



Universidade Estadual de Campinas  
Faculdade de Engenharia Agrícola



Daniel Hideki Shibuya

**Avaliação de Modelos Baseados em Métodos de Aprendizado  
de Máquina para a Classificação de Perfis Espectro-  
Temporais de Imagens do Sensor MODIS**

**Evaluation of Machine Learning-Based Models for the  
Classification of MODIS Image Spectral-Temporal Profiles**

Campinas  
2024

Daniel Hideki Shibuya

**Avaliação de Modelos Baseados em Métodos de Aprendizado  
de Máquina para a Classificação de Perfis Espectro-  
Temporais de Imagens do Sensor MODIS**

**Evaluation of Machine Learning-Based Models for the  
Classification of MODIS Image Spectral-Temporal Profiles**

Dissertação apresenta à Faculdade de Engenharia Agrícola – FEAGRI como parte dos requisitos exigidos para obtenção do título de Mestre em Engenharia Agrícola, na área de Agricultura Digital.

Dissertation presented to the Faculty of Agricultural Engineering - FEAGRI in partial fulfillment of the requirements for the degree of Master in Agricultural Engineering, in the area of Digital Agriculture.

Orientador: Júlio César Dalla Mora Esquerdo

Este trabalho corresponde à versão final  
Dissertação/tese defendida pelo aluno  
Daniel Hideki Shibuya, e orientada pelo  
Prof. Dr. Júlio César Dalla Mora  
Esquerdo.

Campinas  
2024

Ficha catalográfica  
Universidade Estadual de Campinas (UNICAMP)  
Biblioteca da Área de Engenharia e Arquitetura  
Elizângela Aparecida dos Santos Souza - CRB 8/8098

Shibuya, Daniel Hideki, 1996-  
Sh612a Avaliação de modelos baseados em métodos de aprendizado de máquina para a classificação de perfis espectro-temporais de imagens do sensor MODIS / Daniel Hideki Shibuya. – Campinas, SP : [s.n.], 2024.

Orientador: Júlio César Dalla Mora Esquerdo.  
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Agrícola.

1. Aprendizado de máquina. 2. Redes neurais profundas. 3. Generalização.  
I. Esquerdo, Júlio César Dalla Mora, 1977-. II. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Agrícola. III. Título.

Informações Complementares

**Título em outro idioma:** Evaluation of machine learning- based models for the classification of MODIS image spectral-temporal profiles

**Palavras-chave em inglês:**

Machine learning

Deep neural network

Generalization

**Área de concentração:** Agricultura Digital

**Titulação:** Mestre em Engenharia Agrícola

**Banca examinadora:**

Júlio César Dalla Mora Esquerdo [Orientador]

Erivelto Mercante

João Francisco Gonçalves Antunes

**Data de defesa:** 21-06-2024

**Programa de Pós-Graduação:** Engenharia Agrícola

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: 0009-0005-1166-3517

- Currículo Lattes do autor: <https://lattes.cnpq.br/4690432918632045>

Este exemplar corresponde à redação final da **Dissertação de Mestrado** defendida por **Daniel Hideki Shibuya**, aprovada pela Comissão Julgadora em 21 de Junho de 2024, na Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas.

**FEAGRI**

---

**Prof. Dr. Júlio César Dalla Mora Esquerdo – Presidente e Orientador**

---

**Prof. Dr. João Francisco Gonçalves Antunes – Membro Titular**

---

**Prof. Dr. Erivelto Mercante – Membro Titular**

Faculdade de  
**Engenharia Agrícola**  
Unicamp

**A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.**

## **Agradecimentos**

Agradeço a todos os meus familiares que me apoiaram durante o período do Mestrado, que incentivaram e apoiaram a minha decisão de cursar um programa de pós graduação.

Agradeço a todos meus amigos e companheiros de pós-graduação que estiveram comigo durante esses últimos dois anos, dividindo experiências, aprendendo juntos e me auxiliando em momentos de dificuldades

Agradeço a todos os pesquisadores e colaboradores que estiveram envolvidos neste projeto, que dividiram parte do seu conhecimento e me auxiliaram durante o desenvolvimento.

Agradeço o meu orientador Júlio César Dalla Mora Esquerdo e a professora Gleyce Kelly Dantas Araújo Figueiredo pelos valiosos ensinamentos, conselhos, paciência e orientação durante os últimos dois anos.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), bolsa processo 130329/2022-3

## Resumo

Métodos de aprendizado de máquina aplicados aos dados de sensoriamento remoto permitem análises cada vez mais complexas em relação a uma área de estudo. Na superfície terrestre, padrões de culturas agrícolas, pastagem e outros tipos de vegetação podem ser identificados por meio da análise de perfis espectro-temporais obtidos por sensores orbitais, como o MODIS. Apesar dos avanços das técnicas de classificação digital, muitos estudos acabam sendo limitados devido à falta de amostras de campo para o treinamento de modelos de classificação, principalmente os baseados em redes neurais profundas. Neste contexto, o presente trabalho propõe uma abordagem baseada no conceito de Generalização, em que um modelo de classificação treinado a partir de um conjunto representativo de amostras é aplicado para a classificação interanual da série temporal do MODIS. O modelo treinado utilizou amostras obtidas para o ano-safra 2019/2020, a partir dos mapeamentos do projeto TerraClass, disponível nos biomas Cerrado e Amazônia. Os mapas foram adaptados para serem compatíveis com a resolução espacial do MODIS e utilizados como base para a coleta de amostras. Os algoritmos testados foram o *Random Forest*, TempCNN e L-TAE. Junto aos métodos de aprendizado de máquina, uma análise para identificação da qualidade das amostras foi feita por meio do algoritmo SOM, que influenciou positivamente os resultados obtidos. A Generalização dos modelos treinados no ano 2019/20 foi avaliada por meio de amostras de campo disponíveis em anos-safras diferentes daqueles utilizados no treinamento. O processo de avaliação identificou que os algoritmos TempCNN e L-TAE encontraram maior dificuldade em lidar com as variabilidades intraclasses, principalmente nas classes temáticas Agricultura Semiperene e Pastagem. O *Random Forest* apresentou os melhores resultados nas métricas avaliadas, principalmente após o tratamento pelo SOM, apresentando acurácia global de 0,917 e boa adaptação no processo de Generalização.

Palavras-Chaves: Uso e cobertura da terra, *Random Forest*, Redes Neurais Profundas, Generalização.

## **Abstract**

Machine learning methods applied to remote sensing data allow increasingly complex analysis of a study area. On the earth's surface, patterns of agricultural crops, grassland and other types of vegetation can be identified by analyzing Spectro-temporal profiles obtained by orbital sensors such as MODIS. Despite the advances in digital classification techniques, many studies end up being limited due to the lack of field samples for training classification models, especially those based on deep neural networks. In this context, this paper proposes an approach based on the concept of Generalization, in which a classification model trained on a representative set of samples is applied to the interannual classification of the MODIS time series. The trained model used samples obtained for the 2019/2020 crop year from the TerraClass project's mapping, available in the Cerrado and Amazon biomes. The maps were adapted to be compatible with MODIS spatial resolution and used as a basis for sampling. The algorithms tested were Random Forest, TempCNN and L-TAE. Along with the machine learning methods, an analysis to identify the quality of the samples was carried out using the SOM algorithm, which positively influenced the results obtained. The Generalization of the models trained in 2019/20 was assessed using field samples available in crop years other than those used for training. The evaluation process identified that the TempCNN and L-TAE algorithms had more difficulty to deal with intra-class variability, especially in the Semi-perennial Agriculture and Pasture classes. Random Forest showed the best results in the metrics evaluated, especially after treatment by SOM, with an overall accuracy of 0,917 and good adaptation in the Generalization process.

**Key-words:** Land use Land cover, Random Forest, Deep Neural Networks, Generalization.

## Lista de Figuras

<b>Figura 1:</b> Ilustração de arquitetura da rede TempCNN. Adaptado de PELLETIER; WEBB; PETITJEAN (2019). .....	30
<b>Figura 2:</b> Ilustração de arquitetura da rede L-TAE. Adaptado de GARNOT; LANDRIEU (2020). ...	32
<b>Figura 3:</b> Exemplo de agrupamento com dimensão 3x3 realizado pelo SOM, com presença de um neurônio (azul) que possivelmente é um ruído. ....	34
<b>Figura 4:</b> Mapa do estado de Mato Grosso, representando a distribuição dos três biomas ao longo do seu território. Fonte: IBGE (Instituto Brasileiro de Geografia e Estatística). ....	41
<b>Figura 5:</b> Fluxograma ilustrando a metodologia utilizada neste trabalho. ....	46
<b>Figura 6:</b> Figura ilustrando o processo de adaptação do mapeamento TerraClass para uma amostragem baseada nas imagens do MODIS. (A) Adaptação da resolução espacial de 10m (Sentinel-2MSI) para 250m (MODIS). (B) Exemplo de processo de remoção de bordas por meio da função Erode. (C) Alinhamento dos pixels do mapa TerraClass adaptado, com os pixels de uma imagem MODIS. ....	48
<b>Figura 7:</b> Figura comparativa de um perfil espectro-temporal sem correção de nuvens (A) e com correção realizada com o auxílio da banda CLOUD (B). ....	49
<b>Figura 8:</b> Fluxograma do processo de criação do mapa bidimensional e metodologia de análise de ruídos por meio das métricas de Probabilidade Condicional e Posterior. Adaptado de SANTOS et al. (2021a). ....	51
<b>Figura 9:</b> Padrões espectro-temporais das classes temáticas. Curvas com valores referentes à média de cada banda. ....	55
<b>Figura 10:</b> Mapa bidimensional gerado pelo SOM as com amostras de treinamento. Em destaque, exemplos de possíveis outliers identificadas por interpretação visual. ....	56
<b>Figura 11:</b> Gráficos frequência relativa referente a pureza dos neurônios de casa classe temática. ....	57
<b>Figura 12:</b> Exemplos de neurônios com perfis espectrais interpretados como ruídos dentro de suas respectivas classes temáticas. ....	58
<b>Figura 13:</b> Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Temporária de 1 ciclo. ....	62
<b>Figura 14:</b> Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Temporária de 2 ciclos. ....	63
<b>Figura 15:</b> Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Semiperene. ....	64
<b>Figura 16:</b> Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Pastagem. ....	65
<b>Figura 17:</b> Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Vegetação Florestal. ....	66
<b>Figura 18:</b> Figuras referentes a classificação interanual no município de Sorriso-MT resultado da Generalização feita com o modelo Random Forest (A) e TempCNN (B). Em destaque, o ano-safra correspondente a amostra de campo. ....	73
<b>Figura 19:</b> Figuras referentes a classificação interanual no município de Paranatinga-MT resultado da Generalização feita com o modelo Random Forest (A) e TempCNN (B). Em destaque, o ano-safra correspondente a amostra de campo. ....	75

<b>Figura 20:</b> Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Rondolândia-MT (A), Comodoro-MT (B) e Brasnorte-MT (C). Em destaque, o ano-safra correspondente a amostra de campo. ....	77
<b>Figura 21:</b> Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Campo Novo do Parecis-MT (A), Barra dos Bugres-MT (B) e Tangará da Serra-MT (C). Em destaque, o ano-safra correspondente a amostra de campo.....	79
<b>Figura 22:</b> Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Campo Novo do Parecis-MT (A) e Diamantino-MT (B). Em destaque, o ano-safra correspondente a amostra de campo. ....	81

## Lista de Tabelas

<b>Tabela 1:</b> Tabela com as classes temáticas e suas respectivas descrições utilizadas no mapeamento TerraClass, que serviram de base para as cinco classes temáticas utilizadas neste trabalho. ....	44
<b>Tabela 2:</b> Tabela referente a distribuição das amostras de campo das cinco classes temáticas, utilizadas para a etapa de teste da Generalização. ....	45
<b>Tabela 3:</b> Tabela referente à distribuição do número de amostras em cada classe temática, antes e após análise feita pelo SOM. ....	59
<b>Tabela 4:</b> Tabela referente aos valores de acurácia global obtidos para os seis cenários de treinamento do modelo de classificação no base 2019/2020. ....	68
<b>Tabela 5:</b> Matriz de confusão referente aos resultados obtidos pelo classificador RF na Amostragem Original. ....	68
<b>Tabela 6:</b> Tabela referente aos resultados de acurácia global de cada modelo na etapa de teste com as amostras de campo. ....	69
<b>Tabela 7:</b> Tabela referente aos resultados da métrica Precision dos modelos de classificação para cada classe temática. ....	69
<b>Tabela 8:</b> Tabela referente aos resultados da métrica Recall dos modelos de classificação para cada classe temática. ....	70
<b>Tabela 9:</b> Tabela referente aos resultados da métrica F1-Score dos modelos de classificação para cada classe temática. ....	70

## **Lista de Abreviações e Siglas**

Acatama – *Accuracy Assesment of Thematic Maps*

AGR-1 - Agricultura Temporária de 1 ciclo

AGR-2 - Agricultura Temporária de 2 ciclos

AGR-S - Agricultura Semiperene

AO – Amostragem Original

AS – Amostragem SOM

BDC – Brazil Data Cube

CDL – *Cropland Data Layer*

CNN – *Convolutional Neural Network*

CONAB – Companhia Nacional de Abastecimento

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária

EVI – *Enhanced Vegetation Index*

FAO – *Food and Agriculture Organization*

IBGE – Instituto Brasileiro de Geografia e Estatística

ILP – Integração Lavoura-Pecuária

L-TAE – *Lightweight Temporal Attention Encoder*

LSTM – *Long Short-Term Memory*

MLP – *Multi-Layer-Perceptron*

MODIS – *Moderate Resolute Imaging Spectroradiometer*

NASA – *National Aeronautics and Space Administration*

NDVI – *Normalize Difference Vegetation Index*

PAST – Pastagem

PC – Probabilidade Condicional

PP – Probabilidade Posterior

RF – *Random Forest*

RNN – *Recurrent Neural Network*

SATVeg – Sistema de Análise Temporal da Vegetação

SCP – *Semi-Automatic Classification Plugins*

Sits – *Satellite Image Time Serires Analysys for Earth Observation Data Cubes*

SMOTE – *Synthetic Minority Over-sampling Technique*

SOM – *Self-Organizing-Maps*

TempCNN – *Temporal Convolutional Neural Network*

USDA – Departamento de Agricultura dos Estados Unidos

VF – Vegetação Florestal

# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
<b>2</b>	<b>Objetivos Gerais e Específicos</b>	<b>19</b>
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>20</b>
<b>3.1</b>	<b>Séries Temporais de Imagens de Satélites</b>	<b>20</b>
<b>3.1.1</b>	<b>Índices de Vegetação</b>	<b>21</b>
<b>3.1.2</b>	<b>Sensor MODIS</b>	<b>22</b>
<b>3.1.3</b>	<b>Aplicações das Séries Temporais</b>	<b>24</b>
<b>3.2</b>	<b>Métodos de Aprendizado de Máquina</b>	<b>25</b>
<b>3.3</b>	<b>Aprendizado Profundo</b>	<b>28</b>
<b>3.3.1</b>	<b>Tipos de Redes Neurais</b>	<b>29</b>
<b>3.3.2</b>	<b>Transformer</b>	<b>31</b>
<b>3.4</b>	<b>Amostras de Campo e Qualidade do Conjunto de Dados</b>	<b>32</b>
<b>3.5</b>	<b>Metodologias para Lidar com Escassez de Amostras de Campo</b>	<b>35</b>
<b>3.5.1</b>	<b>Transfer Learning</b>	<b>36</b>
<b>3.5.2</b>	<b>Generalização</b>	<b>37</b>
<b>4</b>	<b>Material e Métodos</b>	<b>40</b>
<b>4.1</b>	<b>Material</b>	<b>40</b>
<b>4.1.1</b>	<b>Área de Estudo</b>	<b>40</b>
<b>4.1.2</b>	<b>Pacote Computacional Sits</b>	<b>41</b>
<b>4.1.3</b>	<b>Dados Espectrais</b>	<b>42</b>
<b>4.1.3</b>	<b>Amostras de Treinamento dos Modelos de Classificação</b>	<b>43</b>
<b>4.1.4</b>	<b>Amostras para Teste da Capacidade de Generalização dos Modelos</b>	<b>45</b>
<b>4.2</b>	<b>Métodos</b>	<b>45</b>
<b>4.2.1</b>	<b>Criação do Conjunto de Dados de Treinamento e Teste</b>	<b>46</b>
<b>4.2.2</b>	<b>Aplicação de Filtros</b>	<b>49</b>
<b>4.2.3</b>	<b>Tratamento de Amostras (SOM)</b>	<b>50</b>
<b>4.2.4</b>	<b>Criação dos Modelos de Classificação</b>	<b>52</b>
<b>4.2.5</b>	<b>Validação dos Modelos de Classificação</b>	<b>53</b>
<b>4.2.6</b>	<b>Generalização</b>	<b>53</b>
<b>5</b>	<b>Resultados e Discussões</b>	<b>55</b>
<b>5.1</b>	<b>Análise Exploratória dos Dados</b>	<b>55</b>
<b>5.2</b>	<b>Qualidade das Amostras– SOM</b>	<b>56</b>
<b>5.3</b>	<b>Análise após Remoção de Ruídos</b>	<b>59</b>
<b>5.4</b>	<b>Resultados do Modelo para o Ano Base 2019/2020</b>	<b>67</b>

<b>5.5 Resultados Generalização.....</b>	<b>68</b>
<b>5.5.1 Validação com Amostras de Campo.....</b>	<b>68</b>
<b>5.5.2 Classificação série temporal – 2004 até 2022.....</b>	<b>73</b>
<b>6 Considerações Finais e Conclusão.....</b>	<b>83</b>
<b>7 Referências Bibliográficas.....</b>	<b>85</b>
<b>Apêndice A – Matriz de Confusão (Amostras Teste).....</b>	<b>91</b>
<b>Apêndice B – Camadas de Probabilidade RF x TempCNN.....</b>	<b>92</b>
<b>Apêndice C – Camadas de Probabilidade RF-AS.....</b>	<b>94</b>

## 1 Introdução

O Brasil destaca-se na produção de commodities agrícolas, de acordo com relatórios da FAO (*Food and Agriculture Organization*). No ano de 2023, o país esteve entre os maiores produtores mundiais, com destaque para culturas como cana-de-açúcar, café, soja, algodão e milho (FAO, 2023). Neste contexto, informações referentes a estimativa de safra são importantes para o planejamento logístico e econômico do país. Órgãos como a Companhia Nacional de Abastecimento (CONAB) fornecem essas estimativas e utilizam dados de sensoriamento remoto para auxiliar nesta tarefa.

Nas últimas décadas, dados de sensoriamento remoto têm sido utilizados em estudos que visam monitorar a produção agrícola do país. Imagens obtidas por sensores a bordo de satélites permitem que o usuário monitore lavouras sem precisar estar presente no local (JENSEN, 2011). Os avanços tecnológicos na área de sensoriamento remoto permitiram o lançamento de diversas missões visando o monitoramento da superfície da Terra. O aumento na disponibilidade de produtos derivados dos diversos sensores orbitais permite que o usuário tenha a opção de escolher qual produto se adapta melhor à sua demanda (JENSEN, 2011). Por exemplo, os sensores Sentinel-2/MSI e Landsat-8/OLI são utilizados em estudos que exigem maior detalhamento do alvo analisado, como pequenas propriedades ou regiões com alta dinâmica de culturas (SANTOS et al., 2019; XI et al., 2021; YOU; DONG, 2020).

O *Moderate Resolute Imaging Spectroradiometer* (MODIS), é um sensor que está a bordo dos satélites TERRA e AQUA, com resolução temporal de um a dois dias e resolução espacial a partir de 250m, além de apresentar uma série temporal com mais de 24 anos (JUSTICE et al., 2002). Comparado aos sensores Sentinel-2/MSI (10m) e Landsat-8/OLI (30m), o seu detalhamento espacial é inferior, mas ainda pode ser compatível com regiões de estrutura fundiária baseada em médias e grandes propriedades, sobretudo em áreas de relevos mais planos. Além disso, o MODIS tem boa disponibilidade de dados históricos em uma mesma coleção, com excelente qualidade radiométrica e séries temporais longas e consistentes, e que cumpre as demandas de estudos ou programas de mapeamento em regiões com menores variações de tipos de uso e cobertura da terra, e em áreas de estudos extensas (CHEN et al., 2018; MASSEY et al., 2017; WANG et al., 2023) .

Sensores orbitais realizam a aquisição de novas imagens de sensoriamento remoto diariamente, ampliando a quantidade de dados disponíveis para o usuário. Com o aumento na disponibilidade de dados, cresce também o trabalho do usuário ao manipulá-los (SOILLE et al., 2018). Neste sentido, novas alternativas têm sido disponibilizadas à comunidade científica para

o acesso a coleções de imagens de satélite com alto nível de processamento, como o Google Earth Engine (GORELICK et al., 2017) e o projeto Brazil Data Cube - BDC (FERREIRA et al., 2020), facilitando o uso e a manipulação de longas séries temporais de imagens.

Uma série temporal é uma organização cronológica das imagens para sua análise integrada, podendo ser utilizada para a gerar perfis espectro-temporais ou de produtos mais complexos, derivados de classificações espectro-temporais. Os perfis temporais representam o comportamento espectral de uma unidade amostral da imagem, que pode ser um pixel ou uma porção do território composta por vários pixels. Um exemplo de aplicação das séries temporais das imagens do sensor MODIS para a geração de perfis espectro-temporais é o Sistema de Análise Temporal da Vegetação – SATVeg ([www.satveg.cnptia.embrapa.br](http://www.satveg.cnptia.embrapa.br)), um serviço da Embrapa (Empresa Brasileira de Pesquisa Agropecuária) que disponibiliza dados de índices vegetativos para toda a América do Sul (ESQUERDO et al., 2020).

A abordagem espectro-temporal explora o curto tempo de revisita de alguns sensores orbitais e visa a aquisição mais frequente de informações espectrais da superfície terrestre, trazendo vantagens em relação ao enfoque tradicional, que se baseia em um conjunto restrito de imagens. Na agricultura, a abordagem espectro-temporal pode apoiar a identificação de processos de intensificação agrícola (VIEIRA et al., 2022), o mapeamento de mudanças de tipos de uso e cobertura da terra (SIMOES et al., 2020), a estimativa de produtividade de culturas agrícolas (PIGNÈDE et al., 2021) e a estimativa de datas de colheita e semeadura de culturas agrícolas (ZHANG et al., 2021).

Na visualização dos perfis espectro-temporais são comumente utilizados os índices de vegetação, sendo os mais empregados o *Normalized Difference Vegetation Index* (NDVI) (ROUSE et al., 1974) e o *Enhanced Vegetation Index* (EVI) (HUETE et al., 1994), que apresentam alta correlação com variáveis biofísicas, como a biomassa verde e o índice de área foliar, além de permitirem interpretações referentes ao tipo de cultura agrícola, estágio fenológico e tipo de uso e cobertura da terra (JENSEN, 2011). Perfis espectro-temporais das bandas espectrais de um sensor, como os canais dos espectros do visível e infravermelho também são utilizados em conjunto com os índices de vegetação e auxiliam na interpretação dos dados (RUSSWURM et al., 2020). Porém, interpretar as séries temporais de índices ou bandas espectrais requer conhecimento prévio sobre o comportamento espectro-temporal dos alvos da superfície e suas variações ao longo do tempo. Devido a popularização da computação em nuvem e da geração de novos sensores, a disponibilidade das series temporais tem aumentado; entretanto, o uso, aplicação e conhecimento dos padrões espectro-temporais ficam limitados ao público especializado em sensoriamento remoto.

Aliado ao crescimento e a disponibilidade de grandes volumes de dados, os métodos de aprendizado de máquina passaram a receber destaque pela sua capacidade de lidar com um grande número de informações e aprender padrões (BATTA, 2018; SANTOSH; DAS; GHOSH, 2022). Devido a essa capacidade, algoritmos de classificação como o *Random Forest* (RF) estão presentes em diversos trabalhos relacionados ao sensoriamento remoto (WANG et al., 2023). Estudos apontaram que o RF é capaz de identificar os padrões presentes nas séries temporais e com boa acurácia (PELLETIER et al., 2016; SIMOES et al., 2021). Por outro lado, os métodos de aprendizado de máquina baseados em redes neurais artificiais profundas (aprendizado profundo) também têm ganhado destaque por apresentarem bom desempenho em lidar com dados de sensoriamento remoto, principalmente as séries temporais (KHAN et al., 2022; WEN et al., 2022). A rede *Transformer* (SAINTE FARE GARNOT et al., 2020) e a *Convolutional Neural Network* (CNN) (XI et al., 2021) são exemplos de redes neurais utilizadas para interpretar os padrões presentes em séries temporais.

No sensoriamento remoto, os métodos de aprendizado de máquina utilizam, comumente, amostras de campo, que representam os alvos (como os tipos de culturas agrícolas ou uso e cobertura da terra) a serem analisados junto às séries temporais, caracterizando o aprendizado supervisionado (BATTA, 2018). O uso do aprendizado de máquina aplicado à análise de séries temporais apresenta um grande potencial para as atividades de monitoramento do uso e cobertura da terra, em especial das áreas agrícolas. No entanto, a falta de amostras de campo limita a abrangência de estudos sobre o tema. O processo de aquisição de amostras de campo exige custos de deslocamento e mão de obra. No Brasil, a dificuldade é maior devido a extensão continental do país e a própria complexidade do seu território, que possui seis grandes biomas, variações edafoclimáticas e diferentes estruturas fundiárias (HAO et al., 2020; MA et al., 2024). Neste contexto, metodologias capazes de lidar com a escassez de amostras de campo contribuem para aumentar a abrangência de trabalhos na área de sensoriamento remoto.

Um modelo proveniente de um método de aprendizado de máquina, geralmente, é treinado e testado para realizar uma tarefa voltada para uma área de estudo alvo, e com um banco de dados específico. Neste sentido, criar um modelo capaz de se adaptar a diferentes conjuntos de dados é visto como um desafio para o aprendizado de máquina; em muitos casos, ocorre a perda de eficácia do modelo, gerando resultados imprecisos. (BLANCHARD; LEE; SCOTT, 2011; WANG et al., 2021; ZHOU et al., 2021). Em regiões com limitações de amostragem, estudos apontam que técnicas como o *Transfer Learning* (TL) e a Generalização são opções para aplicar um modelo em diferentes áreas sem haver perdas significativas de desempenho (CHAVES et al., 2021; MA et al., 2024; NASIRI et al., 2023).

O TL e a Generalização são vistos como técnicas semelhantes, principalmente porque exigem o treinamento de um modelo em uma região com alta disponibilidade de amostras; porém, diferenças podem ser encontradas em suas respectivas aplicações. No TL, amostras da região alvo são implementadas no modelo base, sendo utilizadas para a sua adaptação e melhoria. Estudos que adotam essa abordagem têm apontado bons resultados envolvendo classificações baseadas em anos-safras distintos (HU et al., 2022), mapeamento de culturas agrícola para regiões com características semelhantes (XU et al., 2020) e adaptação de modelo criado com base em amostras coletadas nos Estados Unidos e aplicados em regiões de produção agrícola na China (HAO et al., 2020).

A Generalização está relacionada à capacidade do modelo de ser aplicado em condições diferentes daquelas encontradas durante o treinamento, ou seja, à capacidade de adaptação do modelo em diferentes bancos de dados (BLANCHARD; LEE; SCOTT, 2011). Nesta abordagem, o usuário não define uma região alvo para adaptar o modelo, conseqüentemente, novas amostras não são implementadas. Ao contrário do TL, a Generalização busca criar um modelo base que represente bem as características analisadas dentro do conjunto de dados, de modo que o mesmo possa ser aplicado em quaisquer tarefas semelhantes (WANG et al., 2021; ZHOU et al., 2021). Neste sentido, a generalização do modelo possibilita maior alcance, e sem precisar de adaptações. Exemplos de aplicações são encontrados em trabalho que envolvem a classificação interanual de uma área de estudo (LUCIANO; CAMPAGNUCI; LE MAIRE, 2022), e a generalização para áreas de estudos amplas como culturas agrícolas nos Estados Unidos (JOHNSON, 2019) e o Cerrado brasileiro (CHAVES et al., 2021).

Neste contexto, a hipótese deste trabalho é que a partir da abordagem de Generalização aplicada em um modelo de classificação treinado com base em um período de ano-safra, é possível realizar a classificação interanual de diferentes períodos de ano-safra da série temporal do sensor MODIS.

## 2 Objetivos Gerais e Específicos

O presente trabalho teve como objetivo geral utilizar métodos de aprendizado de máquina para criar um modelo de classificação baseado no ano-safra 2019/2020, e efetuar a classificação interanual da série temporal do MODIS. Os objetivos específicos deste trabalho são:

- Testar e validar diferentes modelos de aprendizado de máquina para a identificação e classificação do uso e cobertura da terra em perfis espectro-temporais de imagens MODIS, com o objetivo de criar o modelo base no ano-safra 2019/2020;
- Testar e validar o modelo base para a classificação interanual de outros períodos de ano-safra (2016/17, 2020/21 e 2021/22) da série temporal do sensor MODIS, por meio da aplicação da técnica de Generalização com vistas ao desenvolvimento de um protocolo instantâneo de identificação de padrões de uso e cobertura da terra em séries temporais de imagens MODIS.

### 3 Revisão Bibliográfica

#### 3.1 Séries Temporais de Imagens de Satélites

Séries temporais são conjuntos de dados organizados e estruturados sequencialmente de forma cronológica para sua análise integrada. No sensoriamento remoto orbital são utilizados dados fornecidos por sensores que ficam a bordo de satélites, que registram a energia eletromagnética refletida pelos corpos na superfície da Terra (JENSEN, 2011). Os sensores, ao orbitarem a Terra, registram imagens de uma mesma área repetitivamente, de acordo com o seu tempo de revisita. Essa frequência de aquisição de imagens, que varia de sensor para sensor, dá origem às séries temporais, que podem ser utilizadas para expressar os diferentes padrões espectro-temporais dos alvos da superfície terrestre. Neste sentido, FONSECA et al. (2021) revisaram o processo de classificação de uso da terra para o Cerrado e destacaram a importância de reconhecer e utilizar os padrões presentes nas séries temporais, além do uso deste tipo de dado para tarefas que envolvem a análise da dinâmica territorial ao longo do tempo.

Nas últimas décadas, a quantidade de dados obtidos por meio do sensoriamento remoto tem sido cada vez maior, resultando em séries temporais mais longas. Em 2018, SOILLE et al. (2018) observaram e estimaram um crescimento no volume de dados obtidos pelos sensores orbitais, principalmente, na coleção Sentinel-2/MSI com um volume de armazenamento na ordem de grandeza de *petabytes*, valor muito superior se comparado às coleções do MODIS ou Landsat-8/OLI. Neste contexto, usuários que buscam extrair séries temporais de imagens acabam enfrentando limitações de processamento e armazenamento, reforçando a importância da disponibilidade das imagens de satélites em repositórios online.

No Brasil, o projeto BDC, conforme apresentado por FERREIRA et al. (2020) disponibiliza gratuitamente coleções de diversos sensores no formato de cubo de dados, simplificando o processo de extração das séries temporais. Cubos de dados (ou empilhamento de imagens), são estruturas de imagens de sensoriamento remoto que contém todas as datas disponíveis para os períodos analisados. Nos cubos de dados, as imagens de satélite devem apresentar as mesmas dimensões (linhas e colunas), a mesma abrangência geográfica, ter o mesmo número de atributos (índices de vegetação e bandas), estar organizadas em função do tempo e não haver quebra da série temporal (FERREIRA et al., 2020). O website do projeto BDC ([www.brazildatacube.org](http://www.brazildatacube.org)) possibilita a visualização de séries temporais e o acesso aos cubos de dados de sensores como o MODIS, Landsat-8/OLI, Sentinel-2/MSI e CBERS/WFI.

