentos com graus de pureza acima de 86% (Figura 4).

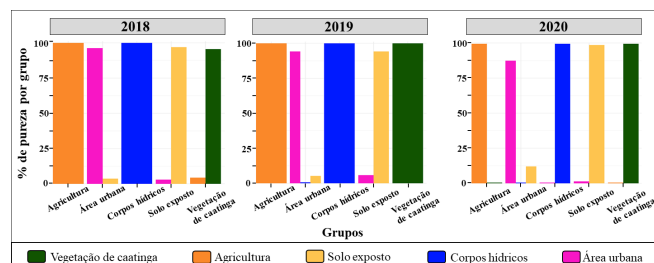


Figura 4: Porcentagem de confusão entre os grupos do agrupamento SOM com grade 9x9 para 2018, 2019 e 2020.

Os agrupamentos de 2018 e 2019 apresentaram maior separabilidade, com 100% de pureza em Agricultura irrigada e Corpos hídricos. Em particular, 2019 apresentou 100% de pureza para Vegetação de caatinga. Os grupos de Área urbana e Solo exposto apresentaram maior confusão, com exceção de 2018, em que, Vegetação de caatinga mostrou confusão

com Agricultura irrigada. A classe Corpos hídricos não foi homogênea em 2020.

Os resultados obtidos com a aplicação da técnica *k-fold* ( $k = 5$ ) (Tabela 1) mostram maiores valores de acurácia utilizando séries temporais, e modelos com acurácia satisfatória nas duas abordagens. Em 2018, as maiores acurácias ocorreram no modelo com 500 árvores (97,9% e 77,6% para séries temporais e única imagem, respectivamente). Por outro lado, o modelo com 100 árvores teve acurácia de 98,0% e 98,5% em 2019 e 2020 usando séries temporais e de 76,9% e 74,9% com a imagem anual.

Utilizando séries temporais					
Acurácia %					
Nº árvores	100	500	1000	1500	2000
2018	97,6	97,9	97,5	97,8	97,4
2019	98,0	97,8	97,7	97,9	97,7
2020	98,5	98,5	98,0	98,0	98,0

Utilizando uma única imagem por ano					
Nº árvores	100	500	1000	1500	2000
2018	76,9	77,6	75,3	76,3	76,0
2019	76,9	75,9	75,7	76,9	76,4
2020	74,6	73,2	72,6	73,7	73,8

**Tabela 1: Validação *k-fold* no conjunto de treinamento com o modelo RF utilizando séries temporais e uma única imagem.**

Os resultados do *F1-score* (Tabela 2) apresentam, para cada classe, maiores valores usando séries temporais (acima de 90% de precisão). Com séries temporais, os menores valores ocorreram para Vegetação de caatinga (93,8%, 94,7% e 95,2%). Para Corpos hídricos, maior exatidão (100% e 99,7% para 2018 e 2019). Para Solo exposto, 99,5% para todos os conjuntos. Utilizando apenas uma imagem, os maiores valores foram das classes Solo exposto, e Corpos hídricos com 97,3% e 94,3%, 94%, respectivamente. Semelhante aos resultados com séries temporais, Vegetação de caatinga teve menor acurácia usando uma única imagem.

Utilizando séries temporais			
Acurácia %			
Classes	2018	2019	2020
Agricultura	99,2	99,2	99,2
Área urbana	95,1	95,5	94,5
Corpos hídricos	100,0	99,7	99,0
Solo exposto	99,5	99,5	99,5
Vegetação de caatinga	94,7	95,2	93,8

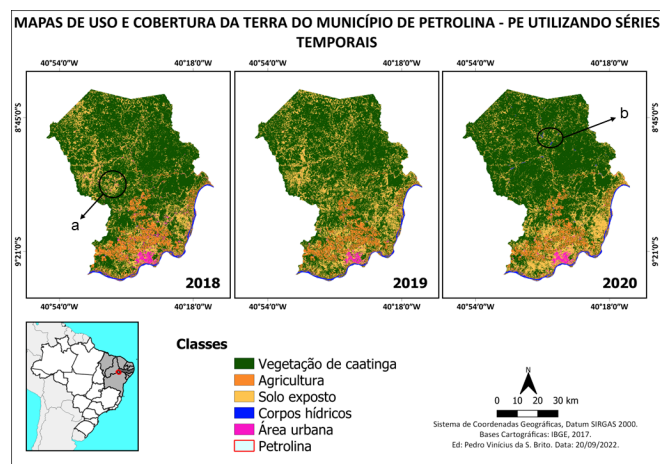
  

Utilizando uma única imagem por ano			
Classes	2018	2019	2020
Agricultura	86,5	61,4	91,0
Área urbana	63,8	88,7	56,9
Corpos hídricos	93,8	94,0	93,9
Solo exposto	97,3	94,3	77,2
Vegetação de caatinga	45,4	53,3	54,0

