

UNIVERSIDADE FEDERAL DO PAMPA

ANDRÉ ROSA SILVEIRA

**IDENTIFICAÇÃO DE ÁREAS
INFESTADAS POR CAPIM-ANNONI
POR MEIO DE ANÁLISE ESPECTRAL**

**Bagé
2025**

ANDRÉ ROSA SILVEIRA

**IDENTIFICAÇÃO DE ÁREAS
INFESTADAS POR CAPIM-ANNONI
POR MEIO DE ANÁLISE ESPECTRAL**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Orientador: Sandro da Silva Camargo
Coorientador: Naylor Bastiani Perez

**Bagé
2025**

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

S581i SILVEIRA, ANDRE ROSA

Identificação de Áreas infestadas por Capim-Annoni por meio
de análise espectral / ANDRE ROSA SILVEIRA.

86 p.

Dissertação(Mestrado)-- Universidade Federal do Pampa,
MESTRADO EM COMPUTAÇÃO APLICADA, 2025.

"Orientação: SANDRO DA SILVA CAMARGO".

1. Eragrostis plana Nees. 2. Sensoriamento Remoto. 3.
Aprendizado de máquina. 4. Validação cruzada. 5. Oversampling.
I. Título.

ANDRÉ ROSA SILVEIRA

IDENTIFICAÇÃO DE ÁREAS INFESTADAS POR CAPIM-ANNONI POR MEIO DE ANÁLISE ESPECTRAL

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Mestre em Computação Aplicada.

Dissertação defendida e aprovada em: 17 de outubro de 2025.

Banca examinadora:

Prof. Dr. Sandro da Silva Camargo
Orientador
(Unipampa)

Prof. Dr. Marcos Correa Neves
(Embrapa)

Prof. Dr. Fábio Marcelo Breunig
(UFPR)

Prof. Dr. Eduardo Ceretta Moreira
(Unipampa)

AGRADECIMENTO

Agradeço ao apoio de minha família que me permitiram dedicar o tempo necessário para a realização desse mestrado. Também à gerência da Caixa que permitiu fazer um horário diferenciado nos dias de aula.

RESUMO

A invasão de espécies exóticas representa uma das principais ameaças à biodiversidade. No Rio Grande do Sul, o capim-annoni, introduzido na década de 1950, ocupa cerca de 20% da vegetação campestre, destacando-se pela elevada capacidade de adaptação e ausência de inimigos naturais. Assim, ações de prevenção e controle são fundamentais para reduzir seus impactos. Este estudo teve como objetivo aplicar e comparar algoritmos de aprendizado de máquina na classificação espectral de amostras vegetais, visando à identificação do capim-annoni (*Eragrostis plana* Nees) em pastagens naturais do Bioma Pampa. Foram coletados dados de reflectância com espectrorradiômetro ASD FieldSpec 4 e, após pré-processamento, gerados três conjuntos de dados: 1.688 bandas, um com 30 bandas selecionadas pelo *Random Forest* e outro com 28 bandas descritas na literatura. Aplicaram-se *Random Forest*, RPart, Máquinas de Vetores de Suporte e Redes Neurais Artificiais, com validação cruzada *Leave-One-Out Cross Validation*. Para lidar com o desbalanceamento amostral, empregaram-se ADASYN e Mixup. RPart apresentou o melhor desempenho no conjunto com 30 bandas (acurácia balanceada de 0,727 e sensibilidade de 0,800) com ADASYN, seguido pelo *Random Forest*. Esses resultados indicam que a análise espectral tem potencial para diferenciar o capim-annoni de espécies nativas, mas os modelos ainda favorecem a classe majoritária, sugerindo a necessidade de estudos adicionais para aprimorar a identificação da classe Annoni.

Palavras-chave: *Eragrostis plana* Nees. Sensoriamento Remoto. Aprendizado de máquina. Validação cruzada. *Oversampling*.

ABSTRACT

The invasion of exotic species represents one of the main threats to biodiversity. In Rio Grande do Sul, *capim-annoni*, introduced in the 1950s, currently occupies about 20% of the grassland vegetation, standing out for its high adaptability and the absence of natural enemies. Therefore, prevention and control measures are essential to mitigate its impacts. This study aimed to apply and compare machine learning algorithms for spectral classification of plant samples, focusing on the identification of capim-annoni (*Eragrostis plana* Nees) in natural pastures of the Pampa Biome. Reflectance data were collected using an ASD FieldSpec 4 spectroradiometer and, after preprocessing, three datasets were generated: the full set with 1.688 bands, one with 30 bands selected by Random Forest, and another with 28 bands described in the literature. Random Forest, RPart, Support Vector Machines, and Artificial Neural Networks were applied with Leave-One-Out Cross Validation. To address sample imbalance, ADASYN and Mixup were employed. RPart achieved the best performance on the 30-band set (balanced accuracy of 0.727 and sensitivity of 0.800) with ADASYN, followed by Random Forest. These results indicate that spectral analysis has potential to discriminate *capim-annoni* from native species, although models still favor the majority class, highlighting the need for further studies to improve the identification of the *Annoni* class.

Keywords: *Eragrostis plana* Nees. Remote sensing. Machine learning. Oversampling. Cross-validation.