Os perfis temporais, quando organizados, armazenados e disponibilizados em bancos de dados geográficos para acesso instantâneo, podem prover informações sobre o uso e cobertura da terra em um determinado ponto da superfície terrestre ao longo de um período específico, conforme demonstraram ESQUERDO et al. (2020) ao apresentarem o Sistema de Análise Temporal da Vegetação (SATVeg), uma ferramenta gratuita para visualização de perfis temporais de dados MODIS em toda a América do Sul. Este tipo de informação pode apoiar a identificação de processos de supressão da vegetação, intensificação agrícola, estimativa das datas de semeadura e colheita, transições entre sistemas pecuários para a agricultura, entre outros.

### **3.1.1 Índices de Vegetação**

As séries temporais obtidas pelo sensoriamento remoto são formadas por bandas espectrais ou por produtos derivados, como os índices de vegetação. Cada alvo interage de forma diferente com as diferentes regiões do espectro eletromagnético, por isso as informações presentes nas bandas espectrais podem ser utilizadas para a identificação de padrões. Por exemplo, RUSSWURM et al. (2020) e ZHAO et al. (2021) utilizaram as bandas espectrais presentes no Sentinel-2/MSI para uma tarefa de classificação de uso e cobertura da terra, envolvendo culturas agrícolas como: milho, algodão e agricultura perene. ZHAO et al. (2021) realizaram uma análise exploratória referente aos perfis temporais das bandas espectrais, e destacaram que as bandas NIR, RED e Red Edge (RE) permitiram maior caracterização de classes de agricultura como milho e algodão em relação as bandas BLUE e GREEN.

Na agricultura, para facilitar a interpretação das séries e sintetizar as informações presentes nas bandas, é comum o uso de índices de vegetação, como o NDVI (ROUSE et al., 1974). O cálculo do NDVI utiliza as bandas do vermelho (região visível do espectro, onde acontece alta absorção pelos pigmentos da folha) e o infravermelho próximo (região do espectro onde a radiação é refletida pela planta) (JENSEN, 2011). Por utilizar duas bandas que apresentam comportamentos distintos ao interagirem com a planta, o NDVI é comumente utilizado para interpretações referentes aos estágios fenológicos e o vigor da cultura, por isso diversos estudos que envolvem o sensoriamento remoto usam como base o NDVI para realizar interpretações sobre uma área de estudo (FERCHICHI et al., 2022; HUANG et al., 2021).

Apesar da sua praticidade de aplicação os valores de NDVI podem ser afetados por efeitos atmosféricos (HUANG et al., 2021). Nesse contexto, outros índices têm sido propostos visando aprimorar o NDVI como, por exemplo, o EVI (HUETE et al., 1995), que visa corrigir interferências atmosféricas que afetam o NDVI. Em seu trabalho de revisão referente aos índices de vegetação, XUE e SU (2017) avaliaram diferentes tipos de índices presentes na literatura, apontando suas vantagens e possíveis limitações. Existe uma ampla disponibilidade de índices, cada um adaptando-se melhor a uma determinada análise. Nesse sentido, a escolha de um índice dependerá da demanda específica do usuário. Por exemplo, o NDVI e EVI apresentam alta correlação com variáveis biofísicas, como a biomassa verde e o índice de área foliar, razão pela qual são comumente utilizados em conjunto para aplicações voltadas para o monitoramento agrícola (PIGNÈDE et al., 2021; XUE; SU, 2017).

Em seu trabalho, HUANG et al. (2021) realizaram uma revisão referente ao uso do NDVI na atualidade e destacaram que, apesar das limitações, o NDVI continua sendo o índice mais conhecido e utilizado. No entanto, reforçaram a possibilidade de combinar diferentes índices ou bandas espectrais para obter maior quantidade de informações. Neste contexto, é comum encontrar trabalhos cujas metodologias incluem o uso de dois ou mais índices de vegetação junto às bandas espectrais do sensor. Por exemplo, LUCIANO et al. (2019) utilizaram quatro bandas espectrais do Landsat-8/OLI e seis diferentes índices de vegetação, dentre eles NDVI e EVI, para uma tarefa de classificação de culturas de cana-de-açúcar. Além disso, os autores realizaram uma análise para verificar quais índices foram os mais representativos. Assim, constataram que a banda SWIR (1.55 – 2.55  $\mu\text{m}$ ) e a banda RED foram as mais importantes na tarefa realizada. Essa análise é cientificamente relevante, pois auxilia outros trabalhos interessados no uso das bandas espectrais em conjunto com os índices de vegetação, além de exemplificar o processo de seleção de atributos relacionados às bandas espectrais.

### **3.1.2 Sensor MODIS**

O MODIS é um produto da *National Aeronautics and Space Administration* (NASA), que está a bordo dos satélites TERRA e AQUA. O TERRA foi lançado em 1999 e possui uma órbita descendente polar sincronizada, temporizada para atravessar o equador aproximadamente às 10:30 da manhã, em horário local, enquanto o AQUA foi lançado em 2002 e possui uma órbita ascendente polar temporizada para atravessar o equador aproximadamente às 13:30 da tarde, ambos apresentam resolução temporal de um a dois dias (JUSTICE et al.,

2002). O produto de índices vegetativos do MODIS está presente nas coleções MOD13Q1 (TERRA) e MYD13Q1 (AQUA), que incluem composição de imagens de 16 dias, com resolução espacial de 250m e correções geométricas e atmosféricas.

Imagens obtidas pelo sensor MODIS apresentam resolução espacial de 250m, sendo compatível para mapeamentos de áreas mais extensas e estrutura fundiária baseadas em monoculturas. Por exemplo, CHEN et al. (2018) mapearam e identificaram padrões nos perfis espectro-temporais do MODIS relacionados a padrões de cultivo no Estado de Mato Grosso, como: soja-milho, soja-pasto, soja-algodão e agricultura temporária de um ciclo. Nesta linha de pesquisa, SILVA JUNIOR et al. (2020) utilizaram o MODIS para realizar um mapeamento da área plantada de soja na região Centro-Oeste do Brasil. ZANG et al. (2020) e XIONG et al. (2017) também optaram pelo uso do MODIS, uma vez que seus trabalhos envolviam o monitoramento de áreas extensas de produção agrícola em regiões da China e África, respectivamente. Já (PELLETIER et al. (2016) e SANTOS et al. (2019) concluíram que no caso de processos de monitoramento em áreas com a presença de diversas culturas, em uma estrutura fundiária mais segmentada e baseada em pequenas ou médias propriedades, sensores com maior resolução espacial, como Sentinel-2/MSI ou o Landsat-8/OLI, são mais adequados.

Considerando o uso integrado das coleções TERRA e AQUA, o MODIS disponibiliza uma série temporal de aproximadamente 22 anos. Dessa forma, o MODIS é predominante em trabalhos que necessitam um estudo aprofundado da série histórica da região (WANG et al., 2023). Neste contexto, PEREIRA et al. (2018) utilizaram o MODIS, e monitoraram mudanças no uso da terra em todo Cerrado brasileiro, no período de 2001 até 2017, e com ênfase no mapeamento de regiões de pastos degradados. Os autores destacaram a importância da série temporal extensa para avaliar possíveis anomalias nos perfis de NDVI, alertando que o processo de degradação ocorre de forma gradual ao longo dos anos. MASSEY et al. (2017) estudaram o mapeamento de culturas agrícolas ao longo do território americano, com destaque nas regiões central e leste, onde estão concentradas as regiões de produção agrícola do país. Os autores optaram pelo uso do MODIS pois o trabalho apresentava características compatíveis com o sensor: área de estudo extensa, mapeamento de monoculturas como soja, milho e trigo e a classificação interanual envolvendo os anos de 2001 até 2004.

### 3.1.3 Aplicações das Séries Temporais

Trabalhos que adotam a abordagem *'time-first, space-later'*, em que a análise temporal é mais relevante, têm apresentado bons resultados para mapeamento e detecção de mudanças no uso e cobertura da terra (CHAVES et al., 2021; CHEN et al., 2018; SIMOES et al., 2021). Por exemplo, PICOLI et al. (2018), buscaram aproveitar a grande disponibilidade de informações presentes nas séries temporais do MODIS e propuseram um modelo de classificação de uso e cobertura da terra. Os autores obtiveram sucesso em classificar classes de vegetação e culturas anuais de dois ciclos como soja-milho e soja-algodão. Além disso, mostraram a utilidade das séries temporais em analisar a vegetação em diferentes escalas de tempo, identificando processos como a intensificação agrícola e o desflorestamento.

Para extrair e sintetizar as informações presentes nas séries temporais de imagens de satélite, diversos trabalhos aplicam métodos de aprendizado de máquina e obtêm resultados satisfatórios em suas respectivas aplicações (BROWN et al., 2013; MIRALLES et al., 2023; ZHAO et al., 2021). Por exemplo, BENDINI et al. (2019) abordaram a metodologia de extração de variáveis fenológicas e sua utilização em classificadores de RF. Os autores destacaram a quantidade de informações presentes nos perfis do EVI e a possibilidade de utilizar parâmetros derivados das séries temporais como base para a classificação de uso e cobertura da terra.

No Brasil, órgãos como a CONAB fornecem ao longo dos anos relatórios referentes a estimativa de safra para verificar o andamento da produção agrícola do país. As informações fornecidas por esses órgãos auxiliam no planejamento logístico, desde o armazenamento até o transporte, e no planejamento financeiro referente a política de preços. Nesse contexto, estudos que visam o monitoramento de áreas agrícolas são de grande importância, pois apresentam metodologias detalhadas de monitoramento em diferentes regiões do Brasil. Resultados e métodos obtidos através desses estudos podem ser úteis em outras pesquisas, bem como para órgãos públicos (Ex: CONAB) que utilizam o sensoriamento remoto em escalas municipal, regional ou nacional. Por exemplo, WANG et al. (2023) revisaram e ressaltaram a importância dos produtos gerados por mapeamentos ao redor do mundo, especialmente em aplicações como monitoramento do uso e cobertura da terra em regiões de diferentes escalas, controle do desmatamento e observação dos impactos das mudanças climáticas.

O Mato Grosso é reconhecido como uma região de intensa produção e expansão agrícola, tornando-se alvo de numerosos estudos de monitoramento. Essa área é frequentemente referida como uma fronteira agrícola devido à sua significativa expansão, sendo objeto de investigações de monitoramento devido à necessidade de um crescimento sustentável e às

mudanças na dinâmica da produção agrícola. SIMOES et al. (2020) utilizaram séries temporais de NDVI, EVI e da banda do infravermelho do MODIS para realizar um estudo de monitoramento no estado de Mato Grosso, observando a expansão agrícola no Cerrado. Os autores utilizaram séries temporais e os resultados foram mapas e perfis temporais de diferentes anos que permitiram visualizar a expansão. Na região do nordeste brasileiro, DA SILVA et al. (2022) apresentaram uma metodologia baseada em séries temporais para monitorar áreas de produção com pivô-central, dando ênfase ao índice MNDWI (XU, 2006), uma variação do índice de NDVI que apresenta maior correlação com variáveis como a umidade do solo. No Estado de São Paulo, LUCIANO et al. (2018, 2019) utilizaram as séries do Landsat-8/OLI para mapear as regiões produtoras de cana-de-açúcar em diferentes safras, caracterizando a distribuição da produção por meio de mapeamento e comparando com dados de fontes, como o Instituto Brasileiro de Geografia e Estatística (IBGE) e mapeamento do projeto Canasat.

No Cerrado, extensas áreas de pastagem são comuns e representam um tipo de vegetação nativa do bioma. Nesse contexto, estudos que abordam o monitoramento de pastagens também contribuem para a avaliação da preservação ambiental da região (MÜLLER et al., 2015; PEREIRA et al., 2018). Por exemplo, JAKIMOW et al. (2018) utilizaram séries temporais das coleções Landsat 7 e 8 para monitorar práticas de manejo que envolviam pastagens em conjunto com cultivos agrícolas, assim como pastos que foram submetidos a queimadas. Manejos complexos também foram mapeados com o auxílio de séries temporais. BUENO et al. (2023) e TORO et al. (2022) realizaram o mapeamento de áreas com sistemas de integração lavoura-pecuária (ILP), um manejo complexo que envolve a rotação entre culturas agrícolas e pastagem, voltado para a criação animal. Através de um estudo detalhado das séries temporais e aliado ao uso de métodos de aprendizado de máquina como o RF, os autores obtiveram sucesso no mapeamento de uma estrutura complexa como a ILP, e que foi mapeada com valores de acurácia global acima de 0,80.

### **3.2 Métodos de Aprendizado de Máquina**

O aprendizado de máquina é um ramo dentro do campo da inteligência artificial e que tem como base o conceito de ensinar a máquina a realizar tarefas por meio do aprendizado de variáveis e padrões em um conjunto de dados volumoso, tarefas de alta dificuldade para um humano realizar sem ajuda da máquina (BATTA, 2018; JANIESCH; ZSCHECH; HEINRICH, 2021). No sensoriamento remoto, uma das aplicações do aprendizado de máquina são os algoritmos voltados para a classificação de imagens, permitindo a elaboração de mapas de uso

e cobertura da terra, cujas classes temáticas podem descrever, de maneira simplificada, o território em questão (HAO et al., 2015; MIRALLES et al., 2023).

Existem diversos tipos de algoritmos que podem ser utilizados na atividade de classificação de dados, sejam tabulares ou na forma de matrizes (imagens). Exemplos de bibliotecas computacionais que permitem o uso de técnicas de aprendizado de máquina são os diversos pacotes disponíveis na linguagem R, o *Scikit-learn* na linguagem Python e o *software* WEKA. Na agricultura, é comum o uso do aprendizado supervisionado, onde amostras de campo são implementados no banco de dados com o objetivo de mostrar ao classificador as respostas esperadas pelo usuário (SANTOS et al., 2021a).

Um dos métodos mais populares é a árvore de decisão. O algoritmo recebe esse nome por ser baseado no processo de tomada de decisão, utilizando uma estrutura de regras ou condições que direcionam aos possíveis resultados obtidos pelo algoritmo (JAMES et al., 2013). A árvore de decisão inicia em um nó, referente a variável que melhor representa o conjunto de amostras, a partir do qual são derivadas diferentes ramificações que representam os caminhos para as possíveis respostas do algoritmo (Ex: Classes temáticas de uso e cobertura da terra). No entanto, o processo de construção de uma árvore de decisão está sujeito ao *overfitting*, ou seja, quando a árvore se adapta a apenas um conjunto de dados. Neste sentido, BREIMAN et al. (2001) propôs o algoritmo RF, com intuito de melhorar o conceito de árvores de decisão.

Apesar de ter sido proposto décadas atrás, o RF continua sendo um dos principais algoritmos voltados para a classificação por aprendizagem de máquina, devido a sua capacidade preditiva que supera algoritmos criados posteriormente (BELGIU; DRĂGU, 2016; PELLETIER et al., 2016). O RF funciona com base num conjunto de classificadores baseados em árvores de decisão, junto a uma amostragem *Bootstrap*, em que diversos subconjuntos derivados do conjunto original são criados e aplicados em árvores de decisão. O RF leva em consideração a decisão da maioria das árvores e, por isso, apresenta resultados relevantes em diversos trabalhos que envolvem a classificação de uma determinada variável. Além disso, a sua implementação é simplificada, pois o algoritmo trabalha com dois hiperparâmetros principais: o *n*tree, variável referente ao número de árvores criadas pelo algoritmo, e o *m*try, que está relacionado ao número de variáveis utilizadas em cada árvore.

Por se tratar de um algoritmo que utiliza amostragem aleatória, o mesmo é capaz de se adaptar a bancos de dados que apresentem dados faltantes e ruídos, além de apresentar baixa tendência ao *overfitting* (JAMES et al., 2013). Trabalhos como os de PELLETIER et al. (2016) exemplificam o uso dos métodos de aprendizado de máquina no sensoriamento remoto. Neste estudo, os autores verificaram a capacidade do RF em realizar a classificação do uso e

cobertura da terra, mesmo em banco de dados extensos e com elevado número de classes. Os resultados obtidos mostraram que o RF foi capaz de classificar mais de 10 classes temáticas de agricultura e vegetação, e com acurácia global acima de 0,80. Em trabalhos cujo o foco é a classificação de culturas agrícolas, o RF também apresentou bons resultados. Por exemplo, SANTOS et al. (2019) utilizaram o RF para classificar uma área no estado de São Paulo, onde existiam diversos tipos de manejos e culturas agrícolas diferentes como pastagem, solo exposto, soja, milho, algodão e trigo. Os autores utilizaram 14 atributos diferentes envolvendo índices de vegetação como o NDVI e EVI, e as bandas espectrais RED, SWIR1(1.55-1.75 $\mu$ m) e SWIR2 (2.09-2.35 $\mu$ m) dos sensores Landsat 7/ETM+ e Landsat 8/OLI. Apesar da grande quantidade de atributos e classes temáticas, o RF apresentou acurácia global 0,84.

Além de classificação de culturas, o RF também apresentou bons resultados em trabalhos com perfis espectrais complexos. TORO et al. (2022) mapearam sistemas de ILP com acurácia global de 0,86, resultado que superou métodos baseados em redes neurais artificiais. O projeto MapBiomas, apresentado por SOUZA et al. (2020), teve como objetivo realizar um mapeamento de uso e cobertura da terra em todo o país, empregando diferentes níveis de classificação. O algoritmo escolhido para esse projeto foi o RF. Apesar da complexidade da tarefa, o MapBiomas apresentando valor de acurácia global de 0,89. LUCIANO; CAMPAGNUCI; LE MAIRE (2022) empregaram o algoritmo RF em uma tarefa de Generalização para o mapeamento de uma série temporal de 33 anos de cana-de-açúcar. Neste estudo, o RF foi capaz de aprender os padrões espectro-temporais da cultura de cana-de-açúcar, o que permitiu que o modelo fosse aplicado numa série temporal extensa.

Apesar de ser amplamente utilizado em tarefas relacionadas ao sensoriamento remoto, o RF é um algoritmo que não considera variações espaciais e temporais do conjunto de dados e, dessa forma, estudos buscam compreender o funcionamento do algoritmo. Em seu trabalho, MILLARD; RICHARDSON, (2015) apontaram que a acurácia do RF pode variar devido à baixa representatividade e distribuição das amostras de treino na área de estudo, principalmente em tarefas que envolvam o mapeamento. BELGIU; DRĂGU, (2016) também revisaram o funcionamento do RF e destacaram a necessidade de cuidados com as amostras de treino a serem utilizadas pelo algoritmo.

No sensoriamento remoto, o problema com dados faltantes é comum, por isso diversos trabalhos buscam formas eficientes de realizar o preenchimento desses espaços (VUOLO; NG; ATZBERGER, 2017). Neste contexto, WANG et al. (2022) propuseram uma metodologia de preenchimento de imagens contaminadas por nuvens do Landsat-8/OLI e Sentinel-2/MSI. O modelo proposto foi baseado no RF, devido a capacidade do algoritmo em

trabalhar com diferentes variáveis. Além disso, demais aplicações podem ser encontradas para o RF como, por exemplo, o pacote *missForest* (STEKHOVEN et al., 2012), escrito na linguagem R, que realiza o preenchimento de dados faltantes em séries temporais, baseando-se no RF e na sua aplicabilidade em tarefas de regressão.

### 3.3 Aprendizado Profundo

O aprendizado profundo é um ramo que pertence ao aprendizado de máquina, mas com algumas diferenças. Ambos conceitos buscam fazer com que a máquina aprenda uma tarefa, baseando-se num conjunto de amostras de treino. Porém, no aprendizado profundo, ao invés de métodos baseados em estatística, são utilizadas as redes neurais artificiais (JANIESCH; ZSCHECH; HEINRICH, 2021). Geralmente, a arquitetura base de uma rede neural é composta por camadas de neurônios conectados entre si, onde os dados de saída de uma camada são os dados de entrada da camada seguinte (SONY et al., 2021). A primeira camada (*input*) são as amostras de treino, e a última é o resultado esperado (*output*) que, para o sensoriamento remoto, comumente, é o resultado da classificação de uso e cobertura da terra. As camadas intermediárias, também chamadas de camadas ocultas, são compostas por uma função de ativação, vetor peso e o viés; esses parâmetros são os responsáveis pelos resultados obtidos pela rede (KATTENBORN et al., 2021).

O aprendizado de redes neurais profundas aplica o conceito chamado de *Back-Propagation*, onde informações utilizadas na primeira interação são utilizadas nas demais. Por se tratar de um método supervisionado, os resultados de cada interação são comparados aos valores reais de referência; na interação seguinte, a rede realiza ajustes nos vetores peso e viés para diminuir a porcentagem de erros na classificação. Por meio desse processo, essas redes são capazes de se ajustarem e aprenderem uma tarefa por conta própria, por isso muitos usuários têm adotado o uso dessas redes ao invés dos métodos estatísticos tradicionais (ISMAIL FAWAZ et al., 2019; KATTENBORN et al., 2021; SANTOSH; DAS; GHOSH, 2022).

As redes neurais, inicialmente, foram criadas para tarefas como interpretações de imagens ou textos. Além disso, são conhecidas por serem capazes de processar uma quantidade massiva de dados e, em muitos casos, superarem algoritmos de aprendizado de máquina (ISMAIL FAWAZ et al., 2019). Devido à capacidade preditiva e ao aumento da disponibilidade de dados, assim como à necessidade por resultados cada vez mais precisos, diferentes arquiteturas baseadas em redes neurais profundas têm sido aplicadas em tarefas de sensoriamento remoto (MA et al., 2019). Por exemplo, FERCHICHI et al. (2022) e WANG et

al. (2023) revisaram o uso das redes neurais profundas no mapeamento de uso e cobertura da terra utilizando dados de sensoriamento remoto. Devido a capacidade de interpretar características dos perfis espectro-temporais das imagens de satélite, os autores destacaram o potencial das redes neurais profundas como a *Convolutional Neural Network* (CNN). No entanto, alertaram que, apesar dos resultados com boas acurácias e capacidade de aprender padrões, as redes neurais profundas apresentam limitações, como a necessidade de um extenso banco de dados, demanda por capacidade de processamento da máquina e maior dificuldade no treinamento dos modelos.

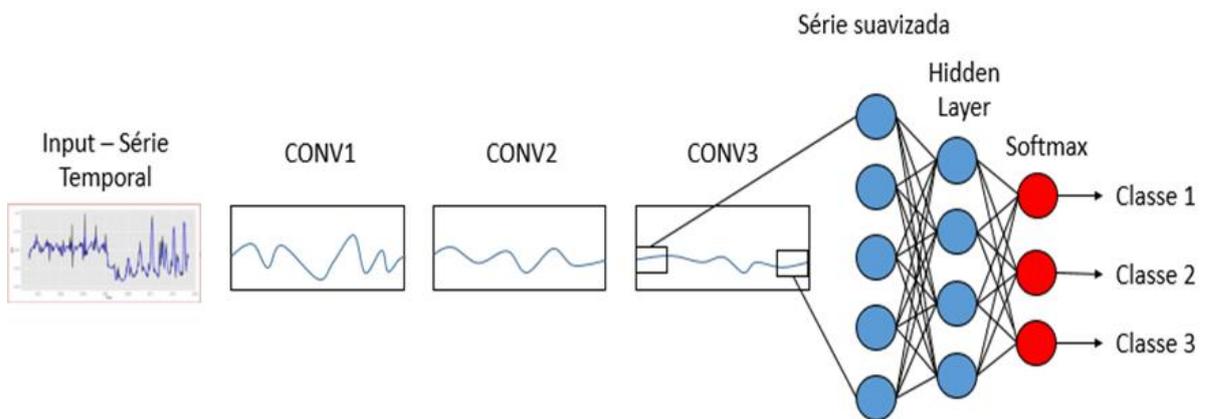
### 3.3.1 Tipos de Redes Neurais

A CNN é uma rede neural profunda inicialmente proposta para a interpretação de imagens ou matrizes de dados, estrutura semelhante das imagens obtidas por sensores orbitais. Por isso, na agricultura, uma das arquiteturas mais utilizadas é a CNN. A rede recebe esse nome, pois apresenta um filtro convolucional, que suaviza os dados de entrada para diminuir a redundância de informações, simplificando o aprendizado da rede (SANTOSH; DAS; GHOSH, 2022). No sensoriamento remoto, a CNN apresenta variações e dependem da demanda do usuário para seu uso. As redes 1D-CNN são utilizadas para análises com dados sequenciais, como as series temporais. Já a 2D-CNN é utilizada para a extração de características espaciais e a 3D-CNN utiliza as abordagens temporal e espacial (KHAN et al., 2022; LIAO et al., 2020).

Os autores DEBELLA-GILO; GJERTSEN (2021), utilizaram as redes 1D-CNN e 2D-CNN para classificar e mapear classes de cereais e gramíneas. Os resultados de acurácia global obtidos por ambas as redes foram semelhantes (0,93) porém apresentaram melhores resultados do que a *Multi-Layer-Perceptron* (MLP) (0,89), uma rede neural com arquitetura mais simples. FIRAT et al. (2022) utilizaram a 3D-CNN para classificar imagens e como resultado na tarefa de classificação obtiveram valores de acurácia global de 0,91 em três conjuntos diferentes de amostras de teste. Além disso, destacaram a importância do filtro convolucional para reduzir a quantidade de informações a serem processadas, e melhorar o desempenho da classificação.

Com o objetivo de melhorar o desempenho da 1D-CNN em séries temporais, PELLETIER; WEBB; PETITJEAN (2019) propuseram uma variação chamada de *Temporal Convolutional Neural Network* (TempCNN). A arquitetura da TempCNN é baseada no domínio temporal dos dados, fazendo com que a rede leve em consideração aspectos temporais do perfil. Além disso, as séries temporais passam por três filtros convolucionais, que simplificam a

informação passada através das camadas, diminuindo a quantidade de pesos e viés ao longo da rede (Figura 1). Após passar pelos filtros, as séries suavizadas são passadas por um conjunto de camadas ocultas, que atribuem a série temporal a uma classe temática. Os autores apresentaram que a TempCNN em uma tarefa de classificação de uso e cobertura da terra, e compararam seus resultados com algoritmos como o RF e a 1D-CNN nas métricas de avaliação *FI-Score*, *Precision* e *Recall*. A classificação de uso e cobertura da terra envolveu 13 classes temáticas e a TempCNN apresentou desempenho superior aos demais classificadores em 11 delas.



**Figura 1:** Ilustração de arquitetura da rede TempCNN. Adaptado de PELLETIER; WEBB; PETITJEAN (2019).

Outro tipo de rede neural que também é comumente utilizada no sensoriamento remoto é a *Recurrent Neural Network* (RNN). Diferente da CNN, a RNN tem como principal característica a capacidade de “lembrar” informações de maior importância para a classificação da série. A RNN, inicialmente, foi proposta para o trabalho com sequências de dados, por isso, no sensoriamento remoto ela é utilizada em tarefas que utilizam as séries temporais (KATTENBORN et al., 2021). Exemplos de RNN são a *Long Short Term Memory* (LSTM) e a *Transformer*. A LSTM é uma rede que tem como característica a capacidade de armazenar padrões temporais e utilizá-los para a classificação de demais períodos da série. XI et al. (2021) exemplificaram e detalharam o uso da LSTM na tarefa de classificação usando séries temporais. Os autores utilizaram como base de comparação métodos de aprendizado de máquina como o RF e, ao comparar os resultados obtidos com a LSTM, os mesmos verificaram que a rede obteve um resultado de acurácia global de 0,84 para a classificação de classes temáticas de vegetação florestal. Apesar dos resultados, os autores destacaram que mais estudos são necessários para aplicar a rede em tarefas com perfis espectro-temporais.

Algoritmos como a TempCNN e o RF foram utilizados no trabalho de ABIDI et al. (2023), que comparou diversos algoritmos diferentes para compara-los em uma tarefa de

classificação de uso e cobertura da terra. Ambos apresentaram acurácia global semelhantes entre si (0,85), superando abordagens baseadas na Resnet (0,80), outra rede comumente utilizada pela comunidade. Em metodologia semelhante, ZHONG; HU; ZHOU (2019) também compararam o desempenho de diversos classificadores, incluindo o RF, a 1D-CNN e a LSTM. Os autores utilizam um conjunto denso de amostras de treino, contendo 14 classes temáticas de culturas agrícolas. Os três algoritmos obtiveram resultados em que a 1D-CNN apresentou maior valor de acurácia global (0,86).

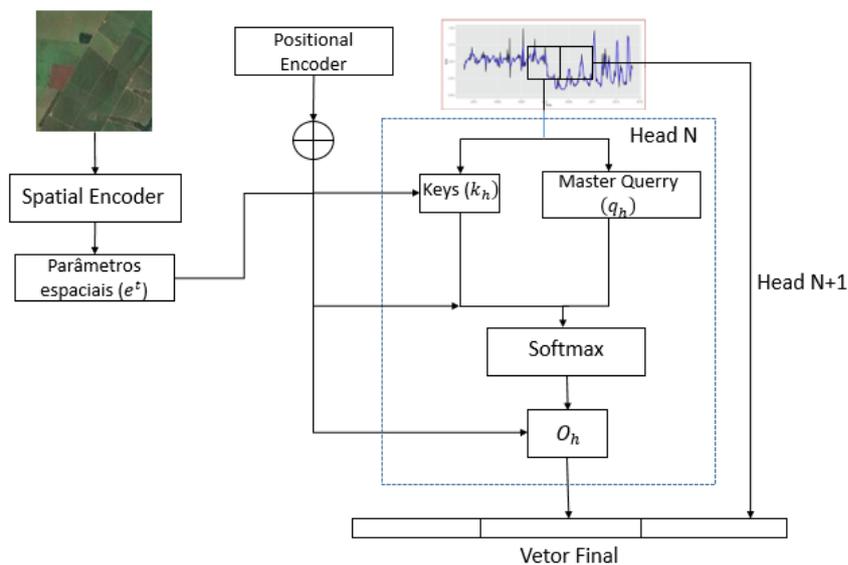
### 3.3.2 Transformer

A *Transformer*, uma variação da RNN que utiliza o conceito *Attention*, foi proposta inicialmente por VELTMAN; PULLE; DE DONCKER (2017). Na arquitetura da *Transformer* são realizadas três etapas anteriores ao processamento dos dados nas camadas ocultas e que a diferenciam dos demais métodos. A primeira é a *Attention*, na qual os parâmetros chamados de *Keys*, *Value* e *Query* são computados e têm como função armazenar as informações mais relevantes de uma sequência de dados. Por se tratar de uma sequência, todos os dados passam por uma função chamada *Positional Encoding*, que ajuda a rede a organizar e interpretar a continuidade da sequência analisada. Por fim, a *Transformer* tem a vantagem de realizar o aprendizado em paralelo, permitindo que diferentes padrões referentes a períodos diferentes da série sejam computados ao mesmo tempo, num processo chamado de *Multi-Head-Attention* (VELTMAN; PULLE; DE DONCKER, 2017). Além disso, assim como a LSTM, a *Transformer* também armazena parâmetros para serem utilizados ao longo da rede, permitindo que as informações relevantes da série sejam utilizadas durante todo o processo de classificação, chamado de *Self-Attention*, e é mais uma etapa que a diferencia das demais redes.

A arquitetura original da *Transformer* foi utilizada por RUSSWURM et al. (2020) na tarefa de identificação de culturas. Os autores verificaram que a *Transformer* obteve melhores resultados que a LSTM, mas destacaram a complexidade em aplicá-los. Neste sentido, GARNOT; LANDRIEU (2020), introduziram a rede *Lightweight Temporal Attention Encoder* (L-TAE), que tem como proposta adaptar a *Transformer* para a análise de padrões espectrais de séries temporais. A L-TAE, inicialmente, divide a série em N períodos (*Heads*), e para cada um deles, os parâmetros *Keys* ( $k_h$ ) e *Querrys* ( $q_h$ ) são computados. Na L-TAE é utilizada apenas uma variável chamada de *Master Query*, que corresponde à média de todas as *Querrys* calculadas nos N períodos. Os parâmetros *Keys* utilizam informações espaciais ( $e^t$ ) da imagem e são calculados em paralelo e organizados por meio da *Positional Encoding*. Os parâmetros

são normalizados pela função *softmax*, e o *output* ( $o_h$ ) é utilizado para construção de um vetor de mesmo tamanho da série temporal, que em seguida passa por uma série de camadas ocultas, responsáveis pela classificação.