**Tabela 2: *F1-scores* no conjunto de treinamento com o modelo RF utilizando séries temporais e uma única imagem por ano.**

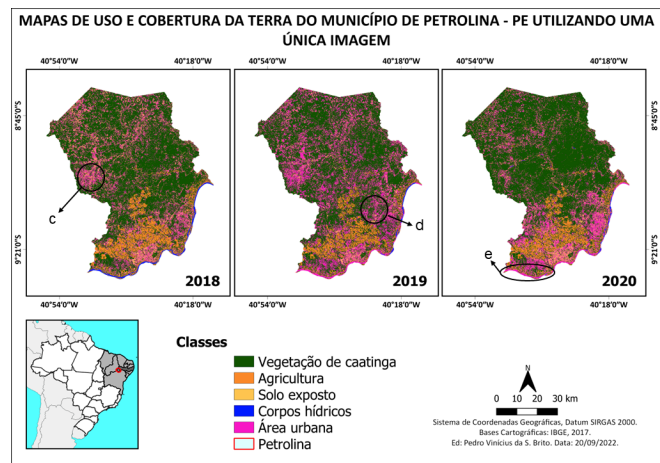
A classificação de uso e cobertura da terra utilizando séries temporais e modelo RF (Figura 5) apresenta melhor identificação de Áreas agrícolas e do rio São Francisco. Por outro lado, observou-se algumas confusões em regiões destacadas (região a) em que, locais de Solo exposto foram

classificados como Área urbana. Em 2020, detectou-se mais Corpos hídricos (região b), o que pode ter relação com a precipitação. Por interpretação visual e conhecimento da região, a classificação de 2019 foi mais precisa.



**Figura 5: Classificações de uso e cobertura da terra via RF para Petrolina-PE utilizando séries temporais.**

A classificação usando uma única imagem anual e modelo RF (Figura 6) mostra maior confusão entre as classes, sobretudo, Solo exposto e Área urbana. As regiões c e d representam locais de Solo exposto que foram classificados como Área urbana. Embora 2018 tenha detectado o leito do rio São Francisco, 2019 e 2020 apresentaram confusões entre Corpos hídricos e Área urbana, predominantemente ao sul.



**Figura 6: Classificações de uso e cobertura da terra via RF para Petrolina-PE utilizando uma única imagem por ano.**

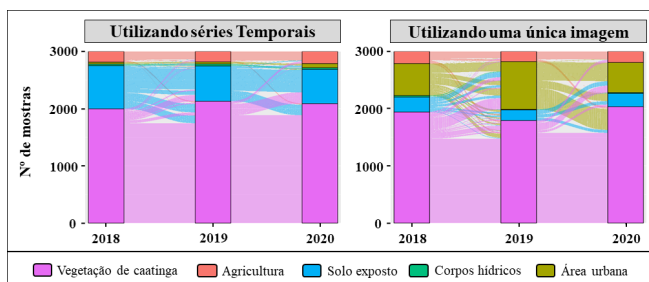
O uso de séries temporais aumentou a acurácia da maioria das classes (Tabela 3). Vegetação de caatinga obteve 100% de acurácia do produtor em todas as situações. No entanto, a classe de Área urbana continuou apresentando confusão. As transições entre classes (Figura 7) mostram, na abordagem de séries temporais, mudanças oriundas da ocupação local, como a transição Vegetação de caatinga-Solo exposto; o que não foi observado com uma imagem por ano.

Os resultados mostraram diferenças entre as abordagens. A abordagem de classificação via séries temporais aperfeiçoou a capacidade de diferenciar classes de uso e cobertura da terra (Tabela 3), diminuindo, conseqüentemente, confusões

Classes	AP %		AU %		AG %	
	ST	UI	ST	UI	ST	UI
<b>2018</b>						
Agricultura	60	55	100	96	87	77
Área urbana	24	56	82	62		
Corpos hídricos	100	73	100	100		
Solo exposto	92	50	87	62		
Vegetação de caatinga	100	99	84	80		
<b>2019</b>						
Agricultura	100	76	100	100	96	79
Área urbana	22	78	89	45		
Corpos hídricos	100	03	100	100		
Solo exposto	99	55	85	84		
Vegetação de caatinga	100	99	100	91		
<b>2020</b>						
Agricultura	71	66	100	96	94	76
Área urbana	30	59	88	29		
Corpos hídricos	99	03	100	100		
Solo exposto	93	39	88	50		
Vegetação de caatinga	100	100	94	91		

ST = Séries temporais, UI = Uma imagem, AP = Acurácia do produtor; AU = Acurácia do usuário, AG = Acurácia Global.