LISTA DE FIGURAS

| | | |
|-----------|---|----|
| Figura 1 | Exemplo de amostra com toiceiras de capim-annoni. | 19 |
| Figura 2 | Exemplo de amostra de Capim-annoni parcialmente seco após uso de herbicida. | 20 |
| Figura 3 | Fluxo da Revisão Bibliográfica. | 37 |
| Figura 4 | Localização da área de coleta de dados espectrais na Embrapa Pecuária Sul, Bagé – RS. | 47 |
| Figura 5 | Espectrorradiômetro ASD FieldSpec® 4 Standard-Res utilizado na coleta dos dados espectrais. | 48 |
| Figura 6 | Espectros representativos das amostras antes e após o pré-processamento. As áreas sombreadas indicam as regiões de forte absorção atmosférica pela umidade, removidas do conjunto de dados conforme Thenkabail (2018) e Macedo (2024). | 55 |
| Figura 7 | Árvore de decisão gerada com o conjunto de 28 bandas da literatura, cenário binário. | 68 |
| Figura 8 | Árvore de decisão gerada com o conjunto de 28 bandas da literatura, cenário triclasse. | 69 |
| Figura 9 | Árvore de decisão gerada com as 30 bandas selecionadas pelo <i>Random Forest</i> , cenário binário. | 70 |
| Figura 10 | Árvore de decisão gerada com as 30 bandas selecionadas pelo <i>Random Forest</i> , cenário triclasse. | 70 |
| Figura 11 | Árvore de decisão gerada com o conjunto completo de 1688 bandas, cenário binário. | 71 |
| Figura 12 | Árvore de decisão gerada com o conjunto completo de 1688 bandas, cenário triclasse. | 72 |

LISTA DE TABELAS

| | | |
|-----------|---|----|
| Tabela 1 | Critérios de qualidade utilizados na seleção dos trabalhos na revisão sistemática..... | 36 |
| Tabela 2 | Trabalhos correlatos identificados na revisão bibliográfica. | 44 |
| Tabela 3 | Bibliotecas R Utilizadas nos Experimentos | 52 |
| Tabela 4 | Resultados dos modelos utilizando 1688 comprimentos de onda e análises com SAM | 61 |
| Tabela 5 | Matriz de confusão do modelo SVM no conjunto binário com 1688 bandas. | 61 |
| Tabela 6 | Métricas de desempenho dos modelos de classificação do conjunto com 30 comprimentos de onda obtidos do Random Forest..... | 65 |
| Tabela 7 | Matriz de confusão do modelo RPart no cenário binário com ADASYN no conjunto de 30 bandas (cenário binário). | 66 |
| Tabela 8 | Resultados dos modelos de classificação sob diferentes técnicas de oversampling no conjunto com 28 bandas da literatura. | 66 |
| Tabela 9 | Matriz de confusão do modelo RPart no conjunto de 28 bandas (cenário binário)..... | 67 |
| Tabela 10 | Bandas com forte correlação (distância até 10 nm) entre os conjuntos analisados..... | 75 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|--------|---|
| ADASYN | <i>Adaptive Synthetic Sampling</i> (Amostragem Sintética Adaptativa) |
| CPU | <i>Central Processing Unit</i> (Unidade Central de Processamento) |
| GPU | <i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico) |
| ha | Hectares |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| IDE | <i>Integrated Development Environment</i> (Ambiente de Desenvolvimento Integrado) |
| LDA | <i>Linear Discriminant Analysis</i> (Análise Discriminante Linear) |
| LOOCV | <i>Leave-One-Out Cross-Validation</i> (Validação Cruzada Leave-One-Out) |
| NIR | <i>Near Infrared</i> (Infravermelho Próximo) |
| PDU | Unidade de Discriminação de Plantas (<i>Plant Discrimination Unit</i>) |
| RGB | <i>Red Green Blue</i> (Vermelho Verde Azul) |
| RNC | Redes Neurais Convolucionais |
| RPA | <i>Remotely Piloted Aircraft</i> (Veículo Aéreo Remotamente Pilotado) |
| RPart | <i>Recursive Partitioning and Regression Trees</i> |
| RNA | <i>Artificial Neural Networks</i> (Redes Neurais Artificiais) |
| SDM | <i>Shoot Dry Mass</i> (Massa Seca da Parte Aérea) |
| SR | Sensoriamento Remoto |
| SRH | Sensoriamento Remoto Hiperespectral |
| SVM | <i>Support Vector Machine</i> (Máquina de Vetores de Suporte) |
| SWIR | <i>Short-Wave Infrared</i> (Infravermelho de Ondas Curtas) |
| UAV | <i>Unmanned Aerial Vehicle</i> (Veículo Aéreo Não Tripulado) |
| VIS | <i>Visible</i> (Visível) |

SUMÁRIO

| | |
|--|----|
| 1 INTRODUÇÃO | 12 |
| 1.1 Problema de Pesquisa..... | 14 |
| 1.2 Hipótese de Pesquisa | 14 |
| 1.3 Objetivos..... | 14 |
| 1.4 Organização do Trabalho | 15 |
| 2 REVISÃO BIBLIOGRÁFICA | 16 |
| 2.1 Bioma Pampa | 16 |
| 2.2 Pastagens | 17 |
| 2.3 Plantas Exóticas Invasoras..... | 18 |
| 2.4 <i>Eragrostis plana</i> Nees (Capim-Annoni) | 18 |
| 2.5 Técnicas de Monitoramento da Vegetação | 21 |
| 2.6 Aparato Tecnológico | 23 |
| 2.7 Sensoriamento Remoto | 23 |
| 2.8 Espectrometria..... | 25 |
| 2.9 Aplicação de Aprendizado de Máquina..... | 25 |
| 2.9.1 <i>Spectral Angle Mapper</i> (SAM) | 26 |
| 2.9.2 Algoritmo do RPart..... | 26 |
| 2.9.3 <i>Random Forest</i> | 27 |
| 2.9.4 Máquinas de Vetores de Suporte (<i>Support Vector Machines</i> - SVM) | 27 |
| 2.9.5 Redes Neurais Artificiais (RNA) | 28 |
| 2.10 Técnicas de Validação e Balanceamento | 29 |
| 2.10.1 Divisão <i>Hold-out</i> (Treinamento/Teste) | 29 |
| 2.10.2 Validação Cruzada <i>k-Fold</i> | 29 |
| 2.10.3 <i>Leave-One-Out Cross Validation</i> (LOOCV) | 30 |
| 2.10.4 <i>Oversampling</i> | 31 |
| 2.10.4.1 Mixup..... | 31 |
| 2.10.4.2 <i>Adaptive Synthetic Sampling Approach</i> (ADASYN) | 32 |
| 2.10.5 Métricas de Avaliação de Modelos | 33 |
| 2.11 Revisão Sistemática da Literatura | 33 |
| 2.12 Trabalhos Correlatos..... | 35 |
| 3 METODOLOGIA | 45 |
| 3.1 Caracterização da Pesquisa | 45 |
| 3.2 Delimitação do Problema | 45 |
| 3.3 Delineamento Experimental | 46 |
| 3.3.1 Coleta de Dados Espectrais..... | 46 |
| 3.4 Riscos e Dificuldades | 48 |
| 3.4.1 Análise de Dados de Espectrometria | 49 |
| 3.5 Ambiente de Desenvolvimento..... | 50 |
| 3.6 Processo de desenvolvimento | 51 |
| 3.6.1 Justificativa dos Modelos Utilizados | 52 |
| 4 RESULTADOS E DISCUSSÕES | 55 |
| 4.1 Tratamento dos Dados..... | 55 |
| 4.2 Análise Estatística..... | 56 |
| 4.3 Seleção de Atributos | 57 |
| 4.4 Modelagem e Avaliação dos Classificadores..... | 58 |
| 4.4.1 Algoritmos Classificatórios Utilizados | 58 |
| 4.4.2 Validação dos Modelos e Tratamento de Desbalanceamento | 59 |