Um exemplo de arquitetura da rede L-TAE é apresentado na Figura 2. GARNOT; LANDRIEU (2020), adaptaram e simplificaram o uso da Transformer e destacaram o menor gasto computacional e rapidez na implementação do algoritmo proposto. Além disso, a L-TAE apresentou acurácia de 0,94 ao classificar séries temporais de imagens de satélite, superando a arquitetura original da *Transformer*.



**Figura 2:** Ilustração de arquitetura da rede L-TAE. Adaptado de GARNOT; LANDRIEU (2020).

### 3.4 Amostras de Campo e Qualidade do Conjunto de Dados

De acordo com DANTAS CHAVES; DE CARVALHO ALVES (2019), a tarefa de mapeamento utilizando imagens de sensoriamento remoto apresenta etapas como: seleção da coleção de imagens e construção da série temporal, caracterização da área de estudo e coleta de dados de campo. A construção da série temporal pode ser feita por meio de linguagens, como o R, que apresentam bibliotecas próprias para o processamento das imagens de satélites, como exemplo o pacote *Satellite Image Time Series Analysis for Earth Observation Data Cubes (sits)* apresentado por SIMOES et al. (2021). Além disso, CHEN et al. (2021) destacaram a importância da construção das séries temporais, pois alguns conjuntos podem apresentar dados faltantes, prejudicando a interpretação da série. Neste sentido, a organização das imagens de satélites no formato de cubo de dados tem relevância para facilitar a interpretação das séries temporais.

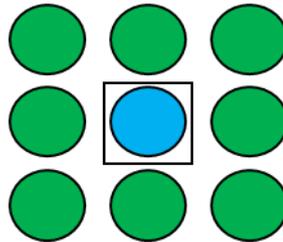
As amostras de campo devem ser coletadas *in situ* e funcionam como referência para os algoritmos de aprendizado de máquina, mostrando quais os resultados esperados pelo usuário durante a tarefa de classificação (BATTA, 2018). Os autores SANTOS et al. (2021) e SIMOES et al. (2021) destacaram que a qualidade dos dados de campo exerce alta influência no desempenho dos classificadores. Eles apontaram que as amostras que apresentam erros são chamadas de *outliers* ou ruídos e sua presença induz os classificadores ao erro. As inconsistências nas amostras de campo ocorrem, principalmente, devido ao erro humano na interpretação das informações em campo, como a identificação de uma cultura ou manejo do solo, e na dificuldade em interpretar zonas de transição entre culturas.

Neste sentido, a análise exploratória dos dados representa uma etapa importante no treinamento de modelos de aprendizado de máquina. Organização dos dados, visualização em gráficos e identificação de padrões podem ajudar a identificar possíveis ruídos no conjunto (JAMES et al., 2013). MILLARD; RICHARDSON (2015) destacaram a necessidade de análise exploratória dos dados, uma vez que os erros na aquisição de amostras de campo podem influenciar na etapa de treino do RF. No aprendizado profundo, ruídos presentes nas amostras também interferem a etapa de treino do algoritmo. ZHAO et al. (2021) testaram cinco algoritmos diferentes baseados em redes neurais, incluindo redes com alto potencial de aprendizado, como a LSTM e a 1D-CNN. Os autores utilizaram séries temporais com presença de dados faltantes causados por nuvens e enfatizaram a importância da caracterização das amostras antes de utilizá-las na etapa de treinamento.

O algoritmo *Self-Organizing-Maps* (SOM), inicialmente proposto por KOHONEN (1990), foi utilizado por SANTOS et al. (2021a, 2021b) com o objetivo de diminuir a influência das *outliers* nos respectivos conjuntos de dados. A arquitetura do SOM utiliza o conceito de *Competitive Learning* para agrupar as amostras por meio do aprendizado não supervisionado (WANG et al., 2023). A arquitetura do algoritmo consiste de uma rede neural que mapeia um conjunto de dados de entrada de alta dimensão para uma de baixa, tendo como resultado um mapa bidimensional, cuja característica é o agrupamento de dados com padrões semelhantes. O mapa gerado é composto por unidades chamadas de neurônios. Os neurônios, inicialmente, são criados com valores ao acaso, e treinados junto com a rede neural por meio do *Back-Propagation* até atingirem um ponto de parada, definido pelo usuário.

Ao final de cada interação, cada série temporal deve ser atribuída a um neurônio. Os neurônios pertencem a uma classe temática e podem conter mais de uma série temporal, desde que estas apresentem comportamento espectral semelhante. No final do processo, neurônios semelhantes são agrupados dentro do mapa bidimensional (SANTOS et al., 2021b).

O agrupamento dos neurônios é calculado por meio da Distância Euclidiana. Dessa forma, neurônios semelhantes ficam agrupados em regiões do mapa, devido a esse agrupamento, assim, neurônios que contenham possíveis ruídos podem ser previamente identificados por meio de identificação visual. A Figura 3, exemplifica a presença de um neurônio que, provavelmente, é um ruído dentro do agrupamento.



**Figura 3:** Exemplo de agrupamento com dimensão 3x3 realizado pelo SOM, com presença de um neurônio (azul) que possivelmente é um ruído.

Além de criar o mapa bidimensional, o algoritmo também calcula métricas para avaliar se um neurônio pode ter séries consideradas como um ruído. Cada neurônio é um agrupamento de séries temporais, podendo receber séries de apenas uma classe ou de diferentes classes. A probabilidade inicial de uma série temporal pertencer a um neurônio é chamada de Probabilidade Condicional (PC). Quando a PC não atinge um valor ideal, definido pelo usuário, a série é automaticamente excluída do conjunto de dados. Outra métrica calculada é a Probabilidade Posterior (PP), que é um valor recalculado da PC por meio de Inferência Bayesiana, e tem como objetivo levar em consideração a influência dos vizinhos próximos. Quando a PP também não atinge o valor desejado pelo usuário, o algoritmo recomenda a revisão deste neurônio.

O processo de revisão é destacado por SANTOS et al. (2021a) em sua aplicação do SOM no Cerrado. Os autores alertaram que, por se tratar de um agrupamento baseado na semelhança entre as séries temporais, fatores como diferenças no manejo de produção (data de colheita e plantio) e efeitos climáticos podem levar uma amostra a ser considerada como ruído. Em casos onde ocorrem variações intraclases o processo de revisão é relevante para os resultados finais. Por exemplo, SANTOS et al. (2021b) realizaram uma segunda análise intraclasse por meio da clusterização hierárquica nas amostras de cada classe temática. O processo ajudou a identificar as variações intraclases e melhorar os resultados finais na classificação de culturas agrícolas e pastagem.

### 3.5 Metodologias para Lidar com Escassez de Amostras de Campo

Amostras de campo com boa representatividade da área de estudo, e em grande quantidade podem contribuir para mapeamentos mais precisos (SIMOES et al., 2020). Porém, em regiões onde ocorrem a limitação de amostras, existe uma maior dificuldade em aplicar a abordagem tradicional de treinar o modelo com amostras referente a área de estudo, principalmente em métodos de aprendizado profundo, que exigem uma grande quantidade de amostras de treino (MA et al., 2024). Por exemplo, WANG et al. (2023) destacou que redes neurais têm maior dificuldade em trabalhar com conjunto de dados, onde há um desbalanceamento no número de amostras presentes em cada classe temática. A *Synthetic Minority Over-Sampling Technique* (SMOTE) (CHAWLA et al., 2002) é uma opção para lidar com o desbalanceamento de amostras. O SMOTE cria amostras artificiais baseadas nos dados disponíveis na classe temática minoritária, melhorando treinamento do modelo. No entanto, essa técnica não se aplica quando não há ou há baixa disponibilidade de amostras na classe minoritária.

Neste contexto, WANG; AZZARI; LOBELL (2019) propuseram uma metodologia de classificação utilizando o aprendizado não-supervisionado, por meio da clusterização, e dados estatísticos de produção disponibilizados por órgãos dos Estados Unidos. Apesar de promissor, o método fica limitado a países que disponibilizam esses dados estatísticos. Por exemplo, no Brasil este tipo de abordagem dificilmente poderia ser aplicado, pois no país dados estatísticos também são escassos. HUANG et al. (2020), buscaram uma solução para a limitação temporal, ao migrar amostras de campo de 2015 para anos anteriores. O método apresentou bons resultados na classificação, mas o processo de migrar dados de campo se limita às amostras que apresentam o mesmo tipo de uso e cobertura da terra ao longo dos anos. Ou seja, estudos que envolvam a mudança de uso e cobertura da terra seriam limitados, pois classes temáticas de interesse podem variar ao longo dos anos.

No sensoriamento remoto, as abordagens de Generalização e TL estão entre as principais estratégias para lidar com a escassez das amostras. Na última década, ambas foram abordadas em diversos estudos e apresentaram resultados promissores, permitindo que usuários pudessem criar modelos para diferentes regiões com base em um banco de dados representativo (MA et al., 2024; ZHOU et al., 2021). Apesar de serem semelhantes, a aplicação de cada uma dessas técnicas depende de variáveis como o tipo de estudo desejado pelo usuário e características do banco de dados disponível.

### 3.5.1 Transfer Learning

O TL, é uma técnica de aprendizado de máquina que envolve a transferência de conhecimento adquirido em um modelo para ajudar no treinamento de um modelo diferente aplicado a uma tarefa semelhante (PAN; YANG, 2010). No sensoriamento remoto, a abordagem significa treinar o modelo em uma região com abundância de amostras de campo (região fonte) e, em seguida, transferir esse conhecimento para ajudar no treinamento do modelo onde ocorre a escassez de dados (região alvo) (MA et al., 2024). O conceito de transferência de conhecimento, é a principal característica que diferencia o TL da Generalização. Além disso, no TL, há uma etapa de adaptação do modelo. Nessa fase, são utilizadas características que se referem ao pequeno número de amostras disponíveis na região alvo para treinar novamente o modelo e adaptá-lo à nova região de interesse.

MA et al. (2024), revisaram a aplicação do TL no sensoriamento remoto e destacaram a importância da etapa de adaptação nos trabalhos que envolvem o TL. Por exemplo, CUI; CHEN; LU (2020) propuseram uma metodologia de adaptação na qual as amostras da região alvo passam por uma etapa adicional, chamada de pré-treino. Durante essa etapa, as características das amostras são extraídas por meio de redes neurais e aplicadas para adaptação do modelo. Apesar de promissora, os autores destacaram a complexidade do processo de adaptação do modelo. Na China, HU et al. (2022) aplicaram o TL em uma tarefa de classificação interanual de culturas no distrito de Hetao, região com presença de culturas agrícolas irrigadas. Os autores concluíram que o modelo proposto, treinado com dados de 2020 e adaptado, permitiu a classificação das culturas para o ano de 2019 atingindo acurácia de 0,89.

A capacidade do TL em adaptar um modelo para diferentes países, foi testada por HAO et al. (2020). Os autores testaram o TL, utilizando o estado de Nebraska nos Estados Unidos como região fonte, e a região de Hengshui, na China, como região alvo. Para extração de amostras foi utilizado o *Cropland Data Layer* (CDL), um mapeamento que cobre as regiões de produção agrícola dos Estados Unidos, disponibilizado pelo Departamento de Agricultura dos Estados Unidos (USDA).. A classificação de culturas agrícolas foi feita por meio do RF e os autores obtiveram um modelo de classificação para uma região com escassez de amostras e com acurácia de 0,98. Porém, identificaram as dificuldades de adaptação devido variáveis temporais como duração do ciclo e datas de plantio e colheita. Em metodologia semelhante, GE et al. (2021) buscaram minimizar os efeitos de diferenças temporais. Neste trabalho a metodologia envolveu a interpretação das séries temporais e a adaptação das janelas de plantio e colheita de culturas de milho e arroz, devido as diferenças nos períodos de cultivo envolvendo

a região das amostras teste (Estados Unidos) e a região alvo (Liaoning, China). O modelo adaptado foi treinado em nove cenários diferentes com uma rede CNN atingindo acurácia global superior a 0,84 em sete cenários. Os autores também realizaram a adaptação treinando o modelo com o RF, e acurácia global foi superior a 0,84 em seis cenários, e em alguns casos superando os resultados da rede CNN

### 3.5.2 Generalização

No aprendizado de máquina, BLANCHARD; LEE; SCOTT (2011) destacaram o desafio de aplicar um modelo a um conjunto de amostras diferentes daquelas utilizadas no treinamento. Embora o modelo geralmente se adapte bem a amostras desconhecidas que são semelhantes às usadas no treinamento, pequenas variações nas amostras podem resultar na perda de sua eficácia. Os autores ressaltaram, portanto, o desafio de criar um modelo que funcione para o maior número possível de aplicações diferentes. Nesse contexto, o termo '*Domain Generalization*', ou Generalização de Domínio, refere-se à capacidade do modelo de manter sua eficácia quando exposto a novos conjuntos de dados, sem a necessidade de um processo de adaptação, que é comum no TL (WANG et al., 2021; ZHOU et al., 2021).

No campo do sensoriamento remoto, a generalização de um modelo permite sua aplicação em classificações em diferentes áreas e anos. Autores como ZHONG et al. (2016) e JOHNSON (2019) apontaram que a semelhança fenológica entre culturas como soja, milho e trigo podem ser aproveitadas e utilizadas para expandir as aplicações dos modelos de aprendizado de máquina. Nesse sentido, os autores destacaram a possibilidade de realizar futuros novos mapeamentos sem a necessidade de recalibração do modelo, o que reduziria o tempo e os custos com novos recursos, além de abranger novas áreas de estudo. NASIRI et al. (2023) também conseguiram aplicar um modelo de Generalização desenvolvido com o objetivo de monitorar áreas de desmatamento florestal por meio do algoritmo RF. Esse estudo demonstra a aplicação da Generalização para análises de mudanças no uso e cobertura da terra.

A Generalização também pode ser aplicada em trabalhos de mapeamentos ou classificação de séries temporais. No entanto, LUCIANO et al. (2018) e ZHONG; GONG; BIGING (2014) apontaram que algumas condições devem ser cumpridas (I) variações dentro das classes temáticas podem ocorrer, porém não devem ser muito diferentes em relação aos padrões encontrados no conjunto de treino; (II) as classes temáticas devem ser as mesmas, ou seja, não é possível adicionar novas classes para a região testada; (III) as variações interanuais devem estar no conjunto de treino (Ex: variação no NDVI causada por um período de seca). As

condições (I) e (III) representam a necessidade por amostras de treino que tenham boa qualidade e sejam representativas, além de apontar para o cuidado em analisar e estudar a área de estudo em que o modelo será aplicado. Por exemplo, JOHNSON (2019), aplicou o conceito de Generalização nas regiões produtoras de soja, milho e trigo nas regiões central e leste dos Estados Unidos. As amostras de treino foram coletadas ao longo de quatro anos (2008 até 2011), e treinadas por meio de árvores de decisão. O modelo foi generalizado para o mapeamento anual nos anos de 1984 até 2007. Em geral, os resultados obtidos apresentaram baixa correlação com os mapeamentos da CDL, o autor destacou perda de acurácia devido a extensão da série e do território estudado, onde variações nos manejos de produção prejudicaram o desempenho do classificador.

Neste contexto, KHANNA et al. (2022) aplicaram a Generalização dividindo os Estados Unidos em quatro domínios e adicionando dados referentes às variáveis geográficas de cada domínio para a rede *Transformer*, melhorando a representatividade do conjunto de treino. Os autores verificaram que o uso da *Transformer* e a divisão dos Estados Unidos em quatro domínios, representam uma alternativa para possíveis novos trabalhos sobre o tema, contribuindo para a caracterização da área de estudo e a Generalização espacial. MASSEY et al. (2017) também utilizaram os Estados Unidos como área de estudo e realizaram uma Generalização com base em árvores de decisão para o período de 2001 até 2014. Para avaliar a qualidade da Generalização, modelos foram criados individualmente para os de 2008, 2010, 2011 e 2012 usando amostras específicas para estas datas, e coletadas ao longo das regiões agrícolas localizadas no centro do país. Os resultados foram comparados com o modelo generalizado. Os autores encontraram diferenças mínimas nos resultados dos modelos individuais (Acurácia média 78%) e generalizado (Acurácia média 75%).

No Brasil, LUCIANO et al. (2018, 2019) treinaram um modelo de classificação RF para mapear áreas de produção de cana-de-açúcar no estado de São Paulo. Modelos foram treinados com base em amostras disponíveis no ano de 2016, e foram capazes de classificar áreas em anos-safras próximos, como o ano de 2015 que foi utilizado para validação. Na mesma linha de pesquisa, LUCIANO; CAMPAGNUCI; LE MAIRE (2022) realizaram a generalização de 33 anos (1986 até 2019) de séries temporais de cana-de-açúcar para todo o estado de São Paulo. Para esta tarefa, os autores utilizaram três coleções Landsat (5, 7 e 8) e treinaram quatro modelos RF, variando o ano base para as amostras de treinamento. O melhor modelo observado, foi aquele em que mais amostras de diferentes anos (2009 até 2014) foram aplicadas, pois isso permitiu que o modelo fosse treinado com um conjunto mais representativo de amostras. Neste trabalho, os autores utilizaram mapeamentos dos projetos MapBiomass e Canasat e compararam

os resultados obtidos por meio da Generalização. Os resultados da Generalização apontaram correlação superior a 0,9 com os dois mapeamentos, exemplificando o potencial da abordagem para classificação de uso e cobertura da terra.

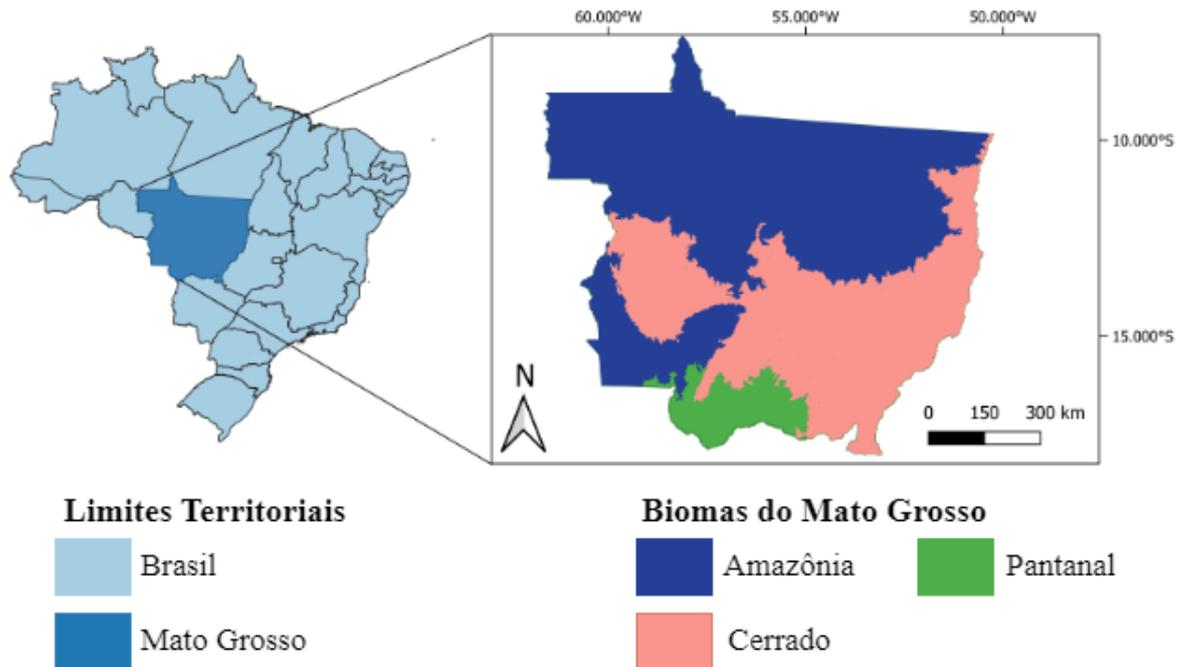
## **4 Material e Métodos**

### **4.1 Material**

#### **4.1.1 Área de Estudo**

A área de estudo abrange todo o estado de Mato Grosso. A região foi escolhida devido a sua importância agrícola para o país, conhecido por ser o principal produtor de grãos, com destaque para a soja, milho e algodão. Durante a safra de verão, a cultura da soja é predominante no estado e, no período de safrinha, as principais culturas praticadas são o milho, o algodão, o sorgo e o girassol. Esse manejo permite a produção de grãos em dois períodos distintos ao longo do ano-safra, sendo considerado o principal sistema de produção presente no estado. Além disso, também é comum a presença de cultivos não comerciais, como a crotalária e o milheto, que servem de cobertura para manejos conservacionistas (ARVOR et al., 2012). No estado também há a presença de lavouras de cana-de-açúcar e regiões de pastos, que podem ser destinados, ou não, para a criação de gado.

Além de culturas agrícolas, no Mato Grosso também é possível encontrar diversos tipos de coberturas de vegetação natural, já que o estado abrange três diferentes biomas: Amazônia, Cerrado e o Pantanal (Figura 4). Além disso, o estado também é conhecido pelo processo de intensificação agrícola, inicialmente marcado pela expansão territorial nos biomas Cerrado e Amazônia, e em seguida caracterizado pela adoção do sistema de produção agrícola de dois ciclos, denominados de safra e safrinha (DANTAS CHAVES; DE CARVALHO ALVES, 2019). Neste sentido, dada a diversidade de biomas, a variedade de culturas e a expansão agrícola, as séries temporais desta região apresentam perfis espectro-temporais variados, sendo um bom objeto de estudos para testar a capacidade preditiva dos algoritmos de aprendizado de máquina.



**Figura 4:** Mapa do estado de Mato Grosso, representando a distribuição dos três biomas ao longo do seu território. Fonte: IBGE (Instituto Brasileiro de Geografia e Estatística).

#### 4.1.2 Pacote Computacional Sits

O *sits* é um pacote computacional desenvolvido na linguagem R (<https://github.com/e-sensing/sits>) que contém diversas funções que permitem ao usuário executar análises, simples ou complexas, de séries temporais de imagens de satélite. SIMOES et al. (2021) destacaram que a proposta do pacote é criar uma biblioteca que permita que o usuário realize todas as etapas de classificação baseadas em séries temporais para a geração de mapas de uso da terra, partindo desde a aquisição das séries temporais até as etapas finais de pós-classificação.

O pacote é compatível com os produtos de sensores como o MODIS, Landsat-8/OLI e Sentinel-2/MSI, desde de que estejam no formato de cubo de dados. A importação dos dados para o ambiente *RStudio* pode ser feita acessando plataformas como o BDC, ou a partir de um diretório local contendo as imagens empilhadas, e que obedeçam aos critérios de um cubo de dados, conforme apresentados por FERREIRA et al. (2020). Após a extração dos dados das imagens, o *sits* tem funções que permitem a análise e pré-processamento das amostras como a plotagem das séries temporais, aplicação de filtros (*Savitsky-Golay*), balanceamento das amostras (*Oversampling* e *Undersampling*) e a análise de ruídos por meio do algoritmo SOM (SIMOES et al., 2021).

Na etapa de treinamento de modelos de classificação do uso e cobertura da terra, geralmente, são utilizados algoritmos de aprendizado de máquina como o RF. Porém, no *sits*, o usuário também tem a opção de criar modelos utilizando algoritmos baseados em aplicações voltadas para o uso de séries temporais no sensoriamento remoto, como as redes neurais profundas L-TAE e TempCNN, propostas por SAINTE FARE GARNOT et al. (2020) e PELLETIER; WEBB; PETITJEAN, (2019), respectivamente.

Dessa forma, devido a praticidade e a quantidade de aplicações para trabalhar com séries temporais, o pacote *sits* foi escolhido como a principal ferramenta para o desenvolvimento deste trabalho, uma vez que nele foi possível realizar etapas importantes da metodologia. Por exemplo, a extração das séries temporais do MODIS utilizando o Data Cube, o uso do SOM para trabalhar a qualidade das amostras de treino, a modelagem utilizando algoritmos de aprendizado de máquina e as métricas para avaliação das classificações.

#### **4.1.3 Dados Espectrais**

Neste trabalho, foram utilizados os produtos do sensor MODIS, que foi escolhido devido aos padrões da produção agrícola no Mato Grosso, onde é predominante formado por estruturas fundiárias extensas e com mesma cultura ao longo do terreno, características compatíveis com a resolução espacial de 250m do MODIS. O estado também apresenta áreas extensas de vegetação nativa, como florestas e pastos. Outro motivo para a escolha do MODIS, é a sua série temporal extensa, sendo um bom objeto de estudo para testar a capacidade de generalização dos modelos de aprendizado de máquina. Em relação aos dados, foram utilizadas as coleções MOD13Q1 e MYD13Q1, referentes às coleções de imagens dos satélites TERRA e AQUA, respectivamente. Os produtos do MODIS disponibilizam imagens com composição de 16 dias utilizando o melhor valor de pixel do período, minimizando o efeito da cobertura de nuvens.

Com o objetivo de simplificar e permitir o trabalho com as séries temporais de forma mais organizada, foram criadas estruturas chamadas de cubos de dados (ou empilhamento de imagens), contendo todas as datas disponíveis para os períodos analisados. As imagens do MODIS foram processadas, empilhadas e fornecidas pela Embrapa Agricultura Digital. O cubo de dados foi criado seguindo metodologia apresentada por FERREIRA et al. (2020) com o auxílio do pacote *sits*, que permite a criação de cubos com imagens em ambientes locais. O cubo de dados é composto pelos índices de vegetação NDVI e EVI, as bandas espectrais (BLUE, MIR, NIR, RED) e a banda CLOUD, referente ao ajuste das séries temporais

em caso de quebras ocasionadas pela presença de nuvens. Esta última é derivada da banda de confiabilidade do pixel, presente nos produtos MOD/MYD13Q1.

Para padronizar os processos de extração das séries temporais, classificação dos perfis espectro-temporais e análises no SOM, todos os cubos de dados criados apresentaram mesmo padrão de início e fim de um ano-safra. De acordo com informações de órgãos como CONAB e IBGE, e na interpretação dos perfis espectro-temporais disponibilizados na plataforma do SATVeg, neste trabalho foi considerado que o ano-safra do Mato Grosso teve início no mês setembro de um ano e fim em setembro do ano seguinte. Por exemplo, para as amostras de treinamento foi utilizado como base a safra 2019/2020, assim, o cubo de dados deste período abrange os meses de 09/2019 até 09/2020.

Para realizar a classificação interanual no *sits*, os demais períodos de ano-safra foram correspondentes ao utilizado no treinamento. Por exemplo, para criar um cubo referente as amostras de teste do ano-safra 2016/2017, o período utilizado foi de 09/2016 até 09/2017. Neste sentido, todas as classificações realizadas seguiram o padrão de início e fim de um ano-safra utilizado nas amostras de treino.

#### **4.1.3 Amostras de Treinamento dos Modelos de Classificação**

Neste trabalho, foram utilizados métodos de aprendizado de máquina baseados em redes neurais, em que uma grande quantidade de amostras de campo foi necessária para treinar os modelos de classificação (ISMAIL FAWAZ et al., 2019). Devido a demanda por dados, a amostragem foi feita com base nos mapeamentos disponibilizados pelos projetos TerraClass Amazônia e TerraClass Cerrado, ano base 2020 (<https://www.terraclass.gov.br>). Outro motivo para a escolha pelos dados do projeto TerraClass foi a necessidade de um banco de dados que representasse as regiões dos biomas Cerrado e Amazônia, dentro do estado de Mato Grosso. Essa condição iria exigir um trabalho complexo de amostragem, caso os dados de treinamento fossem exclusivamente amostras coletadas em campo.

O mapeamento TerraClass, apresenta 16 classes temáticas, das quais foram adotados cinco tipos, conforme Tabela 1. As classes de agricultura deste trabalho estão alinhadas com o padrão utilizado no TerraClass e não precisaram de adaptações. A classe Pastagem (PAST) foi mesclada, incluindo a Pastagem Arbustiva e a Pastagem Herbácea, devido à ausência de Pastagem Arbustiva no mapeamento do bioma Cerrado. Essa fusão teve como objetivo padronizar o processo de amostragem dessa vegetação e incluir amostras com características da Pastagem Arbórea no treinamento do modelo. Já a Vegetação Florestal (VF),

unificaram-se as classes Vegetação Florestal Primária e Secundária, e a de Silvicultura. Essas três classes representam áreas de floresta no estado e foram utilizadas no mesmo agrupamento de classes temáticas por possuírem características semelhantes em perfis espectro-temporais com duração de uma ano-safra.

<b>Classe Temática</b>	<b>Classes TerraClass</b>	<b>Descrição</b>
<b>Agricultura Temporária 1 ciclo (AGR-1)</b>	Agricultura Temporária de 1 ciclo	Culturas agrícolas que apresentam um ciclo de produção no ano-safra.
<b>Agricultura Temporária 2 ciclos (AGR-2)</b>	Agricultura Temporária 2 ou mais ciclos	Culturas agrícolas que apresentam dois ou mais ciclos de produção no ano-safra.
<b>Agricultura Semiperene (AGR-S)</b>	Agricultura Semiperene	Culturas agrícolas com ciclo superior a um ano-safra, representadas, principalmente, pela cana-de-açúcar.
<b>Pastagem (PAST)</b>	Pastagem Herbácea	Pastagens com predomínio de vegetação forrageira herbácea.
	Pastagem Arbustiva	Pastagens com predomínio de vegetação lenhosa, composta por espécies arbustivas/arbóreas.
<b>Vegetação Florestal (VF)</b>	Vegetação Florestal Primária	Vegetação natural caracterizada pelo adensamento de espécies arbóreas que não tenham sofrido com ação humana.
	Vegetação Florestal Secundária	Vegetação natural em processo avançado de regeneração.
	Silvicultura	Culturas de espécies florestais de interesse comercial. Ex: Eucalipto e Pinus.

**Tabela 1:** Tabela com as classes temáticas e suas respectivas descrições utilizadas no mapeamento TerraClass, que serviram de base para as cinco classes temáticas utilizadas neste trabalho.

#### 4.1.4 Amostras para Teste da Capacidade de Generalização dos Modelos

Para verificar e validar a capacidade de generalização do modelo, foram utilizadas amostras de campo disponibilizadas pela Embrapa Agricultura Digital. Este grupo de dados, é referente a pontos em regiões com características conhecidas e coletados com auxílio da plataforma SATVeg, e amostras obtidas através de visitas a campo, sendo distribuídas nas safras 2016/2017, 2020/2021 e 2021/2022. O processo de amostragem realizado pela Embrapa foi conduzido ao nível de cultura. Nesse sentido, as amostras de teste foram adaptadas para as cinco classes temáticas deste trabalho. A adaptação das classes PAST e VF seguiram o mesmo padrão feito com as amostras dos mapeamentos do TerraClass. Para a classe AGR-2, as amostras de soja/milho, soja/algodão, soja/feijão e soja/sorgo foram unificadas, representando dois ciclos de produção agrícola. As amostras de soja/pousio e soja/pasto, apresentaram um ciclo agricultura e foram mescladas na classe AGR-1. Por fim, na classe AGR-S foram incluídas apenas as amostras de cana-de-açúcar. A distribuição do número de amostras de teste em cada classe temática está disponível na Tabela 2.

Classe	Amostras
<b>Agricultura Temporária 2 ciclos</b>	1.628
<b>Agricultura Temporária 1 ciclo</b>	1.076
<b>Agricultura Semiperene</b>	905
<b>Pastagem</b>	1.165
<b>Vegetação Florestal</b>	754
<b>Total</b>	5.528

**Tabela 2:** Tabela referente a distribuição das amostras de campo das cinco classes temáticas, utilizadas para a etapa de teste da Generalização.