**Tabela 3: Valor de acurácia das imagens classificadas a partir da técnica de área ponderada.**



**Figura 7: Mudanças de uso e cobertura da terra dos mapas classificados utilizando séries temporais e uma única imagem.**

espectrais. Previamente, verificou-se similaridade do perfil espectro-temporal entre as classes Solo exposto e Área urbana (Figura 2), apresentando forte confusão entre as mesmas, devido à similaridade de suas assinaturas espectro-temporais.

A presença de neurônio *outlier* na Figura 3 não representa rotulagem errada das amostras coletadas [9]. Isso pode ocorrer devido às amostras com diferentes padrões espaço temporais para a mesma classe ou não-separáveis com o uso de séries temporais NDVI. A heterogeneidade apresentada na vizinhança de alguns agrupamentos pode representar a variação natural entre as amostras da mesma classe e não devem ser removidas do conjunto de amostras.

Fazendo uso do *k-fold* utilizando o mesmo número de árvores nos dois testes, a abordagem de séries temporais obteve maior acurácia (Tabela 1). Assim, foi possível identificar aumento na acurácia com mais amostras representativas. Já nos resultados do *F1-score* (Tabela 2) percebe-se pequena variabilidade no percentual para Vegetação nativa (caatinga), o que pode estar associado à alta heterogeneidade espacial e variabilidade da cobertura vegetal associada à sazonalidade pluviométrica da região. Ademais, observou-se similaridade entre Solo exposto e

Corpos hídricos, alvos pouco alterados durante o ano.

#### 4. CONCLUSÕES

O uso de séries temporais NDVI melhorou a identificação de classes de uso e cobertura da terra em comparação a abordagem de classificação de uma imagem anual e favoreceu a identificação de mudanças na área de estudo. Como o município de Petrolina apresenta dinâmica de conversão envolvendo diferentes classes de vegetação, recomenda-se o uso da abordagem de séries temporais para detectar mudanças relacionadas à sua sazonalidade e dinâmica para o caso de áreas de estudo com características semelhantes.

#### 5. AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Código de Financiamento 001). Os autores também agradecem à FAPESP (processo 2021/07382-2 - Michel E. D. Chaves) e ao projeto Brazil Data Cube.

#### 6. REFERÊNCIAS

- [1] Z. Fernández, J. A. Santos Júnior, R. Amaral, E. França, and et al. Pernambuco caatinga: relevance of soil chemical composition for biodiversity conservation. *Chemistry and Ecology*, 38:108–121, 2022.
- [2] J. Q. Chambers and P. Artaxo. Deforestation size influences rainfall. *Nature Climate Change*, 7:175–176, 2017.
- [3] K. Ferreira, G. Queiroz, L. Vinhas, R. Marujo, and et al. Earth observation data cubes for brazil: Requirements, methodology and products. *Remote Sensing*, 12(24):4033, 2020.
- [4] M. Chaves, M. Picoli, and I. Sanches. Recent applications of landsat 8/OLI and sentinel-2/MSI for land use and land cover mapping: A systematic review. *Remote Sensing*, 12(18):3062, 2020.
- [5] R. Simoes, G. Camara, G. Queiroz, and et al. Satellite image time series analysis for big earth observation data. *Remote Sensing*, 13(13):2428, 2021.
- [6] J. Bezerra, G. Moura, B. Silva, P. Lopes, and Ê. Silva. Parâmetros biofísicos obtidos por sensoriamento remoto em região semiárida do estado do rio grande do norte, brasil. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18:73–84, 2014.
- [7] IBGE. Instituto brasileiro de geografia e estatística, 2022. Disponível em: <<https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>>. Acesso em: 21/10/2022.
- [8] J. D. Galvêncio, S. M. Mendes, Y. C. B. Morais, R. Q. Miranda, and et al. Correlação linear entre a precipitação e o índice de área foliar do bioma caatinga. *Revista Brasileira de Geografia Física*, 13:3304–3314, 2020.
- [9] L. Santos, K. Ferreira, and et al. Quality control and class noise reduction of satellite image time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177:75–88, 2021.
- [10] P. Olofsson and et al. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129:122–131, 2013.



## **PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE**

### **Teses e Dissertações (TDI)**

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

### **Manuais Técnicos (MAN)**

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

### **Notas Técnico-Científicas (NTC)**

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

### **Relatórios de Pesquisa (RPQ)**

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

### **Propostas e Relatórios de Projetos (PRP)**

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

### **Publicações Didáticas (PUD)**

Incluem apostilas, notas de aula e manuais didáticos.

### **Publicações Seriadas**

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Contam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

### **Programas de Computador (PDC)**

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

### **Pré-publicações (PRE)**

Todos os artigos publicados em periódicos, anais e como capítulos de livros.