| | | |
|----------------|---|-----------|
| 4.4.3 | Resultados dos Modelos de Classificação | 60 |
| 4.4.3.1 | Desempenho com o conjunto completo (1.688 bandas)..... | 60 |
| 4.4.3.2 | Desempenho com o conjunto reduzido via <i>Random Forest</i> (30 bandas)... | 63 |
| 4.4.3.3 | Desempenho com o conjunto baseado na literatura (28 bandas)..... | 66 |
| 4.4.3.4 | Análise das Árvores de Decisão | 67 |
| 4.4.3.5 | Análise das abordagens de classificação | 73 |
| 5 | CONCLUSÕES | 76 |
| | REFERÊNCIAS..... | 78 |
| | APÊNDICE A – TABELAS DAS ANÁLISES ESTATÍSTICAS E | |
| | IMPORTÂNCIA DOS COMPRIMENTOS DE ONDA..... | 85 |

1 INTRODUÇÃO

A invasão de espécies exóticas é considerada a segunda maior ameaça à biodiversidade global, ficando atrás apenas da destruição de habitats provocada por atividades humanas. Essas espécies comprometem os serviços ecossistêmicos e reduzem a diversidade biológica nativa por meio de mecanismos como predação, hibridização, competição e efeitos indiretos. Frequentemente, apresentam vantagens competitivas por não possuírem predadores naturais, o que favorece sua proliferação. Além disso, sua capacidade de explorar eficientemente os recursos disponíveis e ocupar nichos ecológicos pouco aproveitados contribui para a modificação dos biomas e das funções ecológicas (FARINHA *et al.*, 2023).

No estado do Rio Grande do Sul, uma das espécies exóticas invasoras mais preocupantes é o capim-annoni (*Eragrostis plana*), originário da África. Introduzido na década de 1950 com finalidades forrageiras, a espécie foi inicialmente considerada promissora para pastagens. No entanto, avaliações agronômicas revelaram sua inadequação ao pastejo, devido à baixa qualidade nutricional e alta resistência à tração mecânica, resultando em baixo desempenho animal. Com rusticidade elevada, prolificidade acentuada e adaptação a solos de baixa fertilidade, o capim-annoni passou a se comportar como uma planta agressivamente invasora (GOULART *et al.*, 2009).

Segundo Medeiros *et al.* (2014) estima-se que cerca de 20% da vegetação campestre do Rio Grande do Sul esteja infestada por capim-annoni, totalizando aproximadamente 3,1 milhões de hectares. A planta é capaz de produzir, em média, 300 mil sementes por indivíduo, e essas sementes podem permanecer viáveis no solo por até 20 anos, dependendo da profundidade em que se encontram. Essa característica torna o controle extremamente desafiador. Além disso, a espécie apresenta efeitos alelopáticos que dificultam a germinação de plantas nativas, comprometendo a regeneração natural das áreas invadidas (MEDEIROS *et al.*, 2014).

A expansão das fronteiras agrícolas e pecuárias tem intensificado a degradação ambiental, criando condições favoráveis à disseminação do capim-annoni. Áreas submetidas ao sobre-pastoreio, à agricultura intensiva ou situadas à margem de rodovias são especialmente suscetíveis à infestação segundo González (2017). Diante desse cenário, torna-se urgente a formulação de estratégias eficazes de controle, com ênfase na prevenção e, sobretudo, na detecção precoce da espécie invasora.

A identificação e o monitoramento das áreas infestadas são fundamentais para a

adoção de medidas de manejo que minimizem os impactos ambientais e econômicos. Conforme destacado por Medeiros e Focht (2007), a prevenção é a abordagem mais eficaz no combate ao capim-annoni, pois, uma vez estabelecido o banco de sementes no solo, sua erradicação torna-se extremamente difícil. Práticas como pastejo rotativo e suplementação alimentar podem atenuar os efeitos da infestação, mas dependem do conhecimento preciso da distribuição da planta. Estudos conduzidos por Amziane *et al.* (2021) e por Holmes *et al.* (2019) utilizaram análise espectral para detecção de plantas daninhas tendo obtido resultados promissores que sugerem que essa abordagem pode distinguir o capim-annoni de outras plantas.

A necessidade de métodos mais eficientes de identificação e monitoramento torna-se evidente diante da ampla distribuição do capim-annoni e das limitações das abordagens tradicionais baseadas em levantamentos de campo. Técnicas de sensoriamento remoto e análise espectral têm se destacado justamente por permitirem a detecção de diferenças fisiológicas, bioquímicas e estruturais entre espécies vegetais, já que propriedades como teor de pigmentos, conteúdo hídrico e características anatômicas influenciam diretamente a forma das curvas de reflectância conforme estudos de Asner (2008), Ustin (2013), Thenkabail, Lyon e Huete (2018b). Essas diferenças, muitas vezes imperceptíveis a olho nu, tornam a análise espectral uma ferramenta promissora para a detecção de plantas invasoras. Além disso, algoritmos de aprendizado de máquina têm demonstrado grande potencial para explorar dados de alta dimensionalidade e auxiliar em aplicações de agricultura de precisão conforme estudo de Liakos *et al.* (2018), reforçando a relevância de integrar métodos computacionais às estratégias de monitoramento do capim-annoni.