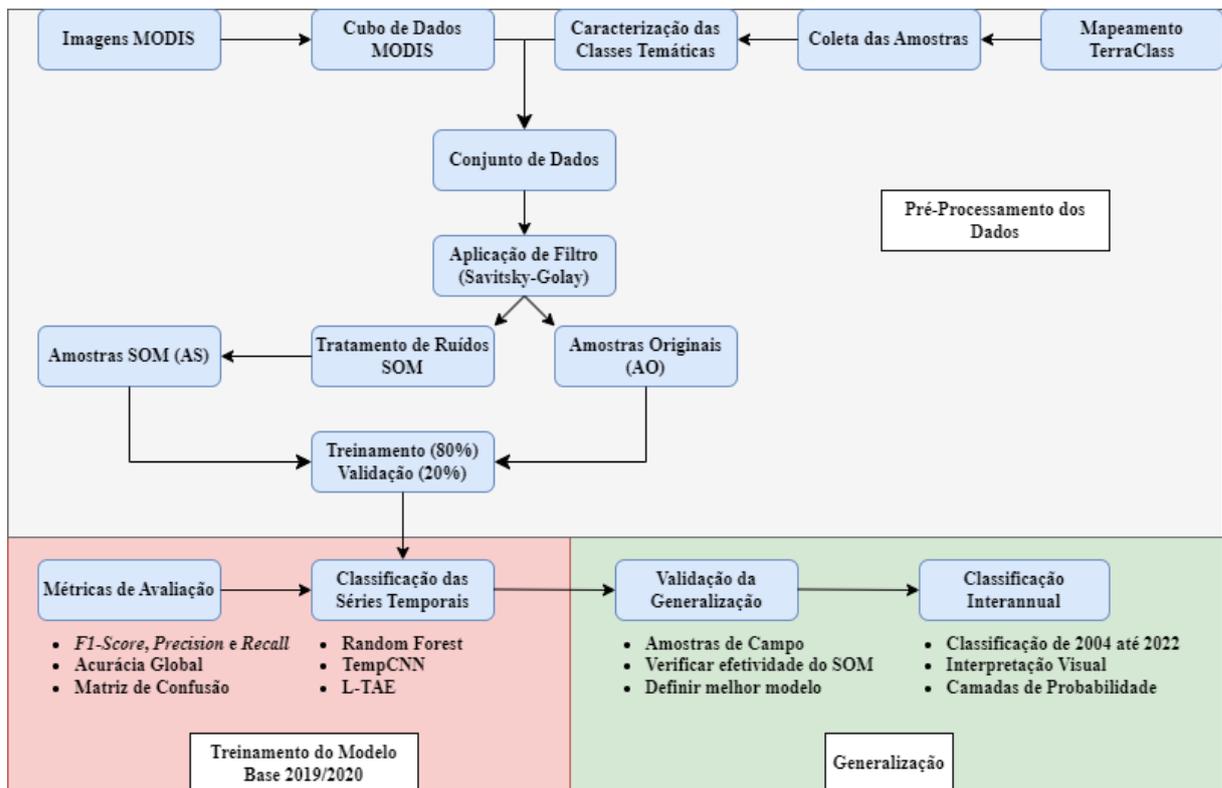
Neste conjunto, formado por três anos-safra distintos, as classes estão balanceadas e nenhuma destas amostras participou da etapa de treinamento do modelo. Por esta razão, este conjunto apresenta uma quantidade aceitável de amostras para validar a capacidade de generalização do modelo nesses anos-safra.

## 4.2 Métodos

A metodologia deste trabalho é apresentada no fluxograma de atividades na Figura 5. Para simplificar o planejamento das atividades, a metodologia foi dividida em quatro etapas: Pré-processamento dos dados, Treinamento do modelo base 2019/20 e a Classificação interanual (Generalização). A etapa de pré-processamento demandou um grande número de

tarefas para garantir os cuidados necessários com os dados de entrada, principalmente por se tratar de um trabalho que envolve o aprendizado de máquina (MAROTO-MOLINA et al., 2013).

As etapas desta metodologia foram orientadas de forma a cumprir o objetivo da Generalização, baseado nas condições necessárias apontadas por LUCIANO et al. (2018). O pré-processamento foi feito com o intuito de obter um banco de dados extenso para aplicação do aprendizado profundo, representativo para o Mato Grosso e livre de ruídos que pudessem levar ao mal desempenho da classificação. O objetivo do treinamento foi criar um modelo capaz de classificar o ano-safra 2019-2020 de forma acurada, sem efeitos de *overfitting*. Seis cenários diferentes foram testados para o treinamento, obtendo seis modelos diferentes. Por fim, foi realizado um processo de validação com amostras de campo para verificar a capacidade de generalização de cada modelo.



**Figura 5:** Fluxograma ilustrando a metodologia utilizada neste trabalho.

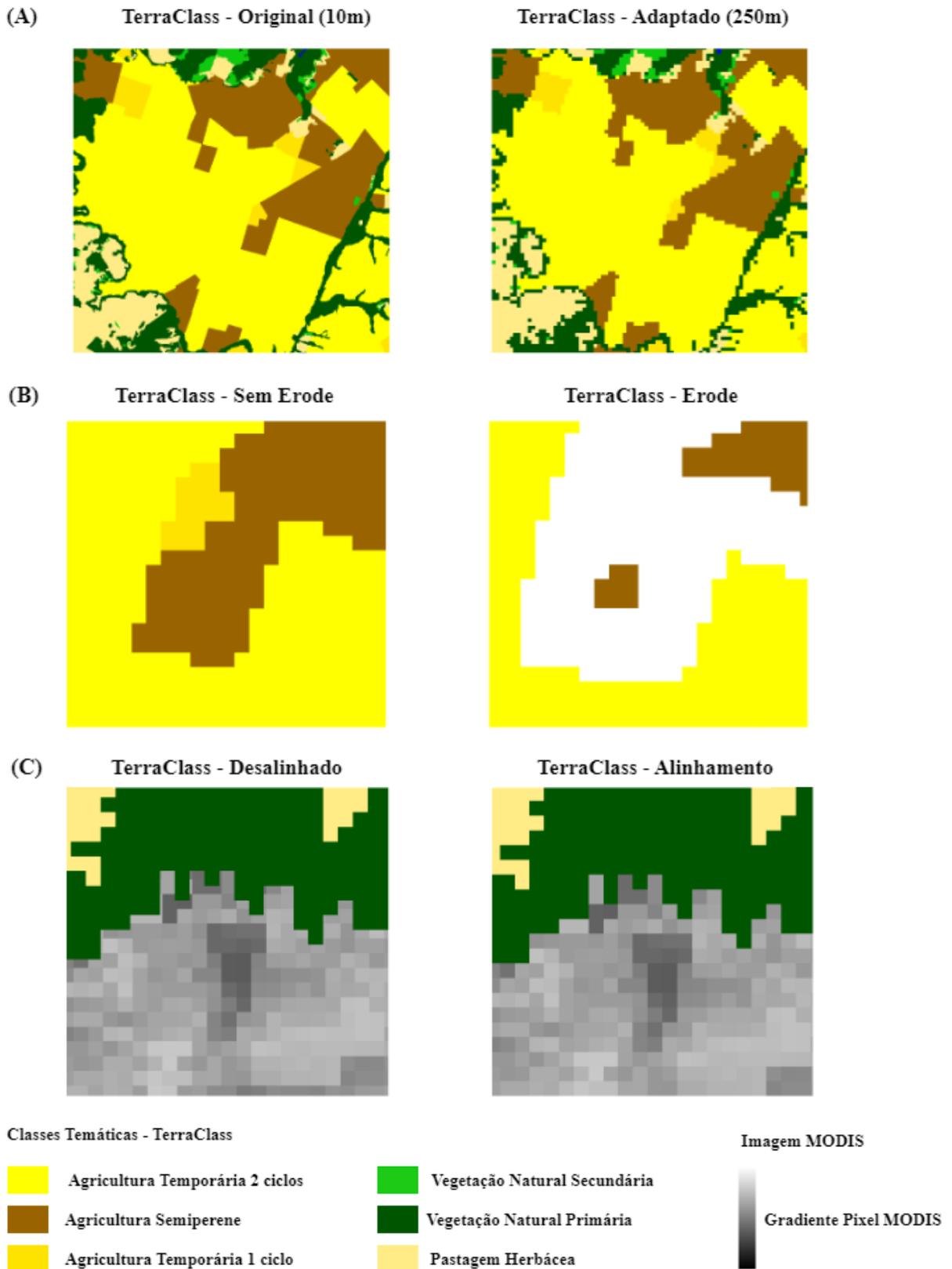
#### 4.2.1 Criação do Conjunto de Dados de Treinamento e Teste

Na primeira etapa de aquisição das amostras, no software QGIS, os mapas TerraClass Cerrado e Amazônia foram integrados, e em seguida, recortados de acordo com os limites territoriais do estado de Mato Grosso. Os mapas TerraClass foram feitos com base em

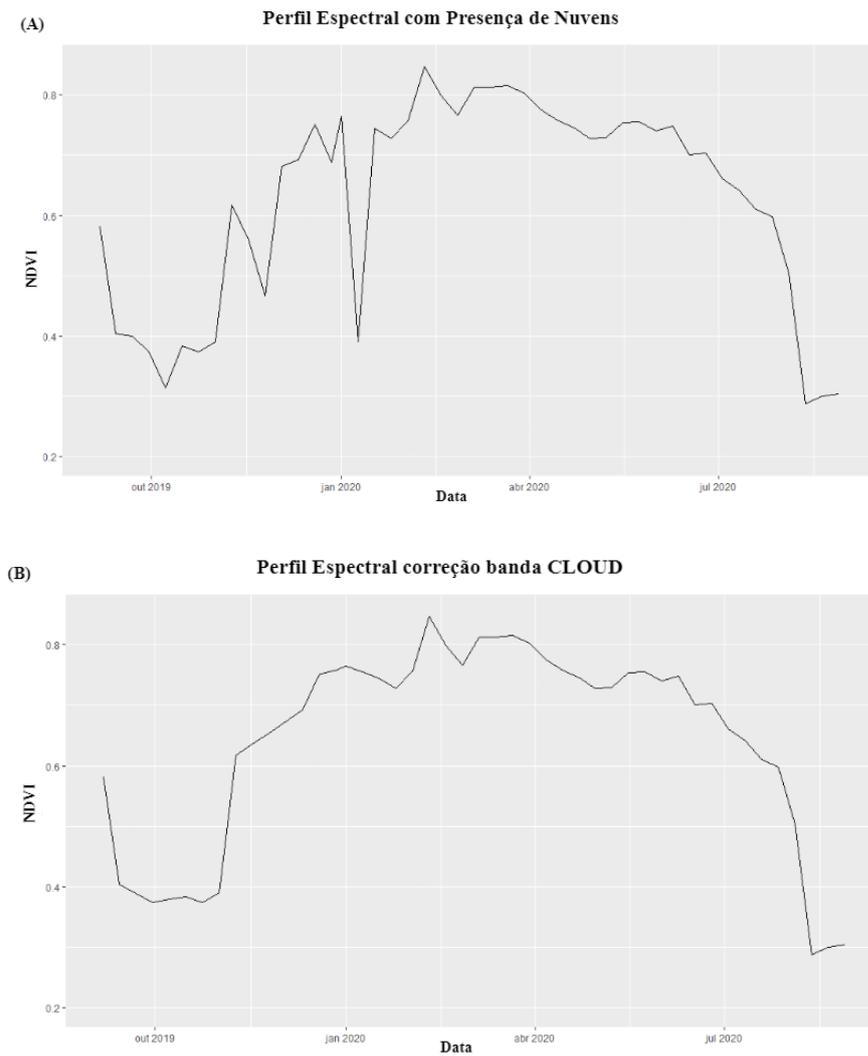
imagens do Sentinel-2/MSI, que possuem resolução espacial de 10m. Como o MODIS apresenta pixels de 250m, os seguintes passos foram realizados: (I) A adaptação da resolução espacial foi realizada utilizando a ferramenta de conversão disponível no QGIS. Nesta etapa, pixels de 10m foram agrupados para criar pixels de 250m, sendo atribuída a classe temática que apresentar maior quantidade de pixels dentro do agrupamento (Figura 6A). (II) Foi realizado um alinhamento dos pixels do mapeamento e de uma imagem do MODIS. A Figura 6C exemplifica o desalinhamento dos pixels, que pode ocorrer após o processo de adaptação de resolução espacial. (III) Para evitar efeitos de borda na amostragem, foi utilizada a função Erode, disponível no plugin SCP (*Semi-Automatic Classification Plugins*) no QGIS. O Erode recorta as bordas de agrupamentos de pixels com mesma classe temática, permitindo que apenas os pixels centrais destes agrupamentos fossem utilizados na amostragem. Neste trabalho o recorte foi de 2 pixels, o equivalente a aproximadamente 500m (Figura 6B).

A amostragem também foi feita no QGIS, utilizando o plugin Acatama (*Accuracy Assessment of Thematic Maps*) (LLANO, 2022). Seguindo recomendações propostas por OLOFSSON et al. (2014), os pontos foram selecionados aleatoriamente ao longo da área estudo, evitando a concentração de amostras em uma região específica. O processo de aquisição foi baseado em pixels, ou seja, cada amostra foi retirada a partir do centroide de cada unidade. Esse método limita a amostragem em um ponto por pixel, evitando a coleta de pontos duplicados e diminui a redundância do conjunto. No final, foi criado um arquivo em que cada amostra fornece as coordenadas geográficas (latitude e longitude) e a classe temática referente a cada ponto, informações necessárias para extração das séries temporais no pacote *sits*.

A extração das séries temporais foi feita por meio da função “*sits\_get\_data*”. Esta função permite a integração das séries temporais presentes no cubo de dados com base nas coordenadas geográficas das amostras de campo. Nesta etapa, também foi incluído o ajuste para a presença de nuvens por meio da banda CLOUD. O *sits* utiliza os dados da banda CLOUD para identificar quebras nas séries temporais causadas por nuvem, e automaticamente ajustar oscilações dentro da série temporal (Figura 7). O resultado da função “*sits\_get\_data*” é um conjunto de dados em que, cada amostra contendo um rótulo da classe temática de uso e cobertura da terra, recebe as séries temporais de índice de vegetação e bandas espectrais referentes ao ponto de coleta.



**Figura 6:** Figura ilustrando o processo de adaptação do mapeamento TerraClass para uma amostragem baseada nas imagens do MODIS. (A) Adaptação da resolução espacial de 10m (Sentinel-2MSI) para 250m (MODIS). (B) Exemplo de processo de remoção de bordas por meio da função Erode. (C) Alinhamento dos pixels do mapa TerraClass adaptado, com os pixels de uma imagem MODIS.



**Figura 7:** Figura comparativa de um perfil espectro-temporal sem correção de nuvens (A) e com correção realizada com o auxílio da banda CLOUD (B).

#### 4.2.2 Aplicação de Filtros

No sensoriamento remoto orbital, a presença de nuvens prejudica a aquisição de dados da superfície, causando interferências no sinal da série temporal. Neste trabalho, as séries temporais são produtos de uma composição de 16 dias, que visa diminuir o efeito das nuvens. Apesar do uso da banda CLOUD, algumas inconsistências ainda podem ser encontradas nas séries temporais e, para evitar que esse tipo de problema interfira nos resultados finais, foi aplicado um filtro de suavização. O filtro aplicado foi o *Savitsky-Golay*, que está presente no pacote *sits* (SAVITSKY; GOLEY, 1964).

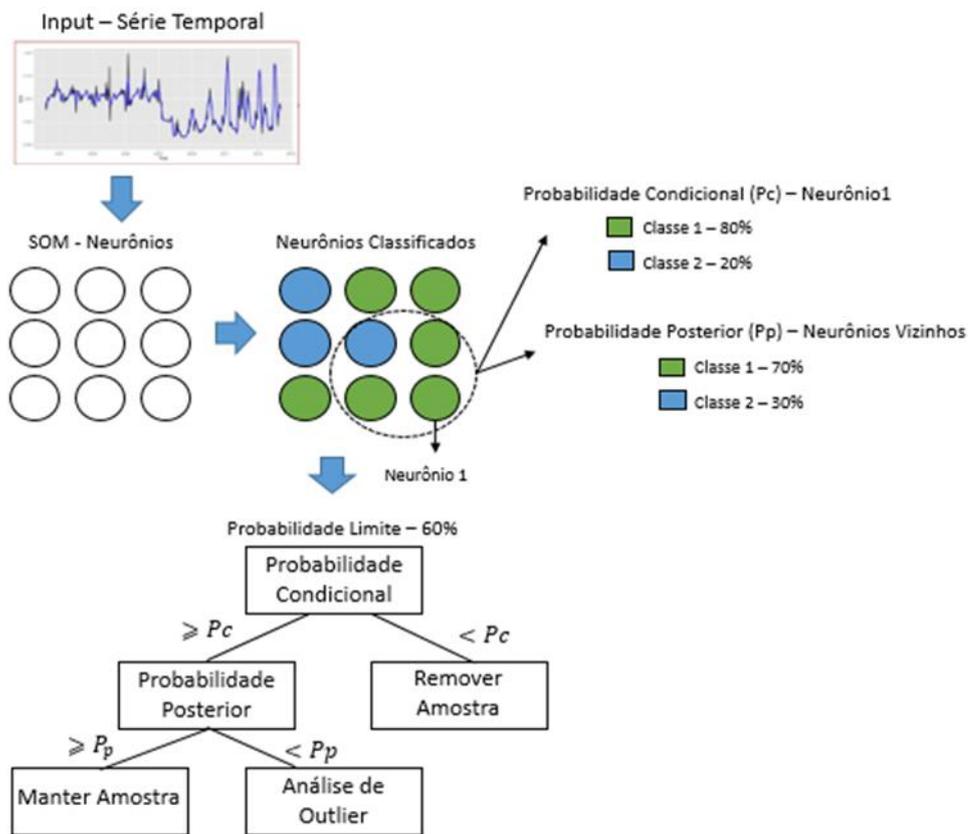
### 4.2.3 Tratamento de Amostras (SOM)

A identificação e remoção de ruídos foi conduzida por meio do método baseado no *Self-Organizing-Maps* (SOM) (SANTOS et al., 2021a). No pacote *sits*, a função “*sits\_som\_map*” utiliza o algoritmo proposto inicialmente por KOHONEN (1990), que permite que o usuário aplique o algoritmo e realize interpretações referentes ao seu banco de dados. Neste trabalho, foram utilizados mapeamentos como base da amostragem, abordagem semelhante a outros trabalhos que também adotaram uma metodologia baseada na Generalização (KHANNA et al., 2022; MASSEY et al., 2017). Mapeamentos como o TerraClass, geralmente, são resultados de modelos baseados em aprendizado de máquina, por isso estão suscetíveis a apresentarem erros de classificação. Nesse contexto, foi considerado que durante o processo de coleta amostral, existia uma alta probabilidade de haver amostras com características incoerentes com sua respectiva classe temática. Portanto, a análise das amostras por meio do SOM foi considerada uma etapa fundamental. O processo metodológico para análise de ruídos está disponível na Figura 8, e foi baseado na metodologia apresentada por SANTOS et al. (2021a, 2021b).

O mapa bidimensional teve dimensões de 30x30, totalizando 900 neurônios. A escolha por este número está relacionada a grande quantidade de dados de treinamento e à necessidade de identificar possíveis variações intraclasses. Para calcular a distância entre neurônios, foi utilizada a distância Euclidiana, métrica recomendada por SANTOS et al. (2021a) devido aos melhores resultados e adaptação com as séries temporais. As duas métricas para análise de ruído, PC e PP, foram definidas como 60%.

Na etapa de identificação de ruídos, neurônios que apresentaram PC menor do que 60% foram excluídos automaticamente do conjunto de amostras de treino. Esta etapa adaptou o processo de identificação, avaliação e exclusão de possíveis ruídos no conjunto de amostras, conforme exemplificados por SANTOS et al. (2021b). Neste sentido, os critérios utilizados para análise dos neurônios foram: (I) Neurônios que contêm amostras com PP inferiores a 60% foram analisados; (II) Neurônios distantes do agrupamento de sua respectiva classe temática; e (III) Neurônios localizados nas bordas de dois ou mais agrupamentos. A terceira condição foi proposta, pois devido ao processo de adaptação do mapeamento TerraClass, a possibilidade de haver um pixel que continha, originalmente, mais de uma classe temática era alta. Neste contexto, amostras com confusão na classe temática poderiam estar presentes, e prejudicar o treinamento do modelo.

Como ferramenta de apoio foi utilizado o SATVeg para verificar a série histórica de um ponto. O objetivo desta verificação foi confirmar se a discrepância entre neurônios da mesma classe foi causada por erro de classificação ou por variações intraclasse. Em caso de variações intraclasse, os neurônios foram mantidos, pois a presença deles significa maior representatividade da área de estudo. Se as informações observadas na série não estavam em conformidade com a classe temática, o neurônio foi excluído ou transferido para outra classe, caso as características do perfil espectral correspondessem a mesma.



**Figura 8:** Fluxograma do processo de criação do mapa bidimensional e metodologia de análise de ruídos por meio das métricas de Probabilidade Condicional e Posterior. Adaptado de SANTOS et al. (2021a).

#### 4.2.4 Criação dos Modelos de Classificação

Nesta etapa, foram criados os modelos base para a classificação dos perfis espectro-temporais com base nas séries temporais do MODIS. O conjunto de dados foi dividido em dois: treinamento (80%) e validação (20%). A divisão foi estratificada, com o objetivo de manter as proporções de amostras dentro das classes temáticas. A divisão teve como intuito validar o modelo 2019/20 com amostras do mesmo ano, por isso o conjunto de validação foi separado da etapa de treinamento dos modelos.

Os modelos de classificação foram orientados ao nível de pixel, conforme o padrão do pacote *sits*. A função utilizada foi a “*sits\_train*”, que tem como principal característica treinar um modelo baseando em um algoritmo de classificação. Neste trabalho, foram utilizados três classificadores: *Random Forest* (RF), *Temporal Convolutional Neural Network* (TempCNN) e a *Lightweight Temporal Attention Encoder* (L-TAE).

No processo de treinamento dos modelos, foi realizado um processo de seleção de hiperparâmetros para auxiliar no desempenho dos algoritmos. Para o algoritmo RF, os valores dos hiperparâmetros *n*tree e *m*try foram definidos como 500 e 15, respectivamente. Por outro lado, nos algoritmos L-TAE e TempCNN, o processo de seleção de hiperparâmetros foi mais complexo em comparação com o RF, devido às arquiteturas mais detalhadas das redes neurais. Nesse sentido, durante o treinamento com esses algoritmos, foi empregada a função '*sits\_tuning\_hparams*' responsável por selecionar os melhores hiperparâmetros para cada modelo.

Nesta etapa, foram desenvolvidos seis modelos para explorar diferentes cenários durante o treinamento e determinar qual dos algoritmos apresentou melhores resultados. Foram criados dois modelos para cada algoritmo de classificação (RF, TempCNN e L-TAE). O primeiro modelo foi treinado usando o conjunto de dados sem qualquer tratamento, enquanto o segundo modelo foi treinado após análise por meio do SOM. O objetivo dessa abordagem foi avaliar qual modelo melhor se adapta a tarefa de Generalização e investigar o impacto do tratamento realizado pelo SOM nos resultados finais.

#### 4.2.5 Validação dos Modelos de Classificação

Para avaliar o desempenho dos classificadores foram utilizadas métricas para a análise de possíveis erros de classificação. As métricas utilizadas foram Acurácia Global, *F1-Score*, *Precision*, *Recall* e a matriz de confusão (CONGALTON, 1991). A avaliação utilizando essas métricas tem como objetivo evitar problemas com *overfitting*, situação em que o modelo é adaptado apenas para as amostras de treinamento, apresentando alta acurácia, porém falha na classificação de outras amostras.

O parâmetro *F1-Score* é uma métrica associada aos valores de *Precision* e *Recall*, sendo utilizado para quantificar a qualidade da predição realizada. *Precision* é um valor associado ao erro de omissão, sendo referente a taxa de verdadeiros positivos de uma classificação. Já o *Recall* está associado ao erro de comissão, que computa a taxa de falsos negativos. Essas métricas podem ser encontradas na matriz de confusão, para avaliar como aconteceu a classificação em cada uma das classes temáticas, assim, permitindo que erros de classificação como a confusão entre classes de comportamento espectral diferentes, sejam encontrados e analisados (CONGALTON, 1991; PELLETIER et al., 2016). Todas as métricas de avaliação utilizadas foram geradas automaticamente ao final de cada classificação gerada por meio das funções do pacote *sits*.

No aprendizado de máquina, quando uma amostra passa por um modelo de classificação, é atribuída a ela uma classe temática. No *sits*, esse processo pode ser visualizado e interpretado por meio das camadas de probabilidade, que representam as probabilidades de um pixel pertencer a uma das classes temáticas. Quanto maior a probabilidade, menor é a confusão com outras classes temáticas. Neste sentido, o classificador não teve dificuldades em identificar esse padrão. Por outro lado, quanto menor o valor da probabilidade, maior a confusão entre classes e, portanto, o resultado da classificação não pode ser visto como confiável.

#### 4.2.6 Generalização

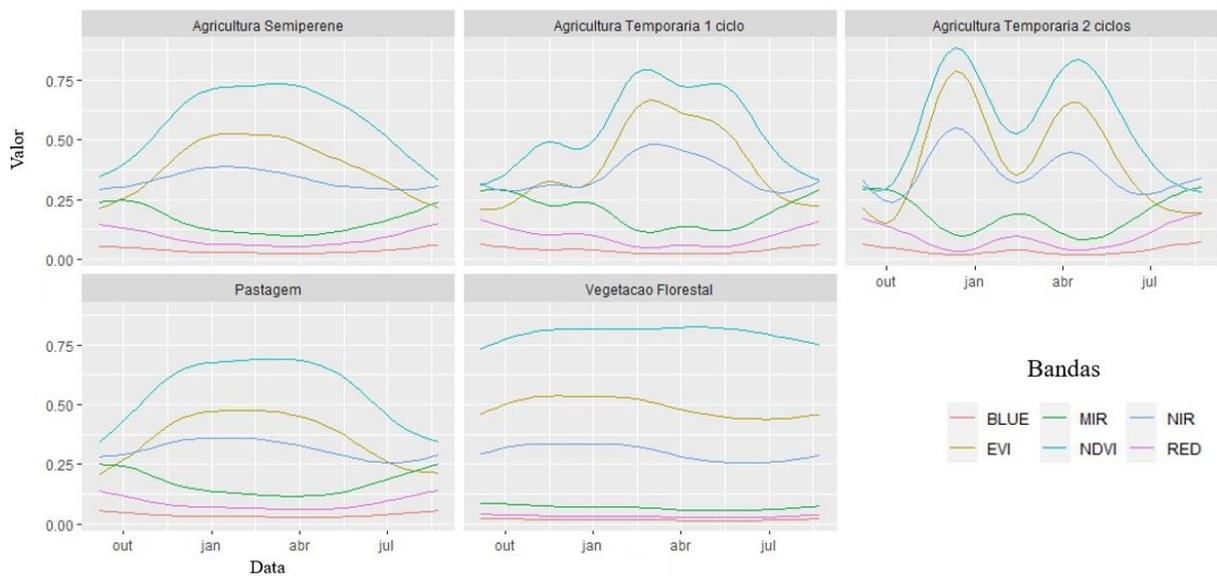
Após a criação e validação do modelo para a safra 2019/2020, o modelo foi aplicado em uma abordagem de Generalização. A classificação dos demais anos safra foi simplificada pelo uso da função “*sits\_classify*”, que divide a séries temporais em períodos iguais aos padronizados na etapa de treinamento (09/2019 até 09/2020). Neste trabalho, a série do MODIS iniciou em 09/2004 e encerrou em 09/2022, totalizando 18 períodos para classificação referentes a um ano-safra para cada pixel.

As métricas de validação utilizadas na Generalização foram as mesmas utilizadas para avaliar o modelo base, porém aplicadas apenas para o conjunto com as amostras de campo de 2016/17, 2020/21 e 2021/22. A partir dos resultados destas métricas o melhor dentre os seis modelos de classificação foi escolhido e aplicado para a classificação ano a ano dos 18 intervalos da série. Nesta classificação, amostras de campo não estavam disponíveis para todos os anos, portanto, o processo de validação ocorreu por interpretação visual com o auxílio de ferramentas como o SATVeg e a interpretação das camadas de probabilidades, com o intuito de verificar se não ocorreram incoerências na classificação.

## 5 Resultados e Discussões

### 5.1 Análise Exploratória dos Dados

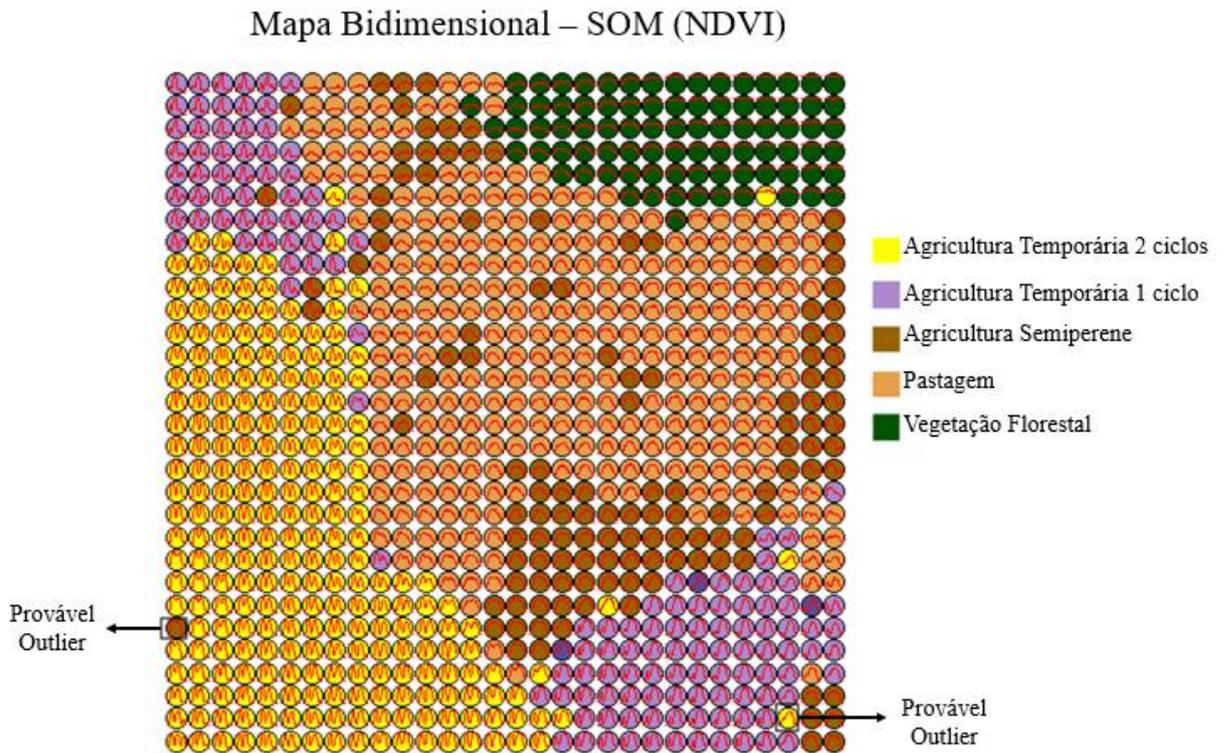
A Figura 9 mostra os padrões espectro-temporais de cada uma das classes temáticas, referentes às amostras de treinamento obtidas através do mapa TerraClass, ano base 2020. As curvas representam a média de cada uma das bandas, e foram utilizadas para uma análise inicial sobre o banco de dados. Por exemplo, a classe de Vegetação Florestal (VF) apresenta valores de NDVI elevados e constantes ao longo do ano. A Agricultura Temporária de 2 ciclos (AGR-2), apresenta os dois ciclos característicos da fenologia de uma cultura praticada no período de safra e outra praticada no período da safrinha no estado de Mato Grosso. As classes de Agricultura Semiperene (AGR-S) e Pastagem (PAST) apresentam comportamento semelhante, porém diferem no período de pico de NDVI. No caso da Agricultura Temporária de 1 ciclo (AGR-1), é possível observar que não há uma definição clara do ciclo fenológico a partir da média dos índices, cuja causa pode ser devida à alta variabilidade das datas de plantio ou duração do ciclo das culturas praticadas. Neste sentido, essa classe foi identificada como sendo uma provável fonte de variabilidade intraclasse, exigindo maiores cuidados na análise feita pelo SOM.