Nesse contexto, este trabalho propõe o uso da análise espectral combinada com técnicas de aprendizado de máquina para a identificação do capim-annoni. Dada a complexidade dos dados espectrais, caracterizados por alta dimensionalidade, presença de ruído e forte correlação entre variáveis, algoritmos de aprendizado de máquina têm se mostrado promissores em aplicações da agricultura de precisão como apontado por Liakos *et al.* (2018). Diferentemente dos estudos anteriores que utilizavam conjuntos restritos de bandas espectrais, como os estudados por Neves *et al.* (2019), esta pesquisa emprega 1.688 comprimentos de onda coletados em campo por meio de espectrorradiômetro.

1.1 Problema de Pesquisa

É viável empregar técnicas de aprendizado de máquina e análise espectral para distinguir o capim-annoni de outras espécies nativas do Bioma Pampa?

1.2 Hipótese de Pesquisa

A hipótese deste estudo é que a análise de dados espectrais, associada a algoritmos de aprendizado de máquina e técnicas estatísticas, permite distinguir com precisão o capim-annoni de outras espécies vegetais com características espectrais semelhantes.

1.3 Objetivos

O objetivo geral é aplicar e comparar diferentes algoritmos de aprendizado de máquina na classificação espectral de amostras vegetais, explorando a riqueza espectral disponível para auxiliar na distinção das amostras de capim-annoni das demais espécies do Bioma Pampa em pastagens naturais do sul do Brasil.

Objetivos específicos

- Avaliar o desempenho de algoritmos supervisionados como RPart, *Random Forest*, *Support Vector Machine* (SVM, Máquina de Vetores de Suporte) e Redes Neurais Artificiais (RNA) na classificação espectral do capim-annoni;
- Identificar os comprimentos de onda mais relevantes para a distinção entre capim-annoni e demais espécies;
- Investigar a viabilidade de aplicação dos modelos em conjuntos de dados reduzidos, utilizando técnicas de validação cruzada (LOOCV) e *Oversampling* (*Mixup* e *ADASYN*).

1.4 Organização do Trabalho

Este trabalho está estruturado da seguinte forma: o Capítulo 2 apresenta a revisão da literatura e os trabalhos correlatos. O Capítulo 3 descreve a metodologia adotada. O Capítulo 4 detalha a abordagem proposta e os resultados obtidos. Por fim, o Capítulo 5 apresenta as conclusões incluindo as contribuições e limitações.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta a fundamentação teórica e os principais conceitos relacionados à pesquisa. Inicia-se com uma descrição geral do Bioma Pampa, destacando sua biodiversidade e relevância ecológica. Em seguida, discute-se a problemática das plantas exóticas invasoras, com ênfase especial no *Eragrostis plana* Nees (capim-annoni), espécie que tem provocado sérios impactos ecológicos e econômicos na região Sul do Brasil. São abordadas também as principais estratégias de controle e manejo, bem como as aplicações de tecnologias como o Sensoriamento Remoto (SR) e os algoritmos de aprendizado de máquina na identificação e monitoramento de espécies invasoras. Por fim, são apresentados trabalhos correlatos que exploram abordagens semelhantes à proposta deste estudo.

2.1 Bioma Pampa

Os campos do sul da América do Sul são os principais campos temperados do planeta, uma vez que apresentam a maior diversidade de gramíneas do mundo. Essa região abrange cerca de 1 milhão de quilômetros quadrados, distribuídos entre Argentina, Brasil e Uruguai. De acordo com o mapa de biomas do Instituto Brasileiro de Geografia e Estatística (IBGE), no Rio Grande do Sul, os campos de altitude integram a Mata Atlântica, enquanto os campos subtropicais fazem parte do Bioma Pampa. Sua área corresponde a aproximadamente 63% do território gaúcho, totalizando 178.243 km² (GONZÁLEZ, 2017).

De acordo com estudo de Machado (2020), no Bioma Pampa, é possível encontrar uma grande variedade de espécies, entre as quais se destacam as *Asteraceae*, principalmente *Baccharis spp.*, *Eupatorium spp.*, *Hypochaeris spp.* e *Vernonia spp.*, que podem ocorrer isoladamente entre as gramíneas ou nas margens das estradas. As leguminosas dos gêneros *Adesmia spp.*, *Lathyrus spp.*, *Trifolium spp.* e *Vicia spp.* também têm presença relevante, geralmente de forma isolada. Já as ciperáceas, mais adaptadas a áreas úmidas, podem formar densas populações. Na primavera, as gramíneas mais representativas são *Briza spp.*, *Poa spp.*, *Bromus spp.* e *Piptochaetium spp.*; no verão, destacam-se *Bothriochloa spp.*, *Panicum spp.*, *Digitaria spp.*, *Setaria spp.* e *Paspalum spp.* (MACHADO, 2020).

Segundo estudo de González (2017), entre as principais atividades agrícolas no

Bioma Pampa estão os cultivos de soja, trigo, arroz, milho e a silvicultura com espécies exóticas de *Pinus spp.* e *Eucalyptus spp.*, além da pecuária, que é a principal atividade desenvolvida sobre a vegetação natural da região. O avanço acelerado da fronteira agrícola e pecuária, os cultivos mecanizados, o sobre-pastoreio e a invasão de plantas exóticas estão entre as maiores ameaças ao Bioma Pampa (GONZÁLEZ, 2017).

De acordo com estudo de Matei e Filippi (2012), a área do Bioma Pampa localizada no sul do Rio Grande do Sul, anteriormente dominada pela pecuária, tem passado por uma significativa expansão das monoculturas de soja e trigo, além de receber investimentos estrangeiros em plantios de espécies exóticas, como *Pinus spp.* e *Eucalyptus spp.*. Essas transformações têm promovido um desenvolvimento desequilibrado, refletindo padrões observados em outros biomas. A introdução de espécies exóticas invasoras e a expansão das monoculturas têm resultado em degradação ambiental, incluindo erosão hídrica e dessecamento dos solos, agravados pela mecanização em áreas de solos frágeis. A biodiversidade do Pampa encontra-se ameaçada, e a literatura aponta que a limitada disponibilidade de informações sobre sua fauna e flora, somada à reduzida extensão de áreas protegidas, contribui para que o bioma seja menos investigado e conservado em relação a outros brasileiros (MATEI; FILIPPI, 2012).

Diante desses desafios, é fundamental buscar um equilíbrio entre as necessidades sociais e econômicas e a conservação ambiental. Compreender as características da vegetação de campos nativos do Pampa é fundamental para contextualizar a distribuição de espécies e orientar a aplicação de técnicas de análise espectral para a identificação de espécies invasoras, tema que será abordado na próxima seção.

2.2 Pastagens

As pastagens podem ser classificadas, em termos gerais, como cultivadas ou nativas. As pastagens cultivadas correspondem a áreas onde forrageiras são semeadas com o objetivo de alimentar animais, especialmente em sistemas de pecuária de corte. Questões relacionadas ao manejo podem afetar a cobertura vegetal e a diversidade de espécies, influenciando a presença de plantas exóticas, como o capim-annoni (COSTA, 2003; GONZÁLEZ; FONSECA; PEREZ, 2017).

A estrutura vegetativa das pastagens naturais reflete uma grande diversidade, influenciada por fatores como clima, tipo de solo e manejo regional. Essas áreas abrigam

uma rica variedade de espécies adaptadas às condições locais, resultado da complexa interação entre os fatores ambientais e o manejo aplicado às pastagens.

Essa riqueza vegetativa, no entanto, enfrenta desafios, especialmente devido à invasão por espécies exóticas. O impacto dessas plantas invasoras nas pastagens, bem como as estratégias de manejo adotadas para enfrentá-lo, será explorado na próxima seção (GONZÁLEZ; FONSECA; PEREZ, 2017; SILVA, 2023).

2.3 Plantas Exóticas Invasoras

As plantas exóticas invasoras são espécies introduzidas fora de sua área de distribuição natural, que conseguem se estabelecer, se multiplicar e impactar negativamente os ecossistemas locais. Sua presença compromete a biodiversidade, altera processos ecológicos e pode causar prejuízos à agropecuária, especialmente quando colonizam áreas de pastagens mal manejadas (COSTA, 2003; MACHADO, 2020).

O avanço dessas espécies está frequentemente associado a distúrbios ambientais, como o sobrepastoreio, que reduzem a cobertura vegetal e abrem espaço para sua proliferação. Uma vez estabelecidas, essas plantas tendem a formar populações densas e persistentes, dificultando a regeneração da vegetação nativa e reduzindo o valor forrageiro da pastagem (GONZÁLEZ; FONSECA; PEREZ, 2017).

Entre as espécies invasoras mais preocupantes no Bioma Pampa destaca-se o *Eragrostis plana* Nees (capim-annoni), que será tratado em detalhe na próxima seção.

2.4 *Eragrostis plana* Nees (Capim-Annoni)

O *Eragrostis plana* Nees (capim-annoni) é uma gramínea perene da família Poaceae, originária da África do Sul, que se destaca como uma das principais espécies invasoras do Bioma Pampa, especialmente no estado do Rio Grande do Sul. Sua rápida expansão nas pastagens decorre da elevada capacidade de propagação vegetativa e por sementes, além da adaptação a diferentes condições ambientais e solos degradados conforme estudado por Costa (2003), Machado (2020). Morfologicamente, apresenta porte ereto e hábito cespitoso, formando touceiras densas e persistentes, com colmos firmes e folhas estreitas, lineares e ásperas ao tato. A inflorescência é uma panícula aberta composta por espiguetas pequenas e numerosas, produtoras de grande quantidade

de sementes leves e viáveis, que se dispersam facilmente pelo vento e por animais. O sistema radicular é profundo e vigoroso, conferindo tolerância à seca e alta capacidade de rebrota após o corte ou o pastejo. Essas características, somadas à baixa palatabilidade e à produção de substâncias alelopáticas, favorecem a dominância da espécie em áreas de campo nativo e dificultam o seu controle (EMBRAPA, 1993).

Uma característica marcante do capim-annoni é a formação de touceiras (ou toiceiras) densas e vigorosas, que ocupam grandes áreas, dificultando o pastejo e o manejo das pastagens. Essas touceiras promovem a supressão das espécies nativas e diminuem a biodiversidade local, alterando significativamente a estrutura e a funcionalidade do ecossistema como estudados por González, Fonseca e Perez (2017). A Figura 1 ilustra a densidade típica dessas touceiras em áreas infestadas.

Figura 1 – Exemplo de amostra com touceiras de capim-annoni.



Fonte: Milkpoint¹.

O método Mirapasto apresentado por Perez (2022) constitui uma estratégia eficaz para o controle do capim-annoni em áreas de pastagem degradadas. Baseia-se no manejo químico direcionado, utilizando herbicidas seletivos, principalmente o glifosato, aplicados em doses e períodos específicos do ciclo fenológico da planta invasora. As aplicações são realizadas preferencialmente em estágios de rebrote ativo, quando a planta apresenta maior suscetibilidade à ação do herbicida, permitindo suprimir sua densidade e vigor, além de reduzir o banco de sementes no solo. Essa abordagem visa restabelecer o equilíbrio da comunidade vegetal e favorecer a reintrodução de espécies forrageiras desejáveis, promovendo a recuperação gradual da pastagem, como ilustrado na Figura 2, que mostra o capim parcialmente seco após a aplicação do herbicida (PEREZ, 2022).

¹<https://www.milkpoint.com.br/img/artigo/conteudo/62319/>.