**Figura 9:** Padrões espectro-temporais das classes temáticas. Curvas com valores referentes à média de cada banda.

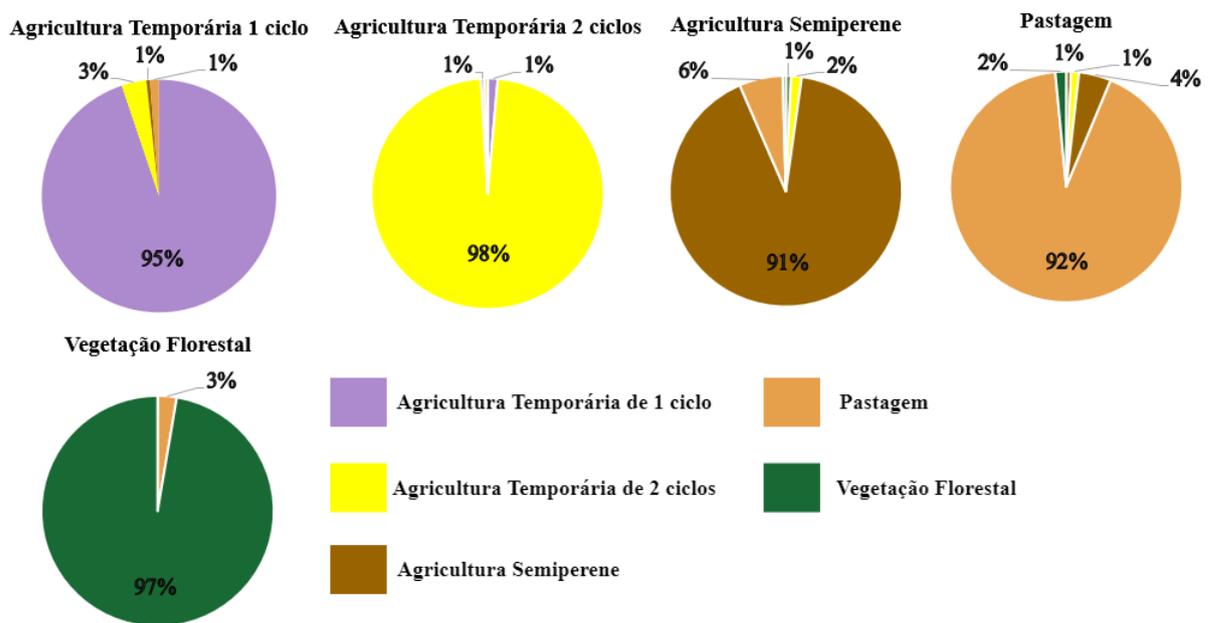
## 5.2 Qualidade das Amostras– SOM

A Figura 10 apresenta o mapa bidimensional resultante do SOM aplicado ao conjunto de dados de treinamento. No mapeamento é possível observar os agrupamentos correspondentes as classes temáticas. As classes AGR-2 e VF, apresentaram um agrupamento bem definido e distantes entre si. Para a AGR-1 foram formados dois grupos diferentes, justificando a ausência de um padrão bem definido para a classe, conforme observado na Figura 9. A classe AGR-S tem dois agrupamentos próximos e diversos neurônios distribuídos, principalmente, próximos a neurônios de PAST. No caso da PAST, seu agrupamento é o único que apresenta vizinhança com todos os demais agrupamentos, sua análise exigiu maior atenção pois ela possuía a chance de apresentar amostras semelhantes com todas as demais classes.



**Figura 10:** Mapa bidimensional gerado pelo SOM as com amostras de treinamento. Em destaque, exemplos de possíveis outliers identificadas por interpretação visual.

Durante o agrupamento das amostras dentro dos neurônios, amostras de diferentes classes podem estar juntas na mesma unidade, contanto que apresentem semelhanças entre si. Padrões semelhantes em diferentes classes podem acontecer e conseqüentemente causar confusões nos algoritmos de classificação. Por isso, foi realizada uma análise de pureza dos neurônios para verificar a quantidade de amostras de classes diferentes dentro dos agrupamentos. O resultado desta análise está disponível na Figura 11.

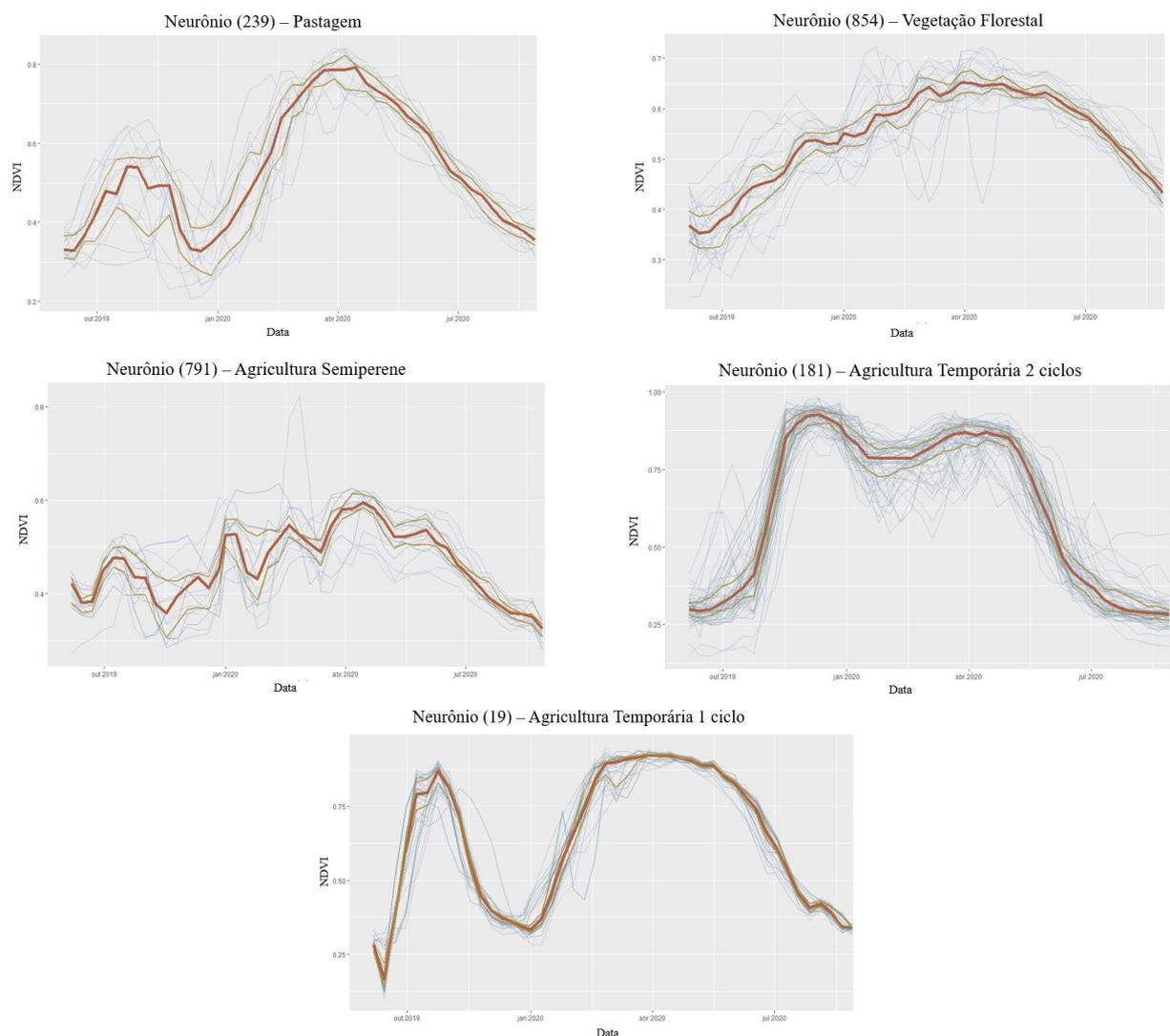


**Figura 11:** Gráficos frequência relativa referente a pureza dos neurônios de casa classe temática.

Os agrupamentos das cinco classes apresentaram porcentagens de pureza superiores a 90%. As menores purezas registradas foram para AGR-S (91%) e PAST (92%). Houve confusão entre essas duas classes, com a maior porcentagem de impurezas em neurônios de PAST confundindo com amostras de Semiperene (4%), e vice-versa, com uma impureza de 6% de amostras de PAST em neurônios de AGR-S. Essa análise concluiu que ruídos estavam presentes nas amostras de treinamento, o que justificou a necessidade deste processo de filtragem de amostras.

No mapa bidimensional, a primeira análise foi a interpretação visual para identificação de neurônios com amostras incoerentes. Em destaque na Figura 10, dois neurônios estão distantes dos agrupamentos de suas respectivas classes e, portanto, apresentam alta chance de apresentarem amostras inconsistentes. Os neurônios exemplificados continham amostras com PC menor que 60% e, portanto, foram excluídas do banco de dados.

Após a exclusão automática realizada pela PC, os demais neurônios foram avaliados seguindo as condições citadas na metodologia. Na Figura 12, estão disponíveis exemplos de neurônios analisados, cujos perfis espectrais foram considerados como ruídos. O neurônio 791, quando comparado ao padrão da Figura 9, não exibe um perfil espectral consistente com a classe AGR-S, principalmente devido à falta de um pico no índice de vegetação próximo ao mês de abril, período correspondente ao início do ciclo da cana-de-açúcar no estado. Neurônios com esse perfil foram excluídos por não representarem adequadamente a classe. Situação semelhante acontece com o neurônio 854, que representa a VF.



**Figura 12:** Exemplos de neurônios com perfis espectrais interpretados como ruídos dentro de suas respectivas classes temáticas.

No caso do neurônio 181, correspondente à AGR-2, era esperado a ocorrência de um “vale” (redução/incremento de NDVI) ao final do mês de janeiro ou durante o mês de fevereiro, representando a colheita da safra e o início do plantio da safrinha, comportamento que não foi observado nessas amostras. Nesse período, com auxílio da plataforma SATVeg foi identificada a presença de nuvens na região onde essas amostras foram coletadas. Devido a correção feita pela banda CLOUD, os dois picos de NDVI característicos da AGR-2 foram mascarados, por isso neurônios com características semelhantes ao neurônio 181 foram excluídos, apesar de pertencerem a classe temática.

Dentro das amostras de treino, também foram observados neurônios com curvas espectro-temporais que correspondem a outras classes temáticas. Por exemplo, o neurônio 19 (Figura 12) é referente à classe de AGR-1, mas seu comportamento apresenta características da AGR-2, com ciclos de safra e safrinha evidentes. Para evitar a perda dessas amostras, optou-se pela mudança da classe temática que melhor se adequa a suas características. Situações semelhantes aconteceram com a classe PAST, onde o neurônio 239 exemplifica um caso em que houve a transição de amostras de pastagem para agricultura de um ciclo.

### 5.3 Análise após Remoção de Ruídos

Após o tratamento das amostras por meio do SOM, o conjunto de treino passou a ter 31.812 observações. Considerando todas as classes, 8.188 pontos foram retirados das amostras de treino, valor que corresponde a aproximadamente 20% do total (Tabela 3). O conjunto resultante foi denominado de “Amostragem-SOM” (AS), enquanto o conjunto sem alterações recebeu o nome de “Amostragem Original” (AO). Para verificar a distribuição das amostras após a remoção de ruídos, uma nova distribuição foi feita pelo SOM com base na AS, com o objetivo de compará-lo com mapeamento da AO. Para facilitar a comparação, o novo mapeamento seguiu os mesmos padrões (grid 30x30 e cálculo com distância Euclidiana).

Classe Temática	Amostragem - Original (AO)	Amostragem - SOM (AS)
Agricultura Temporária 1 ciclo	5.000	1.966
Agricultura Temporária 2 ciclos	10.000	8.066
Agricultura Semipereene	5.000	4.289
Pastagem	10.000	8.025
Vegetação Florestal	10.000	9.466
<b>Total</b>	<b>40.000</b>	<b>31.812</b>

**Tabela 3:** Tabela referente à distribuição do número de amostras em cada classe temática, antes e após análise feita pelo SOM.

Após o SOM, as amostras de AGR-1 continuaram a apresentar dois agrupamentos distintos e distantes entre si, e em menor número, resultado do processo de remoção de ruído e a transição de amostras para a AGR-2. A Figura 13 exemplifica dois neurônios referentes aos agrupamentos observados. Ambos representam apenas um ciclo de agricultura durante o ano, mas em períodos diferentes, ocasionando a distância entre os agrupamentos. As amostras destacadas na Figura 13 comprovaram a existência de variações intraclasse, que foram inicialmente identificadas pela análise exploratória de dados. Apesar de diferentes, ambos os padrões foram considerados fundamentais para representar a classe temática dentro do estado.

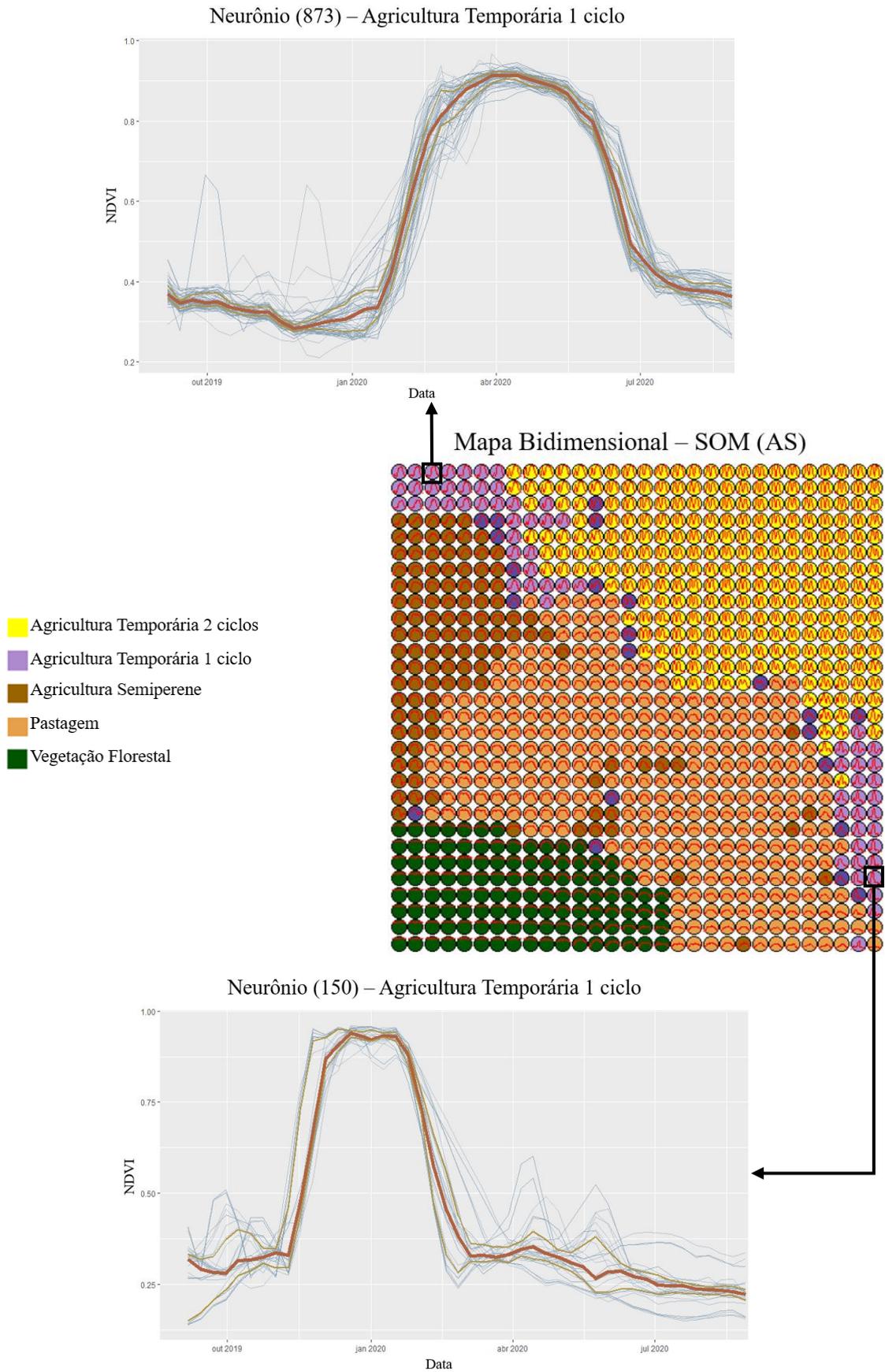
O agrupamento dos neurônios de AGR-2 foi semelhante ao mapeamento da AO. Dentro deste agrupamento, variações referentes ao número de ciclos podem ser encontradas (Figura 14). O neurônio 900 exemplifica o padrão mais recorrente, onde pode ser observada a presença de dois ciclos de safra e safrinha bem definidos. O neurônio 478 apresenta uma variação desse sistema de produção, em que pode ser observado um terceiro ciclo de produção. A presença de um terceiro ciclo é um exemplo do processo de intensificação agrícola do estado, por isso, essa variação foi mantida no conjunto de treinamento.

No caso da AGR-S, a distribuição da OS passa a apresentar apenas um agrupamento, e mais definido do que o da AO. Observa-se que, após a aplicação do SOM, a quantidade de neurônios dispersos no agrupamento de pastagem diminuiu, e os neurônios restantes, mesmo distantes de seu agrupamento, foram mantidos devido a variações durante a produção. A Figura 15 mostra exemplos de variações dentro da classe AGR-S. No neurônio 392, observa-se uma queda no NDVI após o mês de abril, correspondendo ao corte para a colheita da cana-de-açúcar. Já no neurônio 222, o mesmo padrão não é observado, pois a queda do NDVI acontece de forma gradual ao longo dos meses. Diferenças no manejo da cultura podem acontecer num estado com as dimensões do Mato Grosso, por isso o padrão do neurônio 222 não foi considerado como um ruído.

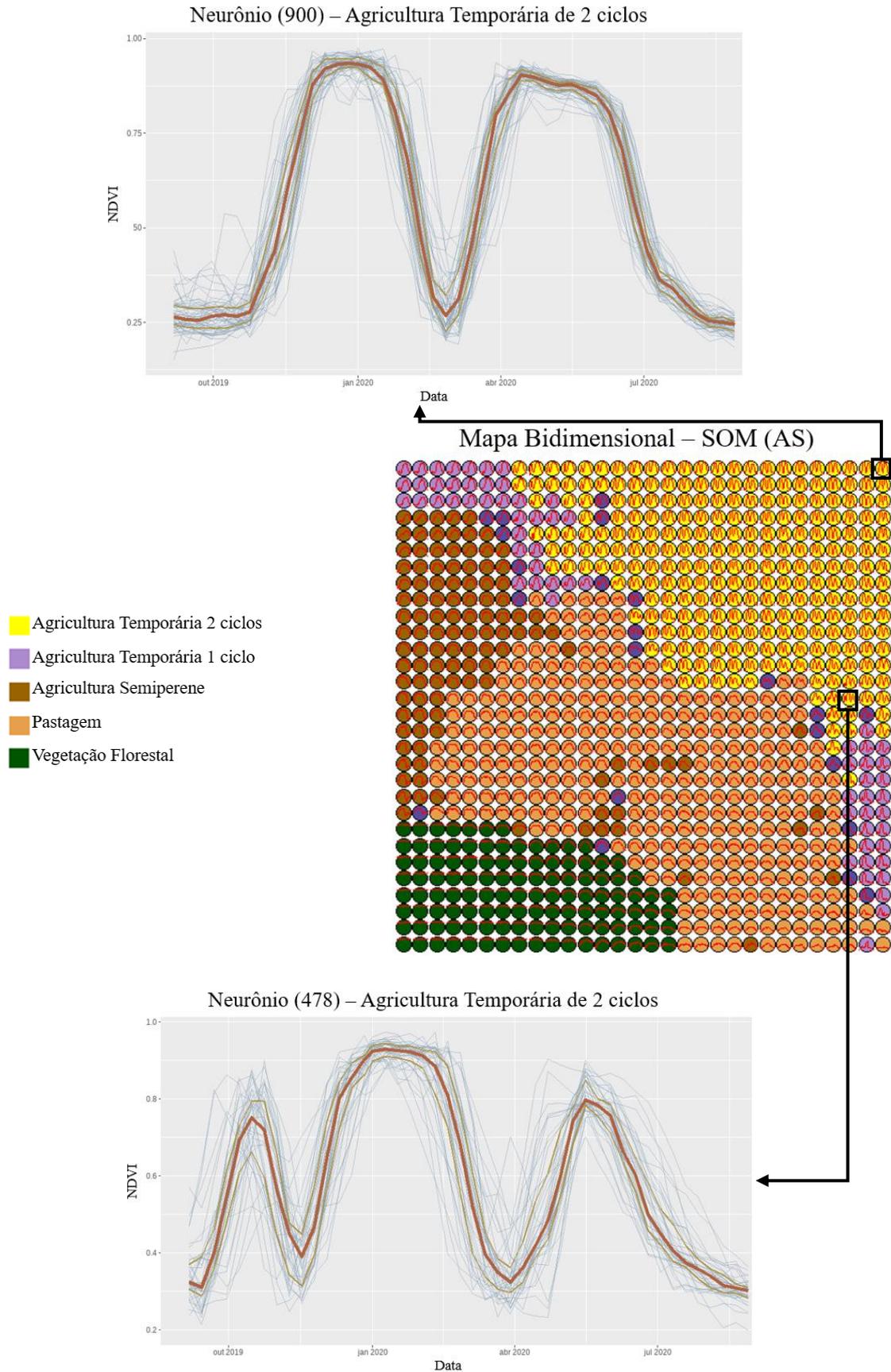
No caso da PAST, a padronização dos neurônios é menor, resultando em um agrupamento mais diversificado. Além disso, a classe continua sendo a única que apresenta vizinhança com as demais classes nos dois mapas bidimensionais analisados. Porém, dentro de seu agrupamento na AS há menor quantidade de neurônios das demais classes, principalmente a AGR-S. Como áreas de pasto estão espalhados por todo o estado, a diversidade de neurônios de pastagem foi interpretada como normal e necessária para representar a classe.

A Figura 16 exemplifica dois padrões distintos dentro da classe PAST. As amostras do neurônio 49 estão localizadas na região Leste do estado de Mato Grosso, onde está situado o Parque Nacional do Araguaia, uma região de preservação ambiental. Por isso, apresenta um perfil constante ao longo do tempo, indicando, provavelmente, baixa interferência humana. Por outro lado, as amostras do neurônio 394 estão localizadas na região central do estado, onde se concentram a maior parte das áreas de produção agrícola, conforme mapeamento do TerraClass. O perfil caracteriza um maior desenvolvimento da vegetação, com características semelhantes ao neurônio 392 (Figura 15), referente a produção de cana-de-açúcar. Em ambos, é possível observar o crescimento do NDVI a partir de outubro de 2019, com pico próximo a 0,8 e constante até o mês de abril de 2020. Devido a essas semelhanças, os dois neurônios estão próximos entre si no mapa bidimensional, porém diferem no momento da colheita da cana-de-açúcar.

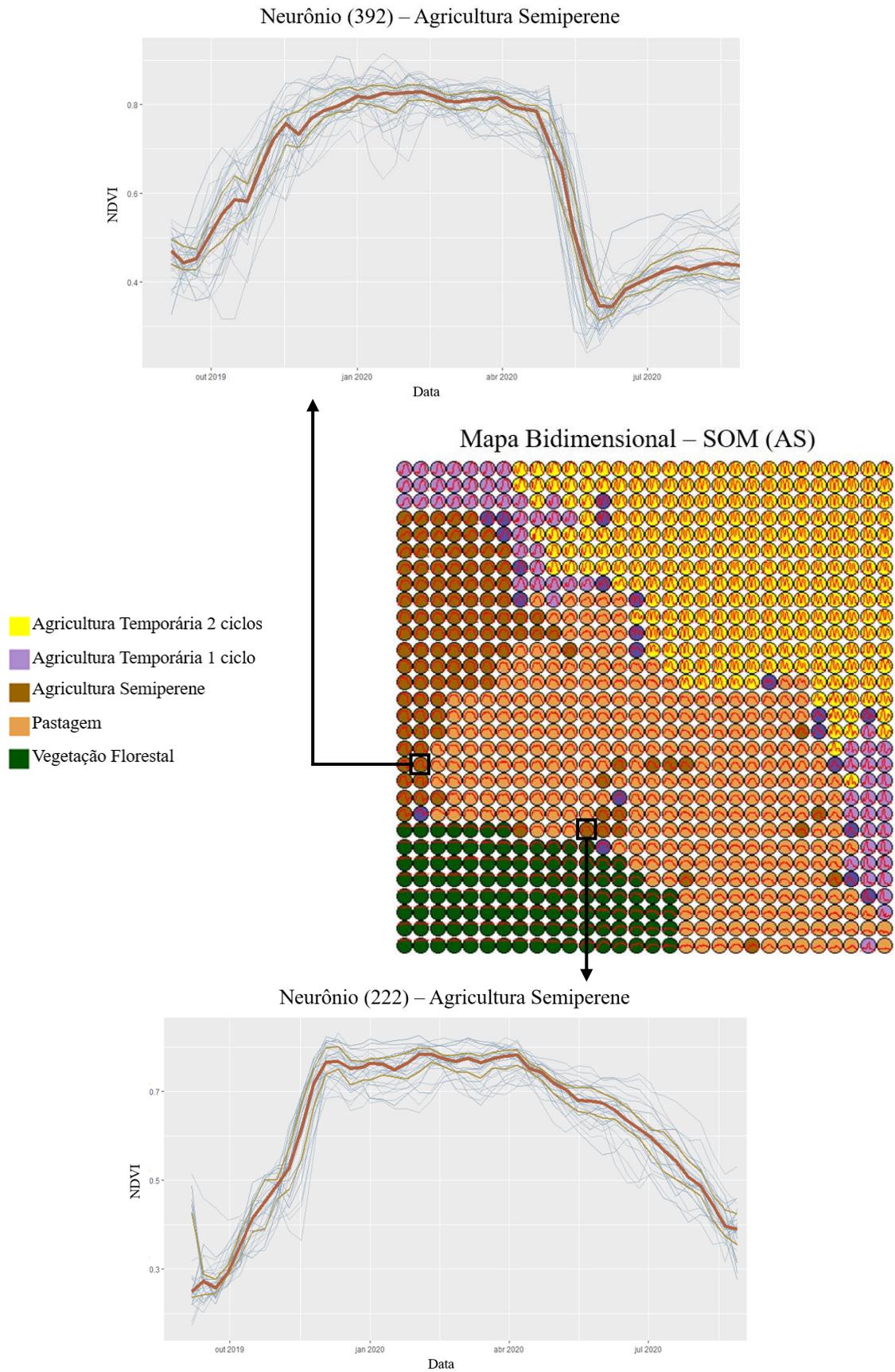
Na classe VF o agrupamento manteve o padrão, com poucas alterações em relação ao conjunto original. Apesar de possuir o maior número de amostras, suas características permanecem consistentes, resultando em um agrupamento menor se comparado as demais classes. A Figura 16, exemplifica os dois tipos de padrões de Vegetação Florestal disponíveis. O neurônio 1 representa uma formação florestal com maior densidade de vegetação, com ciclo de NDVI praticamente constante e que ocorre principalmente no bioma Amazônia. O neurônio 44 representa uma vegetação em estado de recuperação, no mapeamento TerraClass este perfil recebe o nome de Vegetação Natural Secundária (Tabela 1).



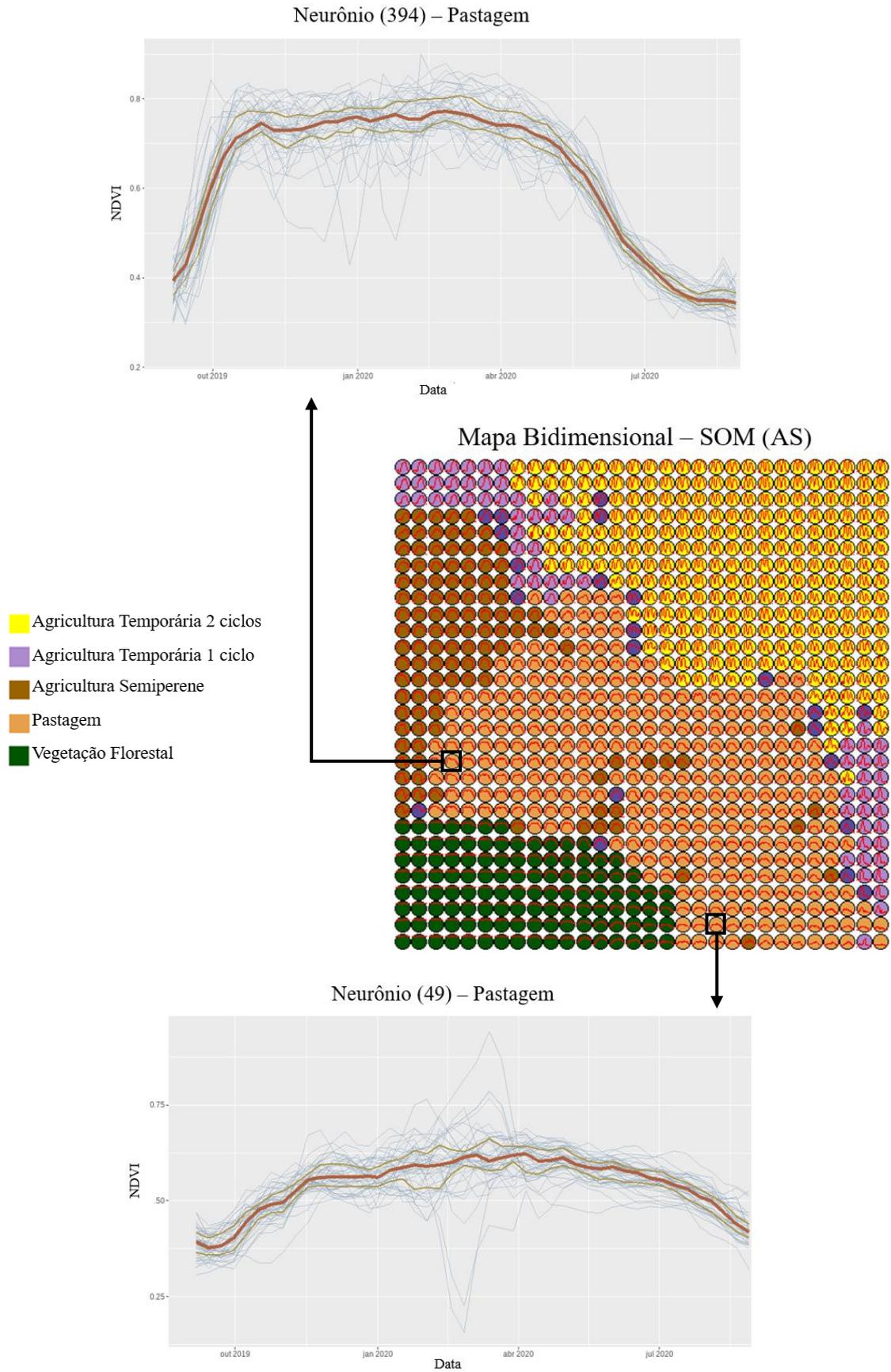
**Figura 13:** Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Temporária de 1 ciclo.