Embora o presente estudo utilize dados espectrais obtidos em nível de campo por meio de um espectrorradiômetro ASD FieldSpec 4, o Sensoriamento Remoto representa uma ferramenta promissora para o monitoramento em larga escala do avanço do capim-annoni. Imagens orbitais e aéreas podem ser empregadas para distinguir áreas infestadas e acompanhar a resposta da vegetação a práticas de manejo, por meio da análise de assinaturas espectrais e de índices de vegetação. Nesse sentido, González (2017) aplicou técnicas de Sensoriamento Remoto para avaliar a suscetibilidade à invasão do capim-annoni no Bioma Pampa, demonstrando o potencial dessas ferramentas para mapear e monitorar a dinâmica espacial da espécie invasora. Assim, a integração do método Mirapasto com abordagens baseadas em Sensoriamento Remoto pode ampliar a eficiência das ações de controle e o planejamento de estratégias de manejo.

Figura 2 – Exemplo de amostra de Capim-annoni parcialmente seco após uso de herbicida.



Fonte: Perez (2022).

Além do controle químico, o Mirapasto incorpora o cultivo temporário de sorgo forrageiro durante o verão. O sorgo compete diretamente com o capim-annoni por recursos como água e nutrientes do solo e, ao mesmo tempo, melhora a qualidade da forragem disponível para os animais. Essa abordagem integrada não só reduz a infestação pela espécie invasora, mas também contribui para a restauração da saúde e da produtividade das pastagens, criando condições mais favoráveis para práticas de pecuária sustentável. A aplicação sistemática do Mirapasto fortalece a resiliência do ecossistema, reduzindo impactos negativos sobre a biodiversidade local e os recursos hídricos (PEREZ, 2022).

O capim-annoni (*Eragrostis plana* Nees) é uma gramínea perene, cespitosa e exótica originária da África do Sul, que se estabeleceu amplamente no Bioma Pampa. Segundo Medeiros *et al.* (2014), a espécie apresenta colmos eretos e delgados, podendo atingir entre 60 e 120 cm de altura, com folhas estreitas, rígidas e ásperas ao tato, de coloração verde-acinzentada. O sistema radicular é profundo e bem desenvolvido, conferindo elevada tolerância à seca e favorecendo o estabelecimento em solos pobres e compactados. O ciclo fenológico caracteriza-se por rápido crescimento vegetativo na primavera e no verão, seguido pela emissão de panículas abertas e ramificadas, com grande produção de sementes pequenas e leves, facilmente dispersas pelo vento e por animais. Essas características, associadas à sua alta capacidade de regeneração e competição, tornam o capim-annoni uma das principais espécies invasoras das pastagens sul-brasileiras. Dessa forma, o conhecimento detalhado sobre sua estrutura, folhagem e ciclo fenológico, aliado a métodos avançados de controle e monitoramento, é essencial para garantir a sustentabilidade das pastagens e a conservação do Bioma Pampa (GONZÁLEZ, 2017; PEREZ, 2022).

2.5 Técnicas de Monitoramento da Vegetação

Segundo Jones e Vaughan (2010), diversas estratégias podem ser aplicadas ao monitoramento da vegetação, muitas das quais tem potencial para detectar e acompanhar a disseminação de espécies invasoras como o capim-annoni (*Eragrostis plana* Nees):

- **Sensoriamento remoto:** Utiliza satélites, aeronaves ou drones com sensores ópticos e/ou de radar para registrar imagens, espectros ou pulsos da vegetação em diferentes comprimentos de onda. Essa abordagem permite analisar a saúde das plantas, a densidade da cobertura vegetal e a distribuição de espécies em grande escala, sendo especialmente útil para mapear áreas infestadas.
- **Monitoramento terrestre:** Baseia-se em observações diretas no campo realizadas por técnicos ou pesquisadores, possibilitando avaliações detalhadas da vegetação, como altura, cobertura e presença de espécies invasoras. Essa técnica é eficaz para identificar focos iniciais de infestação e validar dados de sensoriamento remoto.
- **Parcelas de amostragem:** Compreendem áreas previamente delimitadas onde se realizam medições sistemáticas da vegetação, como contagem de indivíduos, coleta de biomassa e análise do solo. São essenciais para estudos ecológicos mais

aprofundados sobre o impacto de espécies invasoras.

- **Sensores de campo:** Envolvem dispositivos instalados no solo ou nas plantas que registram dados como umidade, temperatura, taxa de transpiração e outros parâmetros fisiológicos. Esses sensores fornecem dados em tempo real que ajudam a compreender as condições ambientais que favorecem a propagação de espécies como o capim-annoni.
- **Monitoramento hidrológico:** Em ecossistemas onde a dinâmica da vegetação está associada ao regime hídrico, como zonas úmidas e áreas alagáveis, esse tipo de monitoramento fornece subsídios para avaliar como a disponibilidade de água influencia a presença e a dispersão de espécies exóticas.
- **Monitoramento participativo:** Refere-se ao envolvimento de comunidades locais na coleta de dados e na observação da vegetação. A colaboração popular, por meio de relatos, entrevistas ou uso de aplicativos móveis, é valiosa para ampliar a cobertura do monitoramento e facilitar a detecção precoce de invasões biológicas.

Os dados utilizados em análises espectrais podem ser obtidos tanto por meio do sensoriamento remoto quanto por equipamentos portáteis, como os espectrorradiômetros. Esses dispositivos permitem uma avaliação detalhada da vegetação, possibilitando a identificação de padrões espectrais característicos de espécies invasoras conforme estudo de Jones e Vaughan (2010). No sensoriamento remoto da vegetação, a resposta espectral é determinada pela interação da radiação eletromagnética com os pigmentos, a estrutura celular e o conteúdo de água das folhas. Em geral, a vegetação saudável apresenta baixa refletância nas regiões do azul (450 nm) e do vermelho (670 nm), devido à absorção pela clorofila, e um pico de refletância no verde (550 nm). No infravermelho próximo (700–1300 nm), a refletância é elevada em função da estrutura interna do mesófilo foliar, enquanto no infravermelho médio (1300–2500 nm) predominam feições de absorção associadas à água e à composição química das folhas conforme estudado por Thenkabail, Lyon e Huete (2018b). No caso do capim-annoni, características como a coloração verde-acinzentada, a presença de tricomas e a rigidez foliar podem modificar o comportamento espectral esperado para gramíneas nativas, resultando em menor refletância nas faixas do visível e variações na intensidade do pico no infravermelho próximo conforme estudo de Medeiros *et al.* (2014). Essas diferenças espectrais, quando analisadas por meio de sensores remotos ou espectrorradiômetros de alta resolução, podem permitir a discriminação de áreas infestadas e compreender a dinâmica de expansão da espécie no Bioma Pampa.

Dessa forma, a combinação de diferentes técnicas de monitoramento contribui significativamente para o manejo integrado da vegetação, especialmente no controle do capim-annoni, ao permitir a detecção precoce, o mapeamento de áreas afetadas e a avaliação da eficácia das ações de manejo. Os avanços nas técnicas de monitoramento da vegetação descritos anteriormente são possíveis graças à evolução dos aparatos tecnológicos utilizados na aquisição e processamento de dados. Na próxima seção, serão apresentados os principais equipamentos e tecnologias empregados no monitoramento da vegetação, com ênfase naqueles voltados à detecção espectral de espécies invasoras como o capim-annoni.

2.6 Aparato Tecnológico

O uso de recursos avançados é fundamental para o monitoramento eficaz e o manejo de ecossistemas naturais, especialmente no combate a espécies invasoras como o capim-annoni. Mouta *et al.* (2023) destacam que ferramentas de monitoramento e análise de dados podem fornecer informações detalhadas sobre a distribuição, densidade e impacto de espécies sobre a vegetação nativa. Essas ferramentas apresentam potencial para auxiliar na detecção precoce e no controle do capim-annoni em pastagens, possibilitando a implementação de estratégias de manejo mais adequadas e adaptáveis a diferentes contextos.

Os avanços nas técnicas de monitoramento descritos anteriormente estão diretamente associados ao desenvolvimento e à aplicação de recursos tecnológicos específicos. A seguir, serão apresentados os principais aparatos utilizados neste estudo, com ênfase no sensoriamento remoto, na espectrometria e na análise de dados, ferramentas fundamentais para a detecção e o controle do capim-annoni em áreas de pastagem no Bioma Pampa.

2.7 Sensoriamento Remoto

Thenkabail, Lyon e Huete (2018b) destacam que o Sensoriamento Remoto (SR) possui potencial para monitorar tanto a diversidade quanto o vigor da vegetação. Essa técnica é amplamente utilizada para analisar a estrutura da vegetação e detectar mudanças no ambiente, fornecendo informações sobre o estado geral da vegetação e

seu crescimento. Uma forma mais avançada, o Sensoriamento Remoto Hiperespectral (SRH), permite uma análise mais detalhada das características da vegetação, utilizando um espectro mais amplo e contínuo de comprimentos de onda para examinar relações entre características filogenéticas, taxonômicas e funcionais. Segundo a mesma fonte, a abordagem espectral do SRH oferece informações mais detalhadas e específicas, sendo apropriada para mapeamento da diversidade filogenética, embora sua aplicação dependa dos objetivos do estudo (THENKABAIL; LYON; HUETE, 2018b).

De acordo com a caracterização morfológica, o capim-annoni pode ser descrito como uma planta com folhas e dosséis erectófilos, com um ângulo de inclinação superior a 60° em relação ao solo. Essa característica é particularmente importante, pois diferentes ângulos de inclinação das folhas desempenham um papel crucial nos estudos de sensoriamento remoto da vegetação. No entanto, esses ângulos podem influenciar diretamente a resposta espectral dos alvos, o que pode representar uma limitação importante na detecção de certas espécies. No caso do capim-annoni, a estrutura mais ereta e a coloração verde-acinzentada contrastam com o padrão observado nas gramíneas nativas do Bioma Pampa, que geralmente apresentam folhas mais arqueadas e dosséis planófilos, com coloração verde intensa. Essas diferenças estruturais e ópticas alteram a forma como a radiação solar é refletida e dispersa, podendo resultar em variações na reflectância, principalmente nas faixas do visível e do infravermelho próximo. Assim, compreender essas distinções é essencial para interpretar corretamente as feições espectrais e aprimorar a discriminação entre áreas infestadas e pastagens nativas (FARINHA *et al.*, 2023; MEDEIROS *et al.*, 2014).

Além disso, o SR fornece informações importantes para monitorar grandes áreas com custos reduzidos. Esses dados são fundamentais para compreender a distribuição espacial e temporal da vegetação, tanto de espécies nativas quanto invasoras. Com essas informações, torna-se possível realizar estudos sobre processos invasivos e degradação dos ecossistemas. Várias informações podem ser obtidas com dados de sensores remotos, como: uso e ocupação da terra, cobertura vegetal, fenologia da vegetação, biomassa, produção primária líquida, teor de umidade na vegetação, topografia, corpos d'água e presença de neve (THENKABAIL, 2018).

Os dados para análise espectral podem ser obtidos por meio de sensoriamento remoto ou de equipamentos portáteis, como os espectrorradiômetros, sendo cruciais para a análise espectral discutida na próxima subseção sobre Espectrometria (GONZÁLEZ; FONSECA; PEREZ, 2017).

2.8 Espectrometria

A espectrometria consiste na medição quantitativa da energia eletromagnética refletida ou emitida por um material em diferentes comprimentos de onda, formando feições espectrais (ou bandas de absorção). No sensoriamento remoto, essa técnica permite detectar pequenas variações nas propriedades ópticas da vegetação, que podem ser associadas a diferenças morfológicas e bioquímicas entre espécies (MOAZZAM *et al.*, 2019).