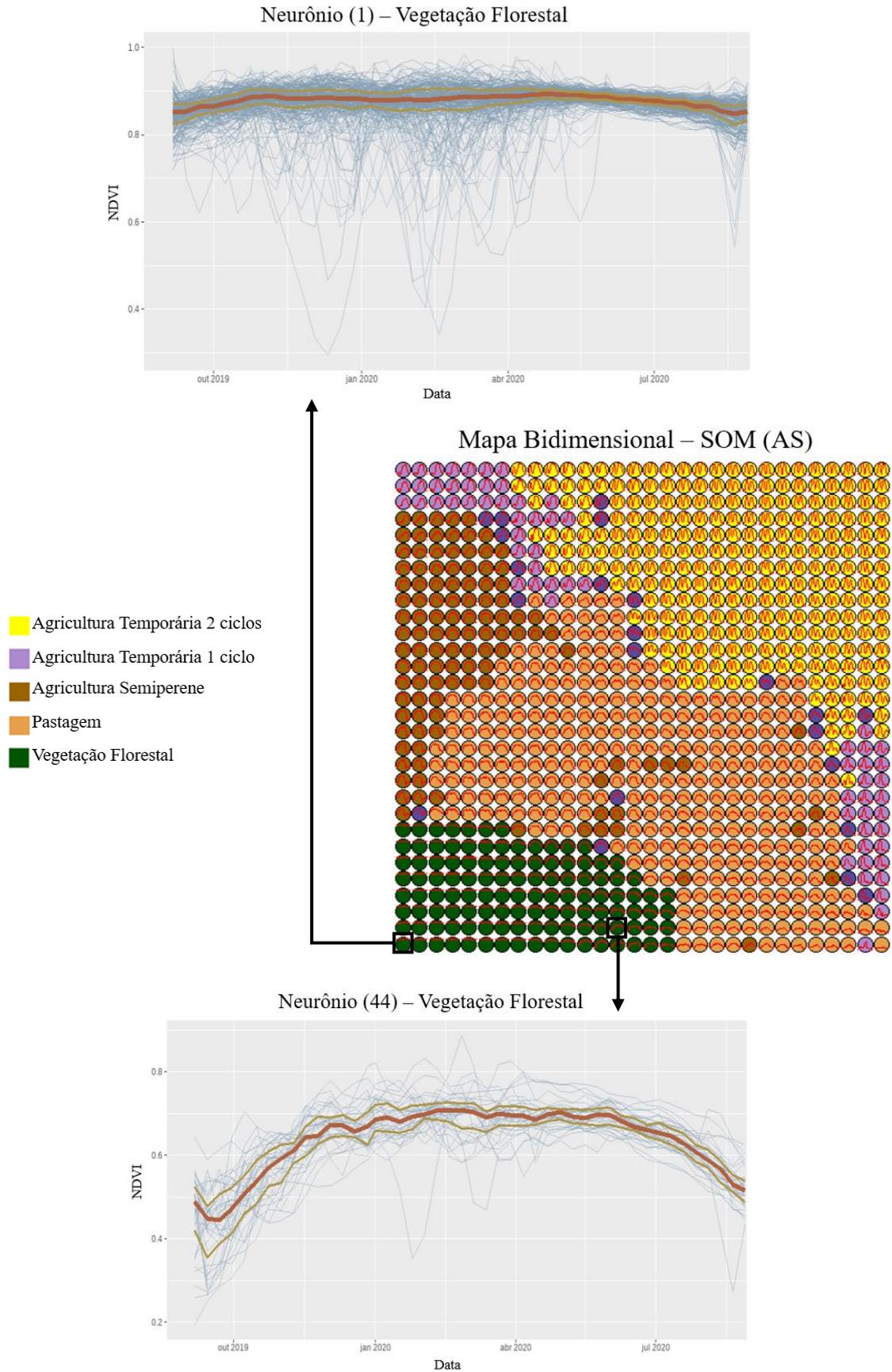
**Figura 14:** Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Temporária de 2 ciclos.



**Figura 15:** Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Agricultura Semiperene.



**Figura 16:** Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Pastagem.



**Figura 17:** Mapa Bidimensional referente ao SOM aplicado no conjunto AS (Amostragem SOM), exemplificando a variabilidade em amostras de Vegetação Florestal.

#### 5.4 Resultados do Modelo para o Ano Base 2019/2020

Os resultados obtidos pelos modelos de classificação para o ano-safra 2019/2020 são apresentados na Tabela 4. A acurácia global foi calculada utilizando 20% do conjunto total de amostras. Todos os três algoritmos utilizados apresentaram altos valores de acurácia global, tanto antes quanto após o tratamento de ruídos com o SOM.

Conforme observado na Tabela 3, existem classes com maior volume de dados, porém as diferenças não foram consideradas suficientes para influenciar nos resultados. Como exemplo, a Tabela 5 representa a matriz de confusão obtida pelo RF ao ser treinado pela AO. Todas as classes apresentaram altos valores de acurácia do produtor (baixo erro de omissão) e acurácia do usuário (baixo erro de comissão), justificando o alto valor da acurácia global. Os demais modelos apresentaram resultados na matriz de confusão semelhantes ao disponibilizado na Tabela 5.

Resultados com boa acurácia eram esperados para os três classificadores, visto que todos têm eficácia testada e comprovada em demais trabalhos de classificação e do uso e cobertura da terra (BELGIU; DRĂGU, 2016; SAINTE FARE GARNOT et al., 2020). Outros trabalhos compararam os resultados de algoritmos baseados em estatística com os obtidos por algoritmos de aprendizado profundo, autores como ABIDI et al. (2023) e FIRAT et al. (2022) verificaram que redes neurais geralmente conseguem resultados melhores devido à melhor capacidade de interpretação dos dados, principalmente as redes CNN (KATTENBORN et al., 2021). No entanto, nesta tarefa, foram registrados valores de AG de 0,96 e sem grandes diferenças entre os algoritmos.

Um provável motivo para este resultado é o número de classes utilizadas. Neste trabalho, existem apenas cinco classes temáticas, e que apresentam características distintas entre si, conforme observado na distribuição de neurônios feita pelo SOM (Figura 10). Neste sentido haviam menos obstáculos que poderiam levar a confusão dos algoritmos. Por exemplo, ZHONG; HU; ZHOU (2019) realizaram uma tarefa que envolvia a classificação de 16 classes temáticas, algumas delas apresentaram maior dificuldade de classificação, por isso seus resultados obtidos estiveram na faixa de 0,8 até 0,85. Além disso, neste trabalho, variações de perfis espectro-temporais intraclasse foram registrados e vistos como possíveis fontes de erros para os algoritmos, porém essas variações representaram uma porção minoritária do conjunto de amostras, por isso para ano base 2019/2020 essas variações não prejudicaram o desempenho dos classificadores.

Modelo	Amostragem Original	Amostragem SOM
Random Forest	0,978	0,994
TempCNN	0,979	0,981
L-TAE	0,979	0,962

**Tabela 4:** Tabela referente aos valores de acurácia global obtidos para os seis cenários de treinamento do modelo de classificação no base 2019/2020.

Random Forest (AO) - Modelo 2019/2020						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semipereene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1885	9	8	9	0	0,986
Agricultura Temporaria 1 ciclo	17	1012	1	8	0	0,975
Agricultura Semipereene	4	1	955	14	2	0,978
Pastagem	22	6	19	1997	29	0,963
Vegetacao Florestal	0	0	0	27	1975	0,987
Recall	0,978	0,984	0,972	0,972	0,985	0,978

**Tabela 5:** Matriz de confusão referente aos resultados obtidos pelo classificador RF na Amostragem Original.

Ao serem treinados com os dados da AS, cada algoritmo teve um comportamento diferente. O RF apresentou uma pequena melhora no seu resultado, enquanto que a TempCNN manteve um resultado semelhante e a L-TAE apresentou uma queda. Apesar dos valores e as pequenas diferenças encontradas, os resultados obtidos continuam semelhantes entre si com variação máxima de 0,022 na acurácia global, por isso não houve conclusões em relação a diferença de cada algoritmo e a efetividade do tratamento de ruídos. No entanto, foi verificado que a metodologia adotada cumpre um dos objetivos específicos deste trabalho, pois permitiu a criação de diferentes modelos base para ano-safra de 2019/2020 e que apresentam boa acurácia.

## 5.5 Resultados Generalização

### 5.5.1 Validação com Amostras de Campo

A distribuição do número de amostras por classe foi apresentada na Tabela 2. Os modelos gerados pelos três algoritmos de classificação, nos dois cenários de amostragem, foram testados para avaliar sua capacidade de Generalização utilizando essas amostras. Os resultados obtidos de acurácia global de cada modelo estão disponíveis na Tabela 6.

	RF-AO	RF-AS	TempCNN AO	TempCNN AS	L-TAE AO	L-TAE AS
<b>Acurácia Global</b>	0,889	0,917	0,858	0,886	0,858	0,859

**Tabela 6:** Tabela referente aos resultados de acurácia global de cada modelo na etapa de teste com as amostras de campo.

Diferente do observado na validação para ano-safra 2019/2020, na Generalização, os algoritmos apresentaram resultados distintos, além de mostrarem uma resposta diferente ao tratamento de ruídos. Dentre os três classificadores, o RF apresentou os melhores resultados de AG nos dois cenários testados, com 0,889 para a AO e 0,917 para a AS. O RF também respondeu de forma positiva ao tratamento de ruídos, apresentando uma melhora na classificação. Este foi o único cenário em que a acurácia global na Generalização foi superior a 0,9.

Os dois classificadores baseados em redes neurais apresentaram a mesma acurácia global (0,858) quando treinados utilizando a AO. No entanto, ao serem treinados com a AS, a TempCNN apresentou melhora na sua classificação, passando a ter uma acurácia de 0,886. No caso da L-TAE, os valores de acurácia dos dois cenários testados não apresentaram alterações significativas. Para realizar uma análise mais detalhada dos resultados, foram utilizadas matrizes de confusão referentes aos modelos dos três algoritmos e nos diferentes cenários (Apêndice A). As métricas extraídas foram a *Precision* (Tabela 7), *Recall* (Tabela 8) e *F1-Score* (Tabela 9).

<b>Precision</b>	RF AO	RF -AS	TempCNN AO	TempCNN AS	L-TAE AO	L-TAE AS
<b>Agricultura Temporaria 2 ciclos</b>	0,951	0,968	0,938	0,966	0,956	0,960
<b>Agricultura Temporaria 1 ciclo</b>	0,956	0,995	0,946	0,954	0,901	0,922
<b>Agricultura Semipereene</b>	0,938	0,906	0,869	0,874	0,741	0,791
<b>Pastagem</b>	0,773	0,811	0,679	0,744	0,750	0,697
<b>Vegetacao Florestal</b>	0,906	0,903	0,929	0,895	0,903	0,918

**Tabela 7:** Tabela referente aos resultados da métrica *Precision* dos modelos de classificação para cada classe temática.

Recall	RF AO	RF -AS	TempCNN	TempCNN	L-TAE	L-TAE
			AO	AS	AO	AS
Agricultura Temporaria 2 ciclos	0,965	0,977	0,965	0,954	0,951	0,926
Agricultura Temporaria 1 ciclo	0,838	0,853	0,837	0,877	0,880	0,852
Agricultura Semiperene	0,839	0,894	0,650	0,783	0,697	0,655
Pastagem	0,913	0,893	0,887	0,867	0,813	0,895
Vegetacao Florestal	0,897	0,942	0,866	0,907	0,889	0,913

**Tabela 8:** Tabela referente aos resultados da métrica *Recall* dos modelos de classificação para cada classe temática.

F1-Score	RF AO	RF -AS	TempCNN	TempCNN	L-TAE	L-TAE
			AO	AS	AO	AS
Agricultura Temporaria 2 ciclos	0,958	0,972	0,951	0,960	0,956	0,944
Agricultura Temporaria 1 ciclo	0,893	0,918	0,889	0,914	0,901	0,903
Agricultura Semiperene	0,886	0,900	0,743	0,826	0,741	0,753
Pastagem	0,837	0,850	0,769	0,801	0,750	0,766
Vegetacao Florestal	0,901	0,922	0,896	0,901	0,903	0,898

**Tabela 9:** Tabela referente aos resultados da métrica *F1-Score* dos modelos de classificação para cada classe temática.

Os valores de *F1-Score* obtidos por cada classificador justificam por que o RF apresentou melhor acurácia do que os demais e demonstram que a diferença entre eles foi registrada principalmente nas classes de Agricultura Semiperene (AGR-S) e Pastagem (PAST). Para a AGR-S, os melhores resultados de *F1-Score* foram obtidos após o SOM, sendo 0,9 para o RF, 0,826 para a TempCNN e 0,753 para a L-TAE. Comparado à L-TAE, a TempCNN obteve melhores resultados, pois apresentou menor erro de comissão (*Precision*). O *Recall* dos dois classificadores foram os menores registrados em todas as classes temáticas, no caso da L-TAE o maior valor foi de 0,697.

Analisando as matrizes de confusão, a principal causa dos baixos valores de *Recall* na classe AGR-S foi a confusão com a classe PAST. Durante a análise exploratória de dados e ao longo do processo de identificação de ruídos, foi observada a semelhança entre perfis espectrais de AGR-S e PAST, conforme os exemplos das Figuras 15 e 16. Essa semelhança não foi um fator limitante na classificação de amostras do mesmo ano-safra, porém, ao generalizar esses perfis, a semelhança entre as classes pode ter causado erros de classificação.

Outra possível fonte de erro, foi a variabilidade presente dentro da classe AGR-S. A Figura 15 exemplificou uma das variações dentro desta classe, porém observa-se que a quantidade de neurônios que representam essas variações é inferior aos demais. Neste sentido, a dificuldade encontrada pela TempCNN e L-TAE em generalizar esta classe pode ter sido causada por falta de amostras de treino. Por exemplo, JOHNSON (2019) alertou que a escassez de dados para representar características referentes às variações intraclases pode prejudicar a Generalização. No caso do RF, o algoritmo é conhecido por apresentar bom desempenho mesmo com classes desbalanceada, justificando os melhores resultados na Generalização da AGR-S (BELGIU; DRĂGU, 2016).

Os resultados de *F1-Score* para Agricultura Temporária de 1 ciclo (AGR-1) foram semelhantes para os três algoritmos. RF e TempCNN apresentaram melhoras após a remoção de ruídos, enquanto a L-TAE manteve os resultados. No caso do RF, a melhora no *F1-Score* está relacionada ao valor da *Precision*, que atinge o valor de 0,995. Este é o menor índice de falsos positivos entre todas as classes e cenários analisados. Ao analisar o *Recall*, o RF apresentou os menores valores em comparação com as redes neurais. Conforme os exemplos da Figura 13, a AGR-1 apresenta variações em que o ciclo da cultura está deslocado no tempo, ou seja, há variações na época de plantio e colheita. Neste sentido, os algoritmos TempCNN e L-TAE, que são capazes de interpretar sequências de dados apresentaram melhores resultados de *Recall* do que o RF.

A Agricultura Temporária de 2 ciclos (AGR-2) foi uma classe que apresentou bons resultados com os três algoritmos. Conforme o exemplo apresentado na Figura 12, foram removidas pelo SOM as amostras onde a correção da banda CLOUD mascarou o período referente ao final da safra e início da safrinha. Ao analisar os resultados, observa-se que a presença dessas amostras não interferiu no desempenho dos classificadores, pois o processo de classificação dentro do pacote *sits* foi capaz de interpretar e considerar a correção de nuvens durante a classificação. Nesse sentido, a melhora do *F1-Score* após o SOM, registrada pelo RF e a TempCNN, esteve relacionada à menor confusão com a AGR-1. No caso da L-TAE, o processo do SOM piorou o resultado de *F1-Score* devido ao aumento da confusão com a classe PAST.

Os resultados para a classe Vegetação Florestal (VF) foram consistentes em todos os cenários. Por ser uma classe que engloba áreas de floresta, seu perfil apresenta características diferentes das demais, principalmente da agricultura, o que facilita sua classificação. Os erros registrados por essa classe foram majoritariamente devido à confusão com a PAST. No SOM, conforme exemplo da Figura 17, observou-se que algumas amostras de VF apresentavam um

perfil espectral que poderia se assemelhar à PAST. Essa semelhança aconteceu devido ao uso da classe Vegetação Natural Secundária (Tabela 1) na classe VF. Como a vegetação secundária é composta por uma vegetação em processo de recuperação, seus índices de NDVI podem se assemelhar a uma PAST, principalmente se a mesma apresentar predominância de vegetação arbórea.

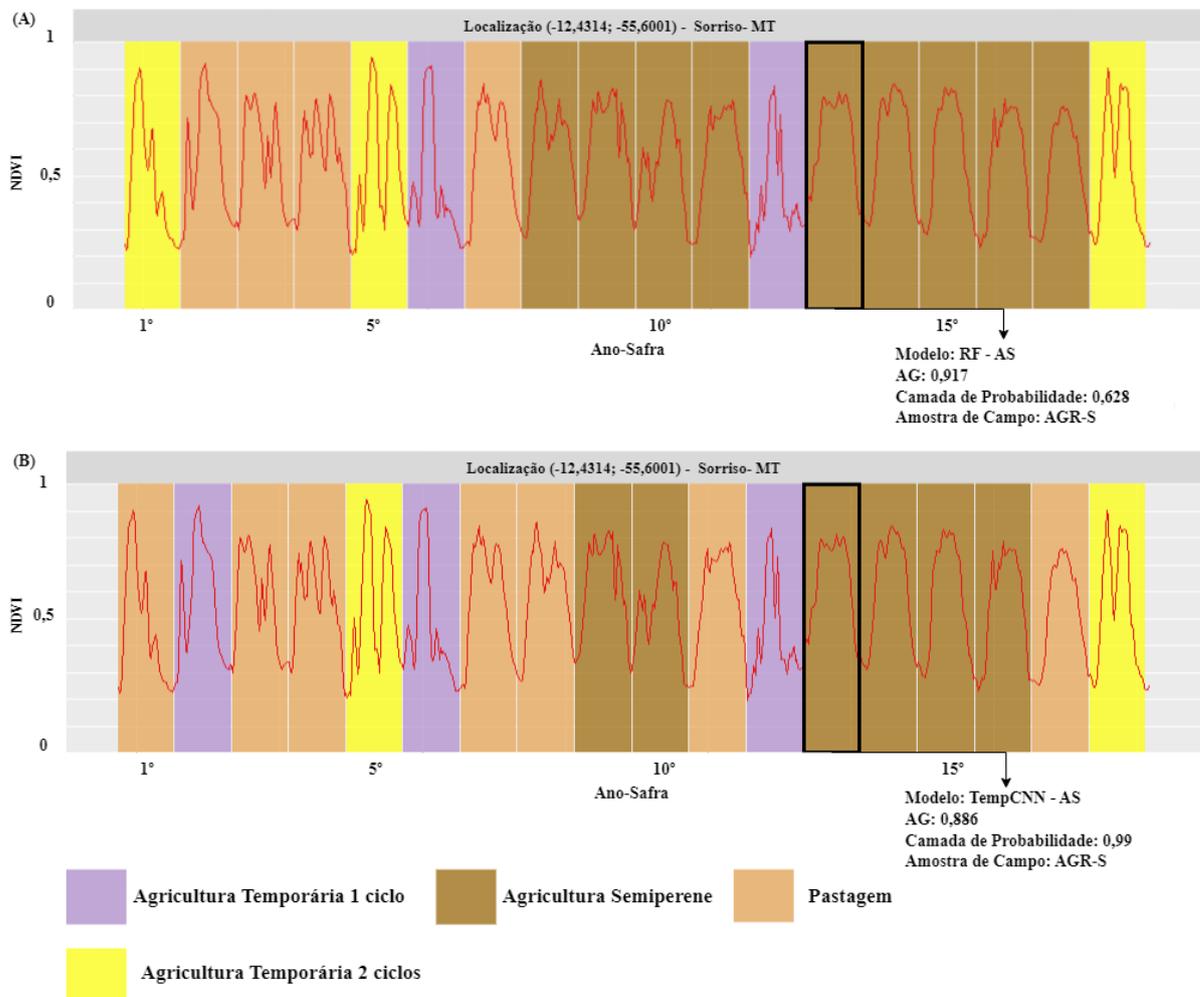
Com base nos resultados obtidos e nas análises realizadas, algumas observações podem ser feitas em relação ao uso do SOM neste trabalho. A primeira é que a redução de ruídos feita pelo SOM ajudou o RF e a TempCNN a melhorarem sua classificação. No caso, da TempCNN a melhora acontece principalmente na AGR-S, em que o *F1-Score* passa de 0,743 para 0,826. No caso da L-TAE não houve mudanças significativas nos seus resultados de acurácia global e *F1-Score*, porém variações significativas puderam ser observadas nas métricas *Precision* e *Recall*. Por exemplo, na AGR-S o *Recall* foi de 0,697 na AO e 0,655 na AS, e na PAST a *Precision* caiu de 0,75 (AO) para 0,697 (AS). Neste sentido, o SOM não proporcionou melhoras para a classificação da L-TAE, e por apresentar resultados inferiores aos demais, concluiu-se que neste trabalho os modelos baseados na L-TAE não apresentaram bom desempenho na Generalização.

Apesar da melhora apresentada após o tratamento de ruídos, os resultados da TempCNN não superaram os do RF. No teste sem o SOM, o RF demonstrou resultados superiores aos demais algoritmos, comprovando sua consistência mesmo com a presença de ruídos. Após a aplicação do SOM, a acurácia global do RF atingiu 0,917, o que foi o maior valor registrado por todos os modelos. A classe em que o RF enfrentou maior dificuldade de classificação foi a PAST, com um *F1-Score* de 0,85. Neste sentido, concluiu-se que essa foi a classe de maior dificuldade para a classificação, justificada pela variabilidade e semelhança com demais classes, identificadas pelo SOM.

O objetivo deste trabalho envolve utilizar a abordagem de Generalização de um modelo de classificação baseado no ano-safra 2019/2020. Os resultados obtidos nos anos onde haviam disponibilidade de amostras de campo, comprovaram a eficácia do modelo em realizar esse processo em três anos-safra diferentes. Foi observado também que o processo de remoção de ruídos por meio do SOM influenciou de forma positiva os algoritmos RF e TempCNN, sendo que o RF apresentou os resultados mais consistentes dentre todos os demais.

### 5.5.2 Classificação série temporal – 2004 até 2022

Os modelos treinados com o RF e a TempCNN apresentaram os melhores resultados nos testes com as amostras de campo, principalmente os modelos RF-AS e TempCNN-AS. Considerando que os resultados deste trabalho podem ser utilizados para a classificação ano a ano de uma série temporal mais extensa (Ex: Séries temporais disponibilizadas pelo SATVeg), foram realizadas análises visuais comparativas entre os dois algoritmos aplicados à classificação de séries temporais do sensor MODIS contendo 18 anos-safra, entre os anos de 2004 e 2022. Para comparar esses dois modelos, foram selecionados pontos das amostras de campo localizados nos municípios de Sorriso-MT (Figura 18) e Paranatinga-MT (Figura19).

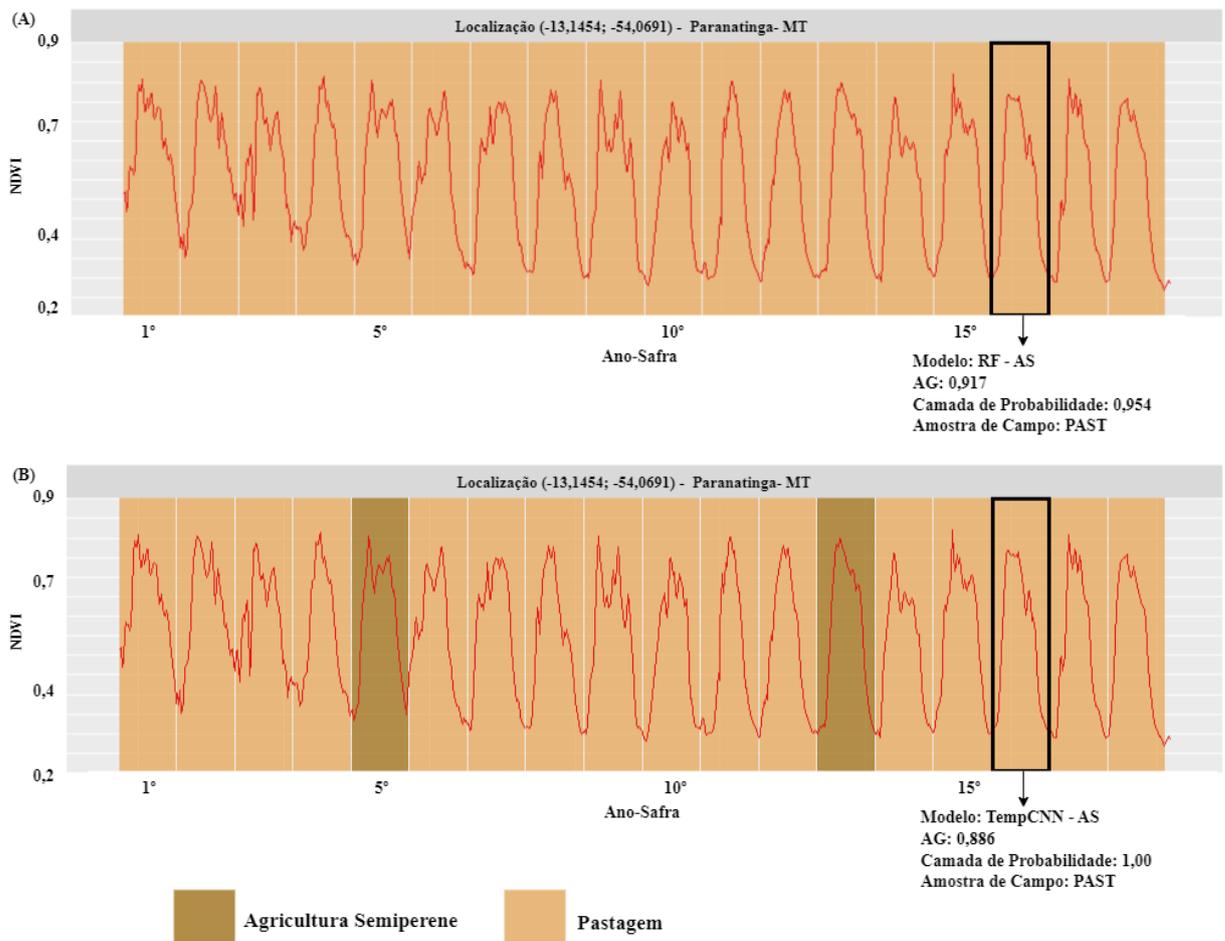


**Figura 18:** Figuras referentes a classificação interanual no município de Sorriso-MT resultado da Generalização feita com o modelo Random Forest (A) e TempCNN (B). Em destaque, o ano-safra correspondente a amostra de campo.

Na Figura 18, o período correspondente a amostra de campo é o 13º ano-safra (2016/17), em que os dois modelos classificaram corretamente o perfil como AGR-S. Porém, a TempCNN apresentou um valor de 0,99 na sua camada de probabilidade, e no RF o valor foi de 0,628. Esses valores representam a probabilidade de o perfil classificado pertencer a classe AGR-S. Ao analisar as camadas de probabilidade do modelo RF-AS (Apêndice B), o mesmo perfil também apresentou uma probabilidade de 0,244 de pertencer a classe PAST. No caso da TempCNN não ocorreu a distribuição das probabilidades entre outras classes temáticas. Essa diferença é explicada pelo funcionamento dos dois algoritmos. Na TempCNN, seu processo de aprendizado é baseado em redes neurais permitindo resultados com maior precisão (ISMAIL FAWAZ et al., 2019). No caso do RF, em que a classificação é feita por meio de um conjunto de árvores de decisão, a resposta do RF é decidida pelo resultado da maioria das árvores, por isso seus resultados apresentam maior variabilidade (BREIMAN et.al 2001).

Apesar de apresentar uma camada de probabilidade mais precisa, é importante destacar que esses resultados com alta precisão da TempCNN estão relacionados ao processo de treinamento da rede. Ou seja, os resultados das métricas obtidas pelos testes com amostras de campo devem ser considerados. Por exemplo, o RF apresenta uma acurácia global de 0,917 e um *F1-Score* superior ao TempCNN em todas classes temáticas, principalmente nas classes AGR-S e PAST.

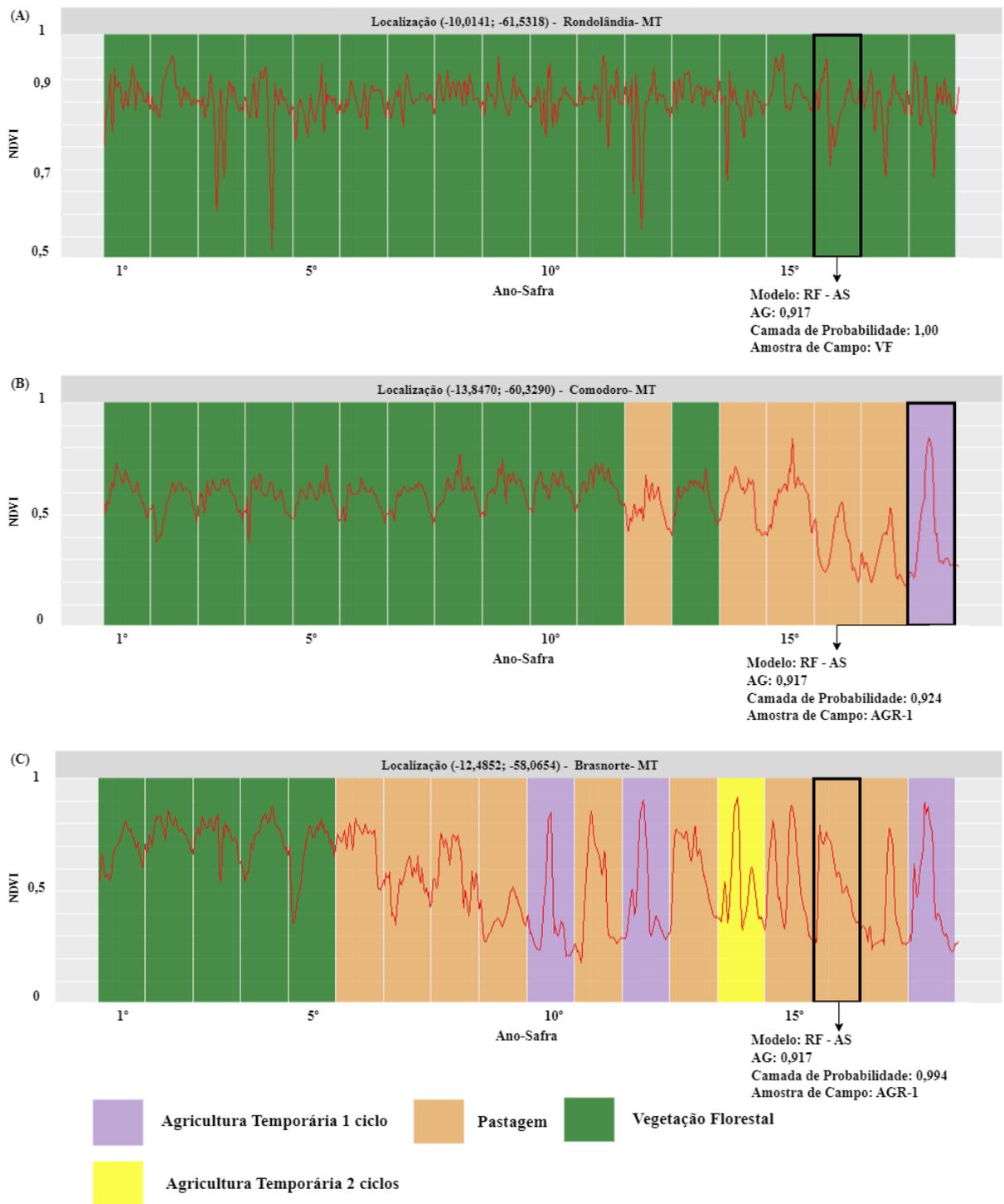
Nos perfis da Figura 18 existem períodos em que a classificação do RF e da TempCNN divergiram nos resultados. Por exemplo, os anos-safra 8 (2011/12), 11 (2014/15) e 17 (2020/21) divergiram em relação as classes AGR-S e PAST. Nesses anos não foi possível interpretar com precisão os resultados, pois não existem amostras de campo para comparar os resultados. Porém, considerando os resultados obtidos na etapa de teste, existe uma maior probabilidade de confusão na classificação feita pela TempCNN, apesar da maior precisão em sua camada de probabilidade.



**Figura 19:** Figuras referentes a classificação interanual no município de Paranatinga-MT resultado da Generalização feita com o modelo Random Forest (A) e TempCNN (B). Em destaque, o ano-safra correspondente a amostra de campo.