No caso do capim-annoni, essas diferenças estão relacionadas, por exemplo, à composição química das folhas, ao teor de pigmentos e à estrutura celular, que são condicionadas pela fenologia e pelas condições ambientais de contorno, fatores que influenciam a forma como a luz é refletida em determinadas faixas do espectro. Assim, a análise espectrométrica permite a identificação de padrões específicos dessa espécie invasora e auxilia na sua distinção em relação às gramíneas nativas (ASNER, 2008; HE; GUO, 2010; USTIN, 2013).

Neste estudo, a variável analisada foi a reflectância, definida como a razão entre a radiação eletromagnética refletida por uma superfície e a radiação incidente sobre ela. A medição dessa variável em diferentes bandas espectrais fornece subsídios para a seleção de comprimentos de onda mais sensíveis à presença do capim-annoni (*Eragrostis plana*), aumentando a precisão dos métodos de detecção (ASNER, 2008; HE; GUO, 2010). O uso da espectrometria nesse contexto não apenas contribui para o reconhecimento da espécie, mas também representa um passo importante para o desenvolvimento de ferramentas de monitoramento que apoiem estratégias de manejo e controle de áreas infestadas, permitindo a identificação de padrões espectrais específicos e a distinção em relação às gramíneas nativas (USTIN, 2013; ASNER, 2008; FARINHA *et al.*, 2023).

2.9 Aplicação de Aprendizado de Máquina

A escolha por métodos como RPart, *Random Forest*, Support Vector Machine e Redes Neurais Artificiais (RNA) fundamenta-se em conceitos consolidados na literatura de aprendizado de máquina, conforme apresentado por Bishop (2006), que aborda os princípios teóricos e as propriedades desses algoritmos. Essa referência fornece o embasamento metodológico para a aplicação desses modelos ao problema estudado.

2.9.1 Spectral Angle Mapper (SAM)

O *Spectral Angle Mapper* (SAM) é uma técnica de classificação supervisionada amplamente utilizada em análises espectrais. O método compara vetores espectrais de amostras desconhecidas com vetores de referência correspondentes a classes conhecidas, calculando o ângulo entre eles. Quanto menor o ângulo formado, maior a similaridade espectral, independentemente da magnitude absoluta da reflectância conforme estudado por Kruse *et al.* (1993). Essa característica torna o SAM robusto a variações de iluminação e sombreamento, sendo útil para discriminar espécies com assinaturas espectrais distintas.

Em estudos de monitoramento de espécies invasoras, o SAM tem sido empregado para identificar assinaturas espectrais específicas em meio à vegetação nativa. Por exemplo, Cho, Mathieu e Debba (2009) aplicaram o SAM para diferenciar espécies invasoras de plantas nativas em dados hiperespectrais de campo, enquanto Veigas (2023) utilizou a técnica para monitorar a detecção de invasoras em ambientes aquáticos. Esses estudos demonstram o potencial do SAM para apoiar o monitoramento e manejo de plantas invasoras.

2.9.2 Algoritmo do RPart

O algoritmo *Recursive Partitioning and Regression Trees* (RPart) é amplamente utilizado na construção de árvores de decisão para problemas de classificação e regressão. No contexto deste estudo, o RPart é aplicado para gerar um modelo capaz de identificar o capim-annoni a partir de dados espectrais, como os comprimentos de onda. A árvore de decisão construída pelo algoritmo é composta por nós que representam divisões com base em variáveis explicativas, buscando separar as amostras em grupos o mais homogêneos possível com relação à classe da espécie (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

A construção da árvore tem início com a seleção da variável que melhor separa as classes, ou seja, aquela que maximiza a pureza dos subconjuntos resultantes. A partir desse nó inicial, o algoritmo continua a dividir os dados de forma recursiva, criando ramificações até que as amostras estejam adequadamente separadas em classes homogêneas ou até que um critério de parada seja atingido. Esses critérios podem incluir o número mínimo de amostras por folha, a profundidade máxima da árvore ou um limiar mínimo de ganho de informação em cada divisão (OKOYE; HOSSEINI, 2024).

Uma das principais vantagens do RPart para a classificação do capim-annoni

é a interpretabilidade do modelo resultante. As árvores de decisão permitem uma visualização clara das regras utilizadas para classificar as amostras, o que facilita a compreensão do processo decisório por especialistas da área. Além disso, o algoritmo é robusto a dados ruidosos e possui a capacidade de identificar as variáveis mais relevantes para a classificação, o que é especialmente vantajoso em estudos com elevado número de variáveis espectrais (FÁVERO; BELFIORE; SOUZA, 2023).

2.9.3 *Random Forest*

O *Random Forest* (RF) é um método de aprendizado de máquina do tipo *ensemble*, ou seja, utiliza a combinação de múltiplos modelos (neste caso, árvores de decisão) para produzir uma previsão mais precisa e estável. O algoritmo gera diversas árvores a partir de subconjuntos aleatórios dos dados e de variáveis explicativas, e combina suas previsões por meio de votação majoritária (no caso de classificação) ou média (em regressão) (YANO *et al.*, 2016; OK; AKAR; GÜNGÖR, 2012).

Essa abordagem melhora a capacidade de generalização do modelo e reduz o risco de *overfitting*, que ocorre quando o modelo se ajusta excessivamente aos dados de treinamento, capturando ruídos e padrões específicos, o que prejudica seu desempenho em novos dados. Como resultado, o RF torna-se mais robusto frente a ruídos nos dados e variações amostrais (YANO *et al.*, 2016).

2.9.4 Máquinas de Vetores de Suporte (*Support Vector Machines - SVM*)

As SVM são algoritmos de aprendizado supervisionado amplamente utilizados para tarefas de classificação e regressão. O princípio básico da SVM é encontrar um hiperplano ótimo que maximize a margem, ou seja, a distância entre o hiperplano e os pontos de dados mais próximos de cada classe, conhecidos como vetores de suporte. Um hiperplano com maior margem tende a melhorar a generalização do modelo, reduzindo o risco de erros em novas amostras (CORTES; VAPNIK, 1995; BHATTACHARYYA