Na Figura 19, é exemplificado uma possível confusão entre as classes PAST e AGR-S. A TempCNN classificou como AGR-S os perfis dos anos-safra 5 (2008/09) e 13 (2016/17), enquanto o RF classificou os mesmos períodos como PAST. O perfil analisado está localizado no município de Paranatinga, localizado na região central do Mato Grosso. De acordo com dados do IBGE, o município não apresenta histórico de produção de cana-de-açúcar, por isso foi concluído que a classificação da TempCNN foi incoerente nesses dois perfis, mesmo que no ano-safra 13 (2016/17) a previsão da TempCNN indique uma probabilidade de 1,00 para a AGR-S (Apêndice B). Neste sentido, é maior a probabilidade de acerto do RF nos perfis divergentes dos anos-safra 8 (2011/12), 11 (2014/15) e 17 (2020/21) na Figura 18, mesmo o RF apresentando menor valor na camada de probabilidade. A partir destes exemplos foi possível observar inconsistências na classificação feita pela TempCNN, por isso para este trabalho, conforme verificado na etapa de teste, o modelo mais indicado para a classificação interanual foi o RF-AS.

Os pontos que foram utilizados para a classificação interanual por meio do modelo RF-AS e os resultados estão disponíveis nas Figuras 20, 21 e 22. Os pontos selecionados para a Generalização de 18 anos-safras são baseados em amostras de campo, e estão destacados em cada uma das séries temporais analisadas. Nesta etapa, a avaliação da classificação nos anos em que não haviam amostras de campo foi conduzida por meio de interpretação visual, em conjunto com o auxílio do SATVeg, identificação dos perfis temporais baseados nos padrões identificados na análise do SOM e a avaliação dos resultados obtidos nas camadas de probabilidade de cada classificação (Apêndice C)

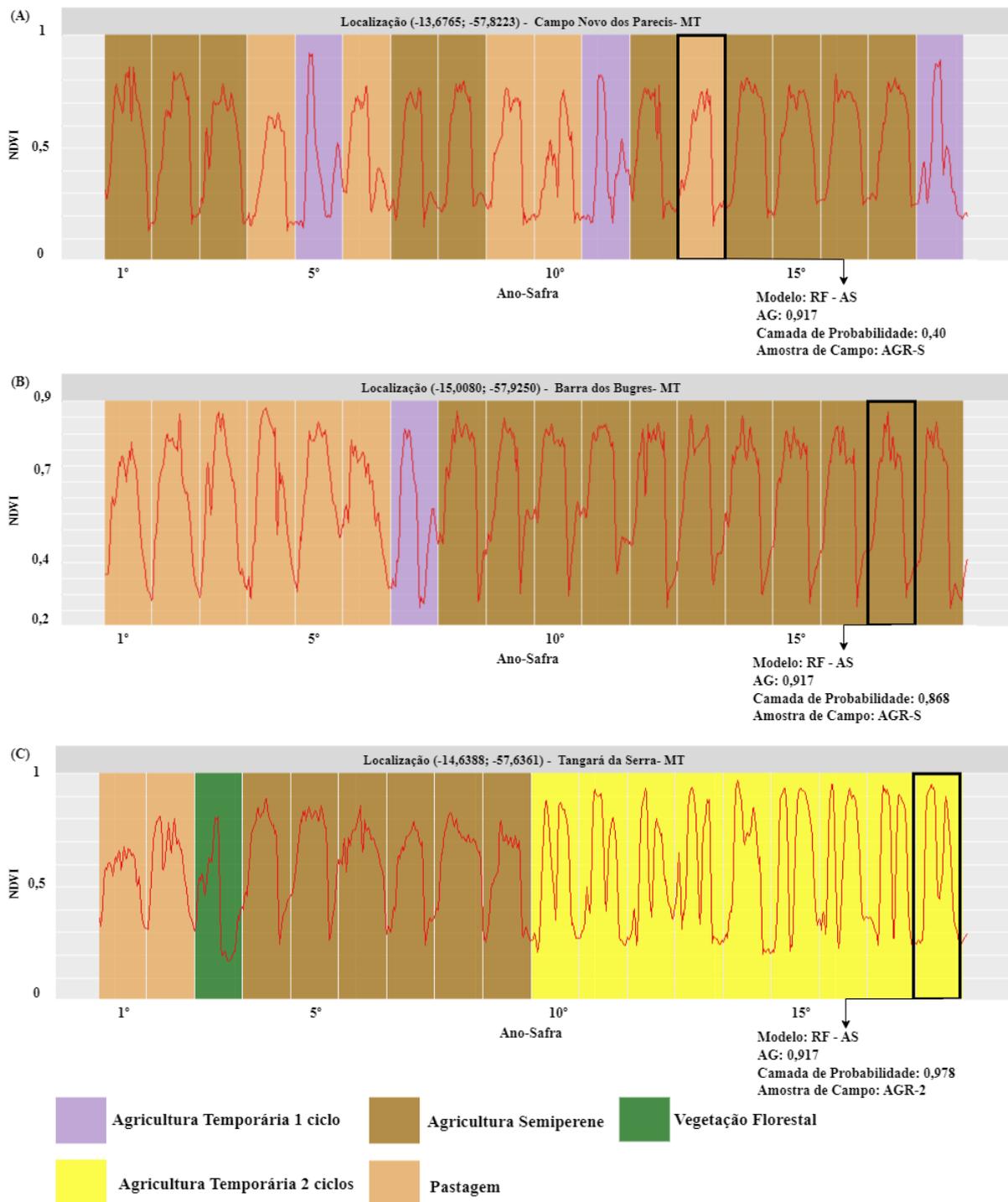


**Figura 20:** Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Rondolândia-MT (A), Comodoro-MT (B) e Brasnorte-MT (C). Em destaque, o ano-safra correspondente a amostra de campo.

Os perfis da Figura 20 são exemplos de perfis espectro-temporais de regiões com predominância da classe VF, apresentando pontos onde houve a transição de vegetação para cultivo agrícola. A Figura 20A, representa um perfil no município de Rondolândia, localizado dentro do bioma Amazônia, com predominância da classe VF. Conforme observado pela análise com o SOM, o perfil espectro-temporal de VF encontrado neste ponto é um perfil característico da classe no estado de Mato Grosso. Neste sentido, o algoritmo não encontrou dificuldades em classificar este tipo de perfil ao longo dos anos, não havendo incoerências nos resultados.

Na Figura 20B, o perfil espectro-temporal pertence a um ponto no município de Comodoro, região com predominância do bioma Cerrado. Nesse ponto, a classe VF apresenta menores valores de NDVI do que o observado na Figura 20A. O perfil de VF encontrado nessa região apresenta características semelhantes ao neurônio 44 (Figura 17), que representa uma variação dentro da classe VF e possui características próximas a classe PAST. Conforme apontado no processo de coleta das amostras de treino, a classe VF considerou a classe temática Vegetação Natural Secundária do mapeamento TerraClass (Tabela 1), que pode apresentar semelhanças com a pastagem, principalmente quando nesta há a predominância da vegetação arbórea. Neste sentido, a possível confusão apresentada entre PAST e VF não foi considerada como incoerente, pois o perfil encontrado nessa região representa uma transição entre as duas classes. Na Generalização, destaca-se que este ponto foi utilizado na etapa de teste como uma amostra de AGR-1 no ano-safra 18 (2021/22). O classificador foi capaz de identificar esse perfil, classificá-lo corretamente de acordo com a amostragem de campo e generalizar a classificação em comportamentos distintos nos anos anteriores.

No perfil apresentado na Figura 20C, ocorreu uma provável mudança de cobertura e uso da terra causada por interferência humana, caracterizada pela transição de classes de vegetação para o uso agrícola. Em destaque, no ano-safra 16 (2019/20), o perfil foi classificado como PAST e a amostragem indica que ela pertence a classe AGR-1. Porém, durante a etapa de avaliação de qualidade das amostras, esse ponto foi considerado como um ruído no banco de dados e não participou da etapa de treino dos modelos com a AS. A PP dessa amostra foi de zero e no processo de análise, ela foi excluída por não apresentar um ciclo semelhante às demais amostras de AGR-1. No entanto, os perfis dos anos-safra 15 (2018/19) e 17 (2020/21) foram classificadas como PAST apesar de apresentarem características das classes AGR-2 e AGR-1, respectivamente. Essas confusões também foram encontradas nas camadas de probabilidade, e por isso as classificações foram consideradas como incoerentes, sendo referências para possíveis melhorias no modelo para estudos futuros.

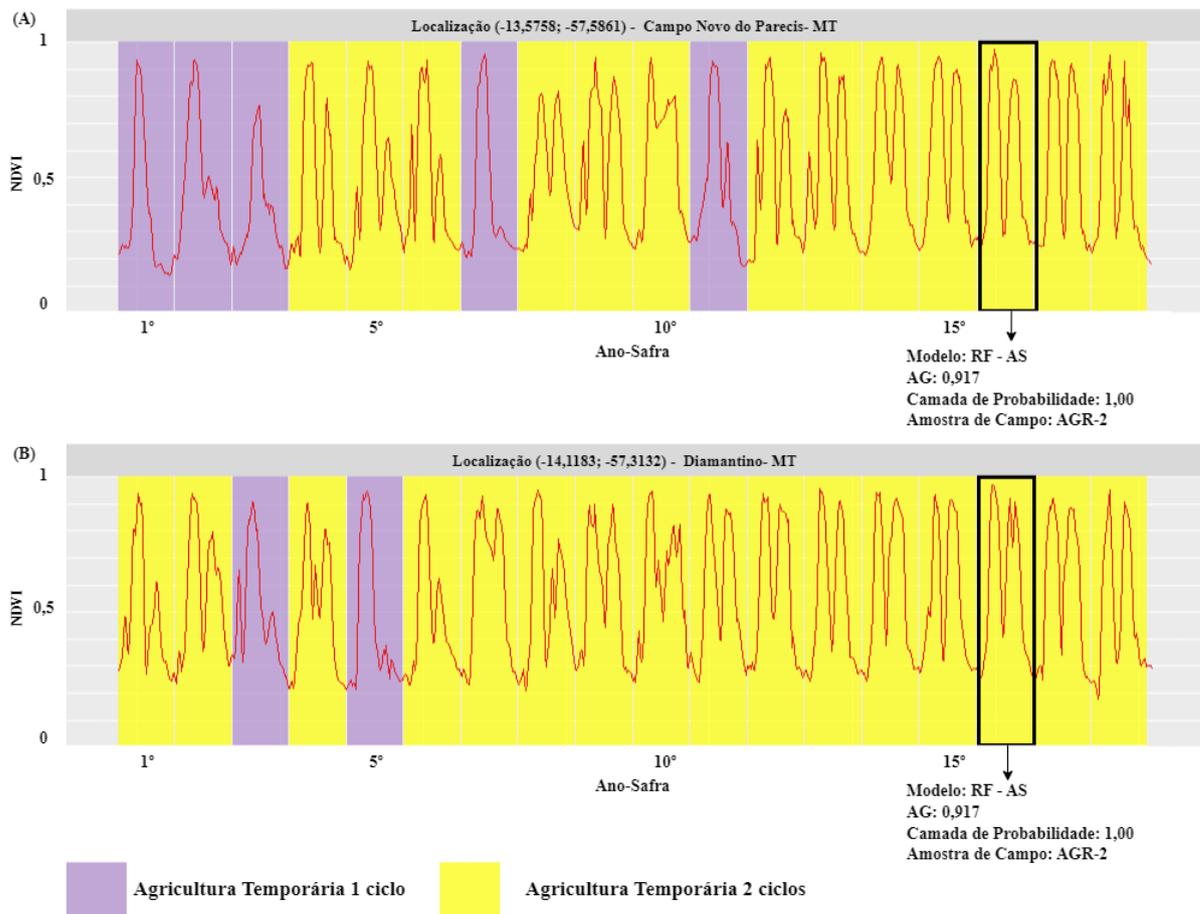


**Figura 21:** Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Campo Novo do Parecis-MT (A), Barra dos Bugres-MT (B) e Tangará da Serra-MT (C). Em destaque, o ano-safra correspondente a amostra de campo.

Os exemplos da Figura 21, são pontos localizados no Leste do Mato Grosso, região onde está concentrada a produção de cana-de-açúcar do estado. Na Figura 21A, também houve uma divergência entre a classe da amostra de campo e o resultado da classificação. No entanto, neste caso a divergência aconteceu por um erro de classificação do modelo. Esse ponto foi incluído do teste da Generalização com as amostras de campo e é um dos pontos em que o modelo classificou uma amostra de AGR-S como PAST. Na camada de probabilidade desse perfil, o valor para a PAST foi de 0,4, enquanto para AGR-S foi de 0,38. Embora o RF tenha apresentado melhores resultados que os demais algoritmos, esse resultado aponta a dificuldade do classificador em distinguir essas duas classes. Nos perfis do anos-safra 4 (2007/08) e 9 (2012/13) uma confusão semelhante aconteceu envolvendo essas duas classes. No perfil também há a presença de períodos de AGR-1, esses períodos foram identificados e classificados sem apresentar incoerências.

Na Figura 21B, o processo de identificação da classe AGR-S ocorreu sem aparentes erros de classificação. A mudança de PAST para a AGR-S foi identificada ao longo do perfil, e a presença de uma transição com a AGR-1 também foi interpretada pelo classificador. As camadas de probabilidade mostraram uma provável confusão no perfil de transição, localizado no ano-safra 7 (2010/11). Nos demais períodos, não foi identificado possíveis incoerências, principalmente para os perfis classificados como AGR-S. O perfil nesta figura é um exemplo em que o modelo foi capaz de classificar com sucesso a classe de AGR-S juntamente com perfis pertencentes a classe PAST, que foi a principal confusão registrada na etapa de teste.

A identificação de perfis da classe AGR-S da Figura 21C também ocorreu sem possíveis erros de classificação. Neste perfil, o modelo conseguiu identificar a transição envolvendo as classes AGR-S e AGR-2. Um possível erro de classificação aconteceu no ano-safra 3 (2006/07), onde, de acordo com a camada de probabilidade, aconteceu uma confusão envolvendo a VF, PAST e AGR-S. Pela análise do SOM, este perfil apresenta características espectro-temporais pouco representada na amostragem de treinamento. Por isso, essa confusão não foi interpretada como incoerente, e foi vista como referência para melhorias futuras do modelo.



**Figura 22:** Figuras referentes a classificação interanual, resultado da Generalização do modelo RF-AS nos pontos localizados nos municípios de Campo Novo do Parecis-MT (A) e Diamantino-MT (B). Em destaque, o ano-safra correspondente a amostra de campo.

Na Figura 22A, o perfil apresentado mostra uma região com dinâmica de transição entre as classes AGR-1 e AGR-2. Durante o ano-safra 10 (2013/14), provavelmente aconteceu um caso de interferência de nuvens e o perfil espectro-temporal foi corrigido pela banda CLOUD, conforme exemplo do neurônio 181, na Figura 12. Apesar dessa correção mudar as características do perfil, o modelo foi capaz de entender que essa correção foi causada pela presença de nuvens e classificar estes perfis como AGR-2. A Generalização também ocorreu sem a presença de incoerências no perfil da Figura 22B. Neste sentido, foi concluído que o modelo não apresentou dificuldades em classificar perfis de AGR-2.

A presença de erros de classificação era esperada devido aos resultados com amostras de campo e análises anteriores à classificação interanual, como o SOM. A maioria dos erros identificados nesta etapa ocorreram em pontos de transição entre duas ou mais classes temáticas (Figura 21A e Figura 20C,) envolvendo principalmente a classe PAST, que apresentou a maior tendência a erros de comissão devido ao seu resultado de 0,811 na métrica

*Precision.* Na distribuição de neurônios feita pelo SOM, também foi possível identificar essa tendência, uma vez que os neurônios de PAST apresentam proximidade com o agrupamento de todas as outras classes temáticas, e por isso foi considerada a classe que estava mais suscetível a erros de classificação.

Os autores MASSEY et al. (2017) e JOHNSON (2019) destacaram a dificuldade da tarefa em criar um modelo representativo para uma área extensa e que classifique diferentes anos. Apesar da presença de erros de classificação, a abordagem de Generalização obteve acurácia global de 0,917 no teste com amostras de campo. Na etapa de avaliação das séries temporais classificadas, os resultados não apresentaram incoerências que pudessem atrapalhar a interpretação das séries temporais. A maioria dos perfis analisados apresentou resultados assertivos, principalmente na identificação de classes como a AGR-2 e AGR-1 (Figura 22). Neste sentido, dada a dificuldade da tarefa de Generalização, foi considerado que os resultados obtidos foram promissores e demonstraram o potencial dessa abordagem para estudos envolvendo dados de sensoriamento remoto.

Os autores LUCIANO; CAMPAGNUCI; LE MAIRE, (2022) também utilizaram o RF como algoritmo base para a Generalização. Com base nos resultados obtidos, a Generalização é uma abordagem que requer que o classificador compreenda as variabilidades intraclasse e se adapte a amostras diferentes das utilizadas no treinamento (KHANNA et al., 2022; ZHOU et al., 2021). Os resultados deste trabalho apontaram que as principais fontes de erros de classificação ocorreram em perfis que representavam uma variação que era minoria dentro de uma classe temática. Neste sentido, a abordagem utilizada pelo RF é mais indicada para a Generalização. O algoritmo utiliza um comitê de diferentes árvores de decisão, o que permite melhor adaptação a classes minoritárias (JAMES et al., 2013; SANTOS et al., 2019). A alternativa de utilizar um algoritmo baseado em redes neurais não é descartada, mas um melhor entendimento sobre as variações dentro da área de estudo e um maior volume de amostras são necessários para que essas redes possam aprender os diferentes padrões intraclasse (KHANNA et al., 2022; WEN et al., 2022).

## 6 Considerações Finais e Conclusão

O presente trabalho apresentou as etapas da criação de um modelo de classificação utilizado numa tarefa de Generalização para a classificação interanual de séries temporais do sensor MODIS. Durante a etapa de pré-processamento foi utilizado o algoritmo SOM, que se mostrou eficaz na redução de ruídos e na identificação de padrões dentro das classes temáticas, auxiliando na caracterização das amostras de campo, no desempenho dos algoritmos (RF) e TempCNN e interpretação dos resultados finais da classificação.

O modelo criado para o ano-safra base 2019/20 apresentou bons desempenhos e atingiu os melhores resultados com o algoritmo RF, obtendo uma acurácia global de 0,994. Nesta etapa, não foram encontradas diferenças significativas entre os algoritmos utilizados para o treinamento (RF- TempCNN- L-TAE), ao contrário de outros trabalhos estudados durante a revisão bibliográfica. Como todos os algoritmos alcançaram bons resultados, foi considerado que a criação do modelo base para o ano-safra 2019/20 atingiu as expectativas e poderia ser utilizado na tarefa de Generalização.

Na Generalização, o modelo treinado com o algoritmo RF foi o cenário que apresentou os melhores resultados em termos de acurácia global (0,917) e métricas como o *F1-Score*, tanto antes quanto após o tratamento de amostras com o SOM. Sua capacidade de lidar com classes desbalanceadas e de adaptar-se a diferentes padrões intraclasse permitiram resultados mais conscientes em comparação a TempCNN e L-TAE. Neste trabalho, a aplicação de redes neurais para a tarefa de Generalização teve como principais obstáculos o aprendizado em pontos com perfis de transição entre as classes temáticas e a presença de variações intraclasse, que eram uma porção minoritária dentro do conjunto de amostras. Com base nas métricas *Precision* e *Recall*, essa dificuldade aconteceu principalmente entre as classes Agricultura Semiperene e Pastagem.

A capacidade de Generalização do modelo também foi testada por meio da classificação interanual de 18 períodos de ano-safra da série temporal do MODIS. Pontos das amostras de campo foram selecionados para verificar a adaptação do modelo RF-AS em perfis espectro-temporais característicos do estado de Mato Grosso. Embora possíveis erros tenham sido apontados, a análise destes pontos mostrou que o modelo foi capaz de realizar a Generalização e classificar as cinco classes temáticas ao longo das séries temporais. Neste sentido, foi concluído que o modelo tem potencial para ser utilizado em conjunto em plataformas que disponibilizam séries temporais de sensores orbitais, como SATVeg, e auxiliar no processo de interpretação dos perfis espectro-temporais.

A eficácia da abordagem de Generalização permite que estudos com imagens de sensoriamento remoto não fiquem limitados a regiões com amostras de campo, permitindo maior abrangência para estudos futuros. Além disso, para melhorar a precisão da classificação e lidar com variações intraclasse, estudos futuros podem se concentrar em aumentar o entendimento das possíveis variações intraclasse presentes no estado de Mato Grosso. Neste sentido, para a abordagem utilizando algoritmos baseados em redes neurais, um maior número de amostras referentes aos padrões minoritários é essencial para auxiliar esses algoritmos. Além disso, a adição de amostras referentes a outros anos também é uma estratégia que pode ser utilizada para melhoria da classificação interanual.

A partir da interpretação dos resultados obtidos as conclusões deste trabalho foram:

- O tratamento de qualidade das amostras feito pelo SOM auxiliou os algoritmos RF e TempCNN na classificação de uso e cobertura da terra dos perfis espectrais do MODIS.
- O melhor modelo foi obtido com o algoritmo RF após o tratamento de amostras feito pelo SOM, em que a acurácia global foi de 0,917.
- A abordagem de Generalização de um modelo com baseado no ano-safra 2019/20 permitiu a classificação de diferentes períodos de ano-safra (2016/17, 2020/21 e 2021/22).
- Na aplicação para a classificação interanual de 18 períodos de ano-safra, o modelo RF-AS apresentou possíveis erros de classificação, porém os resultados mostraram que o modelo foi capaz de classificar as cinco classes temáticas ao longo das séries temporais do MODIS.

## 7 Referências Bibliográficas

- ABIDI, A. et al. Combining 2D encoding and convolutional neural network to enhance land cover mapping from Satellite Image Time Series. **Engineering Applications of Artificial Intelligence**, v. 122, 1 jun. 2023.
- ARVOR, D. et al. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. **Applied Geography**, v. 32, n. 2, p. 702–713, 2012.
- BATTA, M. Machine Learning Algorithms - A Review. **International Journal of Science and Research (IJSR)**, v. 18, n. 8, p. 381–386, 2018.
- BELGIU, M.; DRĂGU, 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.
- BENDINI, H. et al. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 82, n. December 2018, p. 101872, 2019.
- BLANCHARD, G.; LEE, G.; SCOTT, C. **Generalizing from Several Related Classification Tasks to a New Unlabeled Sample**. 2011.
- BREIMAN, Leo. Random forests. *Machine learning*, v. 45, p. 5-32, 2001.
- BROWN, J. C. et al. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. **Remote Sensing of Environment**, v. 130, p. 39–50, 2013.
- BUENO, I. T. et al. Mapping integrated crop-livestock systems in Brazil with planetscope time series and deep learning. **Remote Sensing of Environment**, v. 299, 15 dez. 2023.
- CHAVES, M. E. D. et al. Time-weighted dynamic time warping analysis for mapping interannual cropping practices changes in large-scale agro-industrial farms in Brazilian Cerrado. **Science of Remote Sensing**, v. 3, 1 jun. 2021.
- CHAWLA, N. V et al. **SMOTE: Synthetic Minority Over-sampling Technique** *Journal of Artificial Intelligence Research*. 2002.
- CHEN, Y. et al. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. **International Journal of Applied Earth Observation and Geoinformation**, v. 69, p. 133–147, 1 jul. 2018.
- CHEN, Y. et al. A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 180, n. August, p. 174–190, 2021.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35–46, 1991.
- CUI, B.; CHEN, X.; LU, Y. Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network with Dense Connection. **IEEE Access**, v. 8, p. 116744–116755, 2020.
- DA SILVA, J. L. B. et al. Spatiotemporal dynamics of agricultural areas with central pivot using orbital remote sensing in the Brazilian semiarid. **Smart Agricultural Technology**, v. 2, n. January, p. 100052, 2022.

- DANTAS CHAVES, M. E.; DE CARVALHO ALVES, M. Recent applications of the MODIS sensor for soybean crop monitoring and deforestation detection in Mato Grosso, Brazil. **CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources**, v. 14, n. February, 2019.
- DEBELLA-GILO, M.; GJERTSEN, A. K. Mapping seasonal agricultural land use types using deep learning on sentinel-2 image time series. **Remote Sensing**, v. 13, n. 2, p. 1–17, 2 jan. 2021.
- ESQUERDO, J. C. D. M. et al. SATVeg: A web-based tool for visualization of MODIS vegetation indices in South America. **Computers and Electronics in Agriculture**, v. 175, n. June 2019, p. 105516, 2020.
- FERCHICHI, A. et al. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. **Ecological Informatics**, v. 68, n. January, p. 101552, 2022.
- FERREIRA, K. R. et al. Earth observation data cubes for Brazil: Requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, p. 1–19, 2020.
- FIRAT, H. et al. Spatial-spectral classification of hyperspectral remote sensing images using 3D CNN based LeNet-5 architecture. **Infrared Physics and Technology**, v. 127, n. October, 2022.
- FONSECA, L. M. G. et al. Pattern Recognition and Remote Sensing techniques applied to Land Use and Land Cover mapping in the Brazilian Savannah. **Pattern Recognition Letters**, v. 148, p. 54–60, 2021.
- GARNOT, V. S. F.; LANDRIEU, L. Lightweight temporal self-attention for classifying satellite images time series. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 12588 LNAI, p. 171–181, 2020.
- GE, S. et al. Transferable deep learning model based on the phenological matching principle for mapping crop extent. **International Journal of Applied Earth Observation and Geoinformation**, v. 102, 1 out. 2021.
- GORELICK, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017.
- HAO, P. et al. Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. **Remote Sensing**, v. 7, n. 5, p. 5347–5369, 2015.
- HAO, P. et al. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. **Science of the Total Environment**, v. 733, p. 138869, 2020.
- HU, Y. et al. An Interannual Transfer Learning Approach for Crop Classification in the Hetao Irrigation District, China. **Remote Sensing**, v. 14, n. 5, 1 mar. 2022.
- HUANG, H. et al. The migration of training samples towards dynamic global land cover mapping. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 161, n. January, p. 27–36, 2020.
- HUANG, S. et al. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. **Journal of Forestry Research**, v. 32, n. 1, p. 1–6, 2021.
- HUETE, A.R.; HuiQing Liu; Van Leeuwen, W.J.D. **The use of vegetation indices in forested regions: issues of linearity and saturation**. IGARSS'97. 1997 IEEE Int. Geosci. Remote Sens. Symp. Proceedings. Remote Sens. A Sci. Vis. Sustain. Dev. 1997, 4, 1966–1968; doi:10.1109/IGARSS.1997.609169

- ISMAIL FAWAZ, H. et al. Deep learning for time series classification: a review. **Data Mining and Knowledge Discovery**, v. 33, n. 4, p. 917–963, 2019.
- JAKIMOW, B. et al. Mapping pasture management in the Brazilian Amazon from dense Landsat time series. **Remote Sensing of Environment**, v. 205, p. 453–468, 1 fev. 2018.
- JAMES, G. et al. **Introduction to Statistical Analysis**. 6th. ed. [s.l.: s.n.].
- JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine Learning and Deep Learning. **Ingeniare**, v. 29, n. 2, p. 182–183, 2021.
- JENSEN, J. R. Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres. [S.l.]: Parêntese Editora São José dos Campos, 2011
- JOHNSON, D. M. Using the Landsat archive to map crop cover history across the United States. **Remote Sensing of Environment**, v. 232, n. July, p. 111286, 2019.
- JUSTICE, C. O. et al. **An overview of MODIS Land data processing and product status**. , 2002.
- KATTENBORN, T. et al. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 173, n. December 2020, p. 24–49, 2021.
- KHAN, A. et al. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. **Ecological Informatics**, v. 69, n. February, p. 101678, 2022.
- KHANNA, S. et al. Activation Regression for Continuous Domain Generalization with Applications to Crop Classification. 14 abr. 2022.
- KOHONEN, T. The Self-Organizing Map. Proc. IEEE 1990, 78, 1464–1480;.
- LIAO, C. et al. Synergistic use of multi-temporal RADARSAT-2 and VENUS data for crop classification based on 1D convolutional neural network. **Remote Sensing**, v. 12, n. 5, p. 1–17, 2020.
- LLANO, X. (2022), SMBByC-IDEAM. AcATaMa - QGIS plugin for Accuracy Assessment of Thematic Maps, version 23.4. Available in <https://github.com/SMBByC/AcATaMa>
- LUCIANO, A. C. DOS S. et al. Generalized space-time classifiers for monitoring sugarcane areas in Brazil. **Remote Sensing of Environment**, v. 215, n. October 2017, p. 438–451, 2018.
- LUCIANO, A. C. DOS S. et al. A generalized space-time OBIA classification scheme to map sugarcane areas at regional scale, using Landsat images time-series and the random forest algorithm. **International Journal of Applied Earth Observation and Geoinformation**, v. 80, n. December 2018, p. 127–136, 2019.
- LUCIANO, A. C. DOS S.; CAMPAGNUCI, B. C. G.; LE MAIRE, G. Mapping 33 years of sugarcane evolution in São Paulo state, Brazil, using landsat imagery and generalized space-time classifiers. **Remote Sensing Applications: Society and Environment**, v. 26, 1 abr. 2022.
- MA, L. et al. Deep learning in remote sensing applications: A meta-analysis and review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 152, n. March, p. 166–177, 2019.
- MA, Y. et al. **Transfer learning in environmental remote sensing**. **Remote Sensing of Environment** Elsevier Inc., , 1 fev. 2024.
- MAROTO-MOLINA, F. et al. Data pre-processing to improve the mining of large feed databases. **Animal**, v. 7, n. 7, p. 1128–1136, 2013.

- MASSEY, R. et al. MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types. **Remote Sensing of Environment**, v. 198, p. 490–503, 1 set. 2017.
- MILLARD, K.; RICHARDSON, M. On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. **Remote Sensing**, v. 7, n. 7, p. 8489–8515, 2015.
- MIRALLES, P. et al. ScienceDirect A critical review on the state-of-the-art and future prospects of machine learning for Earth observation operations. n. xxxx, 2023.
- MÜLLER, H. et al. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, p. 490–499, 1 jan. 2015.
- NASIRI, V. et al. Comparing Object-Based and Pixel-Based Machine Learning Models for Tree-Cutting Detection with PlanetScope Satellite Images: Exploring Model Generalization. **International Journal of Applied Earth Observation and Geoinformation**, v. 125, 1 dez. 2023.
- OLOFSSON, P. et al. **Good practices for estimating area and assessing accuracy of land change.** **Remote Sensing of Environment** Elsevier Inc., , 25 maio 2014.
- PAN, S. J.; YANG, Q. **A survey on transfer learning.** **IEEE Transactions on Knowledge and Data Engineering**, 2010.
- PELLETIER, C. et al. 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.
- 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. 1–25, 2019.
- PEREIRA, O. J. R. et al. Assessing pasture degradation in the Brazilian Cerrado based on the analysis of MODIS NDVI time-series. **Remote Sensing**, v. 10, n. 11, 2018.
- 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, n. August, p. 328–339, 2018.
- PIGNÈDE, E. et al. Sugarcane yield forecast in ivory coast (West africa) based on weather and vegetation index data. **Atmosphere**, v. 12, n. 11, p. 1–22, 2021.
- ROUSE, J. W. et al. **Monitoring vegetation systems in the great plains with ERTS.** In: EARTH RESOURCES TECHNOLOGY SATELLITE-1 SYMPOSIUM, 3., 1973, Washington. Proceedings... Washington: NASA, v.1, p.309-317, 1973
- RUSSWURM, M. et al. BREIZHCROPS: A TIME SERIES DATASET FOR CROP TYPE MAPPING. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XLIII-B2-2, n. B2, p. 1545–1551, 14 ago. 2020.
- SAINTE FARE GARNOT, V. et al. Satellite image time series classification with pixel-set encoders and temporal self-attention. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, n. June, p. 12322–12331, 2020.
- SANTOS, C. L. M. DE O. et al. Classification of crops, pastures, and tree plantations along the season with multi-sensor image time series in a subtropical agricultural region. **Remote Sensing**, v. 11, n. 3, p. 1–27, 2019.

- SANTOS, L. A. et al. Quality control and class noise reduction of satellite image time series. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 177, n. July 2020, p. 75–88, 2021a.
- SANTOS, L. A. et al. Identifying spatiotemporal patterns in land use and cover samples from satellite image time series. **Remote Sensing**, v. 13, n. 5, p. 1–21, 2021b.
- SANTOSH, K.; DAS, N.; GHOSH, S. **Deep learning: a review**. [s.l: s.n.].
- SAVISTKY, A.; GOLEY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1639–1643, 1964.
- SILVA JUNIOR, C. A. DA et al. Mapping soybean planting area in midwest Brazil with remotely sensed images and phenology-based algorithm using the Google Earth Engine platform. **Computers and Electronics in Agriculture**, v. 169, n. January, p. 105194, 2020.
- SIMOES, R. et al. Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. **Scientific Data**, v. 7, n. 1, p. 1–10, 2020.
- SIMOES, R. et al. Satellite Image Time Series Analysis for Big Earth Observation Data. p. 1–20, 2021.
- SOILLE, P. et al. A versatile data-intensive computing platform for information retrieval from big geospatial data. **Future Generation Computer Systems**, v. 81, p. 30–40, 1 abr. 2018.
- SONY, S. et al. A systematic review of convolutional neural network-based structural condition assessment techniques. **Engineering Structures**, v. 226, n. January 2020, p. 111347, 2021.
- SOUZA, C. M. et al. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, v. 12, n. 17, 1 set. 2020.
- TORO, A. P. S. G. D. et al. **EVALUATION OF EARLY SEASON MAPPING OF INTEGRATED CROP LIVESTOCK SYSTEMS USING SENTINEL-2 DATA**. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. **Anais...International Society for Photogrammetry and Remote Sensing**, 30 maio 2022.
- VELTMAN, A.; PULLE, D. W. J.; DE DONCKER, R. W. Attention is All You Need. **Power Systems**, n. Nips, p. 47–82, 2017.
- VIEIRA, D. C. et al. Cropland expansion, intensification, and reduction in Mato Grosso state, Brazil, between the crop years 2000/01 to 2017/18. **Remote Sensing Applications: Society and Environment**, v. 28, n. August, p. 100841, 2022.
- VUOLO, F.; NG, W. T.; ATZBERGER, C. Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data. **International Journal of Applied Earth Observation and Geoinformation**, v. 57, p. 202–213, 1 maio 2017.
- WANG, J. et al. Generalizing to Unseen Domains: A Survey on Domain Generalization. 2 mar. 2021.
- WANG, Q. et al. Remote sensing image gap filling based on spatial-spectral random forests. **Science of Remote Sensing**, v. 5, n. March, p. 100048, 2022.
- WANG, S.; AZZARI, G.; LOBELL, D. B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. **Remote Sensing of Environment**, v. 222, n. December 2018, p. 303–317, 2019.
- WANG, Y. et al. **A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing**. **ISPRS Journal of Photogrammetry and Remote Sensing** Elsevier B.V., , 1 dez. 2023.

- WEN, Q. et al. Transformers in Time Series: A Survey. 2022.
- XI, Y. et al. Exploitation of Time Series Sentinel-2 Data and Different Machine Learning Algorithms for Detailed Tree Species Classification. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 7589–7603, 2021.
- XIONG, J. et al. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 126, p. 225–244, 1 abr. 2017.
- XU, J. et al. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. **Remote Sensing of Environment**, v. 247, n. March, p. 111946, 2020.
- XUE, J.; SU, B. Significant remote sensing vegetation indices: A review of developments and applications. **Journal of Sensors**, v. 2017, 2017.
- YOU, N.; DONG, J. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 161, n. January, p. 109–123, 2020.
- ZANG, Y. et al. Remote sensing index for mapping canola flowers using modis data. **Remote Sensing**, v. 12, n. 23, p. 1–19, 2020.
- ZHANG, M. et al. A MODIS-based scalable remote sensing method to estimate sowing and harvest dates of soybean crops in Mato Grosso, Brazil. **Heliyon**, v. 7, n. 7, p. e07436, 2021.
- ZHAO, H. et al. Evaluation of five deep learning models for crop type mapping using sentinel-2 time series images with missing information. **Remote Sensing**, v. 13, n. 14, 2 jul. 2021.
- ZHONG, L. et al. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 151–164, 1 set. 2016.
- ZHONG, L.; GONG, P.; BIGING, G. S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. **Remote Sensing of Environment**, v. 140, p. 1–13, jan. 2014.
- ZHONG, L.; HU, L.; ZHOU, H. Deep learning based multi-temporal crop classification. **Remote Sensing of Environment**, v. 221, n. November 2018, p. 430–443, 2019.
- ZHOU, K. et al. Domain Generalization: A Survey. 3 mar. 2021.

## Apêndice A – Matriz de Confusão (Amostras Teste)

Random Forest (AO) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Acc Usuário
Agricultura Temporaria 2 ciclos	1571	70	4	7	0	0,951
Agricultura Temporaria 1 ciclo	36	902	2	4	0	0,956
Agricultura Semiperene	2	25	759	23	0	0,938
Pastagem	19	79	138	1064	77	0,773
Vegetacao Florestal	0	0	2	67	668	0,906
Acc Produtor	0,965	0,838	0,839	0,913	0,897	0,899

TempCNN (AO) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1571	82	4	9	9	0,938
Agricultura Temporaria 1 ciclo	41	901	2	8	0	0,946
Agricultura Semiperene	4	11	588	71	3	0,869
Pastagem	12	79	309	1033	88	0,679
Vegetacao Florestal	0	3	2	44	645	0,929
Recall	0,965	0,837	0,650	0,887	0,866	0,858

L-TAE (AO) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1549	43	4	16	2	0,960
Agricultura Temporaria 1 ciclo	56	947	11	13	0	0,922
Agricultura Semiperene	10	20	631	133	4	0,791
Pastagem	12	65	258	947	77	0,697
Vegetacao Florestal	1	1	1	56	662	0,918
Recall	0,951	0,880	0,697	0,813	0,889	0,858

Random Forest (AS) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1591	42	1	9	1	0,968
Agricultura Temporaria 1 ciclo	5	917	0	0	0	0,995
Agricultura Semiperene	11	24	809	47	2	0,906
Pastagem	22	90	91	1040	40	0,811
Vegetacao Florestal	0	2	4	69	702	0,903
Recall	0,977	0,853	0,894	0,893	0,942	0,917

TempCNN (AS) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1553	40	2	9	3	0,966
Agricultura Temporaria 1 ciclo	28	944	4	12	1	0,954
Agricultura Semiperene	26	10	709	64	2	0,874
Pastagem	17	78	189	1010	63	0,744
Vegetacao Florestal	4	4	1	70	676	0,895
Recall	0,954	0,877	0,783	0,867	0,907	0,886

L-TAE (AS) - Generalização						
	Agricultura Temporaria 2 ciclos	Agricultura Temporaria 1 ciclo	Agricultura Semiperene	Pastagem	Vegetacao Florestal	Precision
Agricultura Temporaria 2 ciclos	1507	41	5	7	4	0,964
Agricultura Temporaria 1 ciclo	32	917	4	3	0	0,959
Agricultura Semiperene	25	10	593	39	4	0,884
Pastagem	61	107	290	1043	57	0,669
Vegetacao Florestal	3	1	13	73	680	0,883
Recall	0,926	0,852	0,655	0,895	0,913	0,859

## Apêndice B – Camadas de Probabilidade RF x TempCNN

Camada de Probabilidades - Sorriso-MT - Random Forest - Fig 18A									
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal	
1°	05/09/2004	29/08/2005	Agricultura Temp 2 ciclos	0,032	0,358	0,542	0,068	0	
2°	06/09/2005	29/08/2006	Pastagem	0,102	0,108	0,272	0,518	0	
3°	06/09/2006	29/08/2007	Pastagem	0,034	0,002	0,22	0,742	0,002	
4°	06/09/2007	28/08/2008	Pastagem	0,186	0,016	0,14	0,624	0,034	
5°	05/09/2008	29/08/2009	Agricultura Temp 2 ciclos	0,054	0,014	0,88	0,036	0,016	
6°	06/09/2009	29/08/2010	Agricultura Temp 1 ciclo	0,024	0,852	0,08	0,044	0	
7°	06/09/2010	29/08/2011	Pastagem	0,062	0	0,338	0,6	0	
8°	06/09/2011	28/08/2012	Agricultura Semiperene	0,404	0,006	0,208	0,382	0	
9°	05/09/2012	29/08/2013	Agricultura Semiperene	0,674	0,004	0,028	0,282	0,012	
10°	06/09/2013	29/08/2014	Agricultura Semiperene	0,512	0,072	0,104	0,28	0,032	
11°	06/09/2014	29/08/2015	Agricultura Semiperene	0,81	0,008	0,05	0,08	0,052	
12°	06/09/2015	28/08/2016	Agricultura Temp 1 ciclo	0,24	0,442	0,06	0,254	0,004	
13°	05/09/2016	29/08/2017	Agricultura Semiperene	0,628	0,024	0,104	0,244	0	
14°	06/09/2017	29/08/2018	Agricultura Semiperene	0,84	0,032	0,032	0,094	0,002	
15°	06/09/2018	29/08/2019	Agricultura Semiperene	0,778	0,01	0,03	0,176	0,006	
16°	06/09/2019	28/08/2020	Agricultura Semiperene	0,984	0,002	0	0,012	0,002	
17°	05/09/2020	29/08/2021	Agricultura Semiperene	0,846	0	0,01	0,13	0,014	
18°	06/09/2021	29/08/2022	Agricultura Temp 2 ciclos	0,222	0,01	0,682	0,086	0	

Camada de Probabilidades - Sorriso-MT - TempCNN - Fig 18B									
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal	
1°	05/09/2004	29/08/2005	Pastagem	0	0	0	1	0	
2°	06/09/2005	29/08/2006	Agricultura Temp 1 ciclo	0	1	0	0	0	
3°	06/09/2006	29/08/2007	Pastagem	0	0	0	1	0	
4°	06/09/2007	28/08/2008	Pastagem	0	0	0	1	0	
5°	05/09/2008	29/08/2009	Agricultura Temp 2 ciclos	0	0	1	0	0	
6°	06/09/2009	29/08/2010	Agricultura Temp 1 ciclo	0	1	0	0	0	
7°	06/09/2010	29/08/2011	Pastagem	0	0	0	1	0	
8°	06/09/2011	28/08/2012	Pastagem	0	0	0,281	0,719	0	
9°	05/09/2012	29/08/2013	Agricultura Semiperene	1	0	0	0	0	
10°	06/09/2013	29/08/2014	Agricultura Semiperene	1	0	0	0	0	
11°	06/09/2014	29/08/2015	Pastagem	0	0	0	1	0	
12°	06/09/2015	28/08/2016	Agricultura Temp 1 ciclo	0	0,997	0	0,003	0	
13°	05/09/2016	29/08/2017	Agricultura Semiperene	0,999	0	0	0,001	0	
14°	06/09/2017	29/08/2018	Agricultura Semiperene	1	0	0	0	0	
15°	06/09/2018	29/08/2019	Agricultura Semiperene	0,974	0	0	0,026	0	
16°	06/09/2019	28/08/2020	Agricultura Semiperene	1	0	0	0	0	
17°	05/09/2020	29/08/2021	Pastagem	0,418	0	0	0,582	0	
18°	06/09/2021	29/08/2022	Agricultura Temp 2 ciclos	0	0	1	0	0	

Camada de Probabilidades - Paranatinga-MT - Random Forest - Fig 19A								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Pastagem	0,134	0	0,004	0,828	0,034
2°	06/09/2005	29/08/2006	Pastagem	0,224	0,004	0,07	0,688	0,014
3°	06/09/2006	29/08/2007	Pastagem	0,006	0,002	0	0,984	0,008
4°	06/09/2007	28/08/2008	Pastagem	0,008	0,002	0,002	0,988	0
5°	05/09/2008	29/08/2009	Pastagem	0,032	0	0	0,96	0,008
6°	06/09/2009	29/08/2010	Pastagem	0,004	0	0	0,996	0
7°	06/09/2010	29/08/2011	Pastagem	0	0	0	1	0
8°	06/09/2011	28/08/2012	Pastagem	0,026	0,014	0,048	0,912	0
9°	05/09/2012	29/08/2013	Pastagem	0,004	0,004	0,006	0,986	0
10°	06/09/2013	29/08/2014	Pastagem	0,01	0	0,002	0,988	0
11°	06/09/2014	29/08/2015	Pastagem	0,014	0,008	0,006	0,972	0
12°	06/09/2015	28/08/2016	Pastagem	0,024	0,006	0,004	0,966	0
13°	05/09/2016	29/08/2017	Pastagem	0,08	0,006	0,076	0,838	0
14°	06/09/2017	29/08/2018	Pastagem	0	0	0	1	0
15°	06/09/2018	29/08/2019	Pastagem	0,01	0	0	0,99	0
16°	06/09/2019	28/08/2020	Pastagem	0,012	0,004	0,03	0,954	0
17°	05/09/2020	29/08/2021	Pastagem	0,028	0,016	0,022	0,934	0
18°	06/09/2021	29/08/2022	Pastagem	0,144	0,012	0,05	0,794	0

Camada de Probabilidades - Paranatinga-MT - TempCNN - Fig 19B								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Pastagem	0	0	0	1	0
2°	06/09/2005	29/08/2006	Pastagem	0,01	0	0	0,99	0
3°	06/09/2006	29/08/2007	Pastagem	0	0	0	1	0
4°	06/09/2007	28/08/2008	Pastagem	0	0	0	1	0
5°	05/09/2008	29/08/2009	Agricultura Semiperene	0,54	0	0	0,46	0
6°	06/09/2009	29/08/2010	Pastagem	0	0	0	1	0
7°	06/09/2010	29/08/2011	Pastagem	0	0	0	1	0
8°	06/09/2011	28/08/2012	Pastagem	0	0	0	1	0
9°	05/09/2012	29/08/2013	Pastagem	0,01	0	0	0,99	0
10°	06/09/2013	29/08/2014	Pastagem	0	0	0	1	0
11°	06/09/2014	29/08/2015	Pastagem	0	0	0	1	0
12°	06/09/2015	28/08/2016	Pastagem	0	0	0	1	0
13°	05/09/2016	29/08/2017	Agricultura Semiperene	0,86	0	0	0,14	0
14°	06/09/2017	29/08/2018	Pastagem	0	0	0	1	0
15°	06/09/2018	29/08/2019	Pastagem	0	0	0	1	0
16°	06/09/2019	28/08/2020	Pastagem	0	0	0	1	0
17°	05/09/2020	29/08/2021	Pastagem	0	0	0	1	0
18°	06/09/2021	29/08/2022	Pastagem	0	0	0	1	0

## Apêndice C – Camadas de Probabilidade RF-AS

Camada de Probabilidades - Rondolândia - Fig 20A								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Vegetacao Florestal	0	0	0	0	1
2°	06/09/2005	29/08/2006	Vegetacao Florestal	0	0	0	0	1
3°	06/09/2006	29/08/2007	Vegetacao Florestal	0	0	0	0	1
4°	06/09/2007	28/08/2008	Vegetacao Florestal	0	0	0	0	1
5°	05/09/2008	29/08/2009	Vegetacao Florestal	0	0	0	0	1
6°	06/09/2009	29/08/2010	Vegetacao Florestal	0	0	0	0	1
7°	06/09/2010	29/08/2011	Vegetacao Florestal	0	0	0	0	1
8°	06/09/2011	28/08/2012	Vegetacao Florestal	0	0	0	0	1
9°	05/09/2012	29/08/2013	Vegetacao Florestal	0	0	0	0	1
10°	06/09/2013	29/08/2014	Vegetacao Florestal	0	0	0	0	1
11°	06/09/2014	29/08/2015	Vegetacao Florestal	0	0	0	0	1
12°	06/09/2015	28/08/2016	Vegetacao Florestal	0	0	0	0	1
13°	05/09/2016	29/08/2017	Vegetacao Florestal	0	0	0	0	1
14°	06/09/2017	29/08/2018	Vegetacao Florestal	0,002	0	0	0	0,998
15°	06/09/2018	29/08/2019	Vegetacao Florestal	0	0	0	0	1
16°	06/09/2019	28/08/2020	Vegetacao Florestal	0	0	0	0	1
17°	05/09/2020	29/08/2021	Vegetacao Florestal	0	0	0	0	1
18°	06/09/2021	29/08/2022	Vegetacao Florestal	0	0	0	0	1

Camada de Probabilidades - Comodoro-MT - Fig 20B								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Vegetacao Florestal	0,01	0	0,002	0,074	0,914
2°	06/09/2005	29/08/2006	Vegetacao Florestal	0,108	0	0,002	0,246	0,644
3°	06/09/2006	29/08/2007	Vegetacao Florestal	0,046	0	0,004	0,166	0,784
4°	06/09/2007	28/08/2008	Vegetacao Florestal	0,012	0,002	0,004	0,186	0,796
5°	05/09/2008	29/08/2009	Vegetacao Florestal	0,022	0	0,004	0,092	0,882
6°	06/09/2009	29/08/2010	Vegetacao Florestal	0,022	0	0	0,22	0,758
7°	06/09/2010	29/08/2011	Vegetacao Florestal	0,034	0	0,004	0,13	0,832
8°	06/09/2011	28/08/2012	Vegetacao Florestal	0,014	0	0	0,048	0,938
9°	05/09/2012	29/08/2013	Vegetacao Florestal	0	0	0	0	1
10°	06/09/2013	29/08/2014	Vegetacao Florestal	0,002	0	0,002	0,002	0,994
11°	06/09/2014	29/08/2015	Vegetacao Florestal	0	0	0	0	1
12°	06/09/2015	28/08/2016	Pastagem	0,246	0	0,006	0,556	0,192
13°	05/09/2016	29/08/2017	Vegetacao Florestal	0,034	0	0,002	0,208	0,756
14°	06/09/2017	29/08/2018	Pastagem	0,058	0	0,014	0,772	0,156
15°	06/09/2018	29/08/2019	Pastagem	0,156	0	0,012	0,616	0,216
16°	06/09/2019	28/08/2020	Pastagem	0,016	0,058	0,018	0,902	0,006
17°	05/09/2020	29/08/2021	Pastagem	0,036	0,052	0,036	0,876	0
18°	06/09/2021	29/08/2022	Agricultura Temp 1 ciclo	0,044	0,924	0,008	0,024	0

Camada de Probabilidades - Brasnorte-MT - Fig 20C								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Vegetacao Florestal	0,214	0	0,006	0,246	0,534
2°	06/09/2005	29/08/2006	Vegetacao Florestal	0	0	0	0	1
3°	06/09/2006	29/08/2007	Vegetacao Florestal	0,012	0	0,002	0,01	0,976
4°	06/09/2007	28/08/2008	Vegetacao Florestal	0,002	0,004	0	0	0,994
5°	05/09/2008	29/08/2009	Vegetacao Florestal	0,008	0	0	0,01	0,982
6°	06/09/2009	29/08/2010	Pastagem	0,148	0	0,01	0,468	0,374
7°	06/09/2010	29/08/2011	Pastagem	0,108	0,008	0,004	0,744	0,136
8°	06/09/2011	28/08/2012	Pastagem	0,19	0,004	0,032	0,494	0,28
9°	05/09/2012	29/08/2013	Pastagem	0,036	0,066	0,044	0,844	0,01
10°	06/09/2013	29/08/2014	Agricultura Temp 1 ciclo	0,052	0,55	0,038	0,36	0
11°	06/09/2014	29/08/2015	Pastagem	0,204	0,254	0,184	0,358	0
12°	06/09/2015	28/08/2016	Agricultura Temp 1 ciclo	0,036	0,926	0,002	0,036	0
13°	05/09/2016	29/08/2017	Pastagem	0,03	0	0,006	0,868	0,096
14°	06/09/2017	29/08/2018	Agricultura Temp 2 ciclos	0,034	0,136	0,74	0,086	0,004
15°	06/09/2018	29/08/2019	Pastagem	0,086	0,106	0,316	0,488	0,004
16°	06/09/2019	28/08/2020	Pastagem	0,004	0	0,002	0,994	0
17°	05/09/2020	29/08/2021	Pastagem	0,108	0,144	0,34	0,406	0,002
18°	06/09/2021	29/08/2022	Agricultura Temp 1 ciclo	0,042	0,878	0,01	0,07	0

Camada de Probabilidades - Campo Novo do Parecis-MT - Fig 21A								
Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Agricultura Semiperene	0,534	0,01	0,064	0,348	0,044
2°	06/09/2005	29/08/2006	Agricultura Semiperene	0,672	0,02	0,078	0,218	0,012
3°	06/09/2006	29/08/2007	Agricultura Semiperene	0,6	0,004	0,024	0,338	0,034
4°	06/09/2007	28/08/2008	Pastagem	0,16	0,068	0,106	0,652	0,014
5°	05/09/2008	29/08/2009	Agricultura Temp 1 ciclo	0,094	0,522	0,28	0,092	0,012
6°	06/09/2009	29/08/2010	Pastagem	0,366	0,168	0,058	0,402	0,006
7°	06/09/2010	29/08/2011	Agricultura Semiperene	0,47	0,166	0,054	0,306	0,004
8°	06/09/2011	28/08/2012	Agricultura Semiperene	0,546	0,04	0,102	0,312	0
9°	05/09/2012	29/08/2013	Pastagem	0,372	0,044	0,086	0,474	0,024
10°	06/09/2013	29/08/2014	Pastagem	0,29	0,042	0,212	0,416	0,04
11°	06/09/2014	29/08/2015	Agricultura Temp 1 ciclo	0,226	0,328	0,26	0,168	0,018
12°	06/09/2015	28/08/2016	Agricultura Semiperene	0,62	0,036	0,026	0,306	0,012
13°	05/09/2016	29/08/2017	Pastagem	0,38	0,078	0,14	0,4	0,002
14°	06/09/2017	29/08/2018	Agricultura Semiperene	0,884	0,006	0,012	0,088	0,01
15°	06/09/2018	29/08/2019	Agricultura Semiperene	0,914	0,002	0,02	0,064	0
16°	06/09/2019	28/08/2020	Agricultura Semiperene	0,966	0	0,014	0,02	0
17°	05/09/2020	29/08/2021	Agricultura Semiperene	0,918	0,018	0,018	0,044	0,002
18°	06/09/2021	29/08/2022	Agricultura Temp 1 ciclo	0,09	0,62	0,062	0,226	0,002

Camada de Probabilidades - Barra dos Bugres-MT - Fig 21B

Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Pastagem	0,052	0,008	0,008	0,932	0
2°	06/09/2005	29/08/2006	Pastagem	0,054	0,002	0,064	0,88	0
3°	06/09/2006	29/08/2007	Pastagem	0,146	0,026	0,078	0,75	0
4°	06/09/2007	28/08/2008	Pastagem	0,25	0,016	0,178	0,556	0
5°	05/09/2008	29/08/2009	Pastagem	0,158	0,002	0,006	0,826	0,008
6°	06/09/2009	29/08/2010	Pastagem	0,002	0	0	0,998	0
7°	06/09/2010	29/08/2011	Agricultura Temp 1 ciclo	0,418	0,432	0,088	0,062	0
8°	06/09/2011	28/08/2012	Agricultura Semiperene	0,828	0,002	0,018	0,148	0,004
9°	05/09/2012	29/08/2013	Agricultura Semiperene	0,912	0,008	0,008	0,066	0,006
10°	06/09/2013	29/08/2014	Agricultura Semiperene	0,896	0	0,012	0,08	0,012
11°	06/09/2014	29/08/2015	Agricultura Semiperene	0,944	0,002	0,006	0,046	0,002
12°	06/09/2015	28/08/2016	Agricultura Semiperene	0,892	0,008	0,018	0,082	0
13°	05/09/2016	29/08/2017	Agricultura Semiperene	0,898	0,002	0,038	0,06	0,002
14°	06/09/2017	29/08/2018	Agricultura Semiperene	0,904	0,004	0,016	0,074	0,002
15°	06/09/2018	29/08/2019	Agricultura Semiperene	0,936	0,004	0,006	0,054	0
16°	06/09/2019	28/08/2020	Agricultura Semiperene	0,994	0	0,002	0,004	0
17°	05/09/2020	29/08/2021	Agricultura Semiperene	0,868	0,004	0,024	0,104	0
18°	06/09/2021	29/08/2022	Agricultura Semiperene	0,536	0,064	0,144	0,256	0

Camada de Probabilidades - Tangará da Serra-MT - Fig 21C

Ano-safra	Início	Fim	Classe	Agricultura Semiperene	Agricultura Temp 1 ciclo	Agricultura Temp 2 ciclos	Pastagem	Vegetacao Florestal
1°	05/09/2004	29/08/2005	Pastagem	0,03	0	0	0,956	0,014
2°	06/09/2005	29/08/2006	Pastagem	0,026	0	0,018	0,952	0,004
3°	06/09/2006	29/08/2007	Vegetacao Florestal	0,118	0,078	0,054	0,324	0,426
4°	06/09/2007	28/08/2008	Agricultura Semiperene	0,744	0,008	0,032	0,208	0,008
5°	05/09/2008	29/08/2009	Agricultura Semiperene	0,73	0,012	0,04	0,206	0,012
6°	06/09/2009	29/08/2010	Agricultura Semiperene	0,542	0	0,006	0,414	0,038
7°	06/09/2010	29/08/2011	Agricultura Semiperene	0,788	0,01	0,038	0,164	0
8°	06/09/2011	28/08/2012	Agricultura Semiperene	0,544	0,002	0,064	0,148	0,242
9°	05/09/2012	29/08/2013	Agricultura Semiperene	0,71	0,014	0,044	0,228	0,004
10°	06/09/2013	29/08/2014	Agricultura Temp 2 ciclos	0,106	0,01	0,784	0,08	0,02
11°	06/09/2014	29/08/2015	Agricultura Temp 2 ciclos	0,144	0,02	0,724	0,098	0,014
12°	06/09/2015	28/08/2016	Agricultura Temp 2 ciclos	0,092	0,012	0,634	0,238	0,024
13°	05/09/2016	29/08/2017	Agricultura Temp 2 ciclos	0,088	0,012	0,786	0,094	0,02
14°	06/09/2017	29/08/2018	Agricultura Temp 2 ciclos	0,14	0,004	0,7	0,15	0,006
15°	06/09/2018	29/08/2019	Agricultura Temp 2 ciclos	0,002	0,002	0,99	0,006	0
16°	06/09/2019	28/08/2020	Agricultura Temp 2 ciclos	0	0	0,998	0,002	0
17°	05/09/2020	29/08/2021	Agricultura Temp 2 ciclos	0,034	0,018	0,926	0,022	0
18°	06/09/2021	29/08/2022	Agricultura Temp 2 ciclos	0,008	0	0,978	0,014	0

<b>Camada de Probabilidades - Campo Novo do Parecis-MT - Fig 22A</b>									
<b>Ano-safra</b>	<b>Início</b>	<b>Fim</b>	<b>Classe</b>	<b>Agricultura Semiperene</b>	<b>Agricultura Temp 1 ciclo</b>	<b>Agricultura Temp 2 ciclos</b>	<b>Pastagem</b>	<b>Vegetacao Florestal</b>	
1°	05/09/2004	29/08/2005	Agricultura Temp 1 ciclo	0,046	0,848	0,048	0,056	0,002	
2°	06/09/2005	29/08/2006	Agricultura Temp 1 ciclo	0,112	0,488	0,262	0,136	0,002	
3°	06/09/2006	29/08/2007	Agricultura Temp 1 ciclo	0,106	0,498	0,048	0,346	0,002	
4°	06/09/2007	28/08/2008	Agricultura Temp 2 ciclos	0,022	0,022	0,942	0,014	0	
5°	05/09/2008	29/08/2009	Agricultura Temp 2 ciclos	0,098	0,08	0,666	0,15	0,006	
6°	06/09/2009	29/08/2010	Agricultura Temp 2 ciclos	0,08	0,304	0,472	0,142	0,002	
7°	06/09/2010	29/08/2011	Agricultura Temp 1 ciclo	0,006	0,976	0,016	0,002	0	
8°	06/09/2011	28/08/2012	Agricultura Temp 2 ciclos	0,19	0,038	0,55	0,218	0,004	
9°	05/09/2012	29/08/2013	Agricultura Temp 2 ciclos	0,036	0,01	0,928	0,026	0	
10°	06/09/2013	29/08/2014	Agricultura Temp 2 ciclos	0,126	0,012	0,766	0,096	0	
11°	06/09/2014	29/08/2015	Agricultura Temp 1 ciclo	0,092	0,646	0,128	0,132	0,002	
12°	06/09/2015	28/08/2016	Agricultura Temp 2 ciclos	0,058	0,14	0,718	0,082	0,002	
13°	05/09/2016	29/08/2017	Agricultura Temp 2 ciclos	0,01	0	0,982	0,008	0	
14°	06/09/2017	29/08/2018	Agricultura Temp 2 ciclos	0	0	1	0	0	
15°	06/09/2018	29/08/2019	Agricultura Temp 2 ciclos	0	0	1	0	0	
16°	06/09/2019	28/08/2020	Agricultura Temp 2 ciclos	0	0	1	0	0	
17°	05/09/2020	29/08/2021	Agricultura Temp 2 ciclos	0,002	0,002	0,994	0,002	0	
18°	06/09/2021	29/08/2022	Agricultura Temp 2 ciclos	0,034	0,02	0,898	0,048	0	

<b>Camada de Probabilidades - Diamantino-MT - Fig 22B</b>									
<b>Ano-safra</b>	<b>Início</b>	<b>Fim</b>	<b>Classe</b>	<b>Agricultura Semiperene</b>	<b>Agricultura Temp 1 ciclo</b>	<b>Agricultura Temp 2 ciclos</b>	<b>Pastagem</b>	<b>Vegetacao Florestal</b>	
1°	05/09/2004	29/08/2005	Agricultura Temp 2 ciclos	0,034	0,3	0,572	0,094	0	
2°	06/09/2005	29/08/2006	Agricultura Temp 2 ciclos	0,012	0,002	0,958	0,028	0	
3°	06/09/2006	29/08/2007	Agricultura Temp 1 ciclo	0,038	0,716	0,178	0,068	0	
4°	06/09/2007	28/08/2008	Agricultura Temp 2 ciclos	0,058	0,012	0,868	0,062	0	
5°	05/09/2008	29/08/2009	Agricultura Temp 1 ciclo	0,028	0,784	0,168	0,018	0,002	
6°	06/09/2009	29/08/2010	Agricultura Temp 2 ciclos	0,024	0,336	0,546	0,094	0	
7°	06/09/2010	29/08/2011	Agricultura Temp 2 ciclos	0,164	0,01	0,758	0,066	0,002	
8°	06/09/2011	28/08/2012	Agricultura Temp 2 ciclos	0,006	0,014	0,956	0,024	0	
9°	05/09/2012	29/08/2013	Agricultura Temp 2 ciclos	0,036	0,004	0,936	0,024	0	
10°	06/09/2013	29/08/2014	Agricultura Temp 2 ciclos	0,06	0,01	0,87	0,046	0,014	
11°	06/09/2014	29/08/2015	Agricultura Temp 2 ciclos	0,016	0,014	0,946	0,024	0	
12°	06/09/2015	28/08/2016	Agricultura Temp 2 ciclos	0	0	1	0	0	
13°	05/09/2016	29/08/2017	Agricultura Temp 2 ciclos	0,002	0	0,998	0	0	
14°	06/09/2017	29/08/2018	Agricultura Temp 2 ciclos	0,004	0	0,996	0	0	
15°	06/09/2018	29/08/2019	Agricultura Temp 2 ciclos	0,012	0,01	0,974	0,004	0	
16°	06/09/2019	28/08/2020	Agricultura Temp 2 ciclos	0	0	1	0	0	
17°	05/09/2020	29/08/2021	Agricultura Temp 2 ciclos	0,004	0	0,994	0,002	0	
18°	06/09/2021	29/08/2022	Agricultura Temp 2 ciclos	0,018	0	0,962	0,02	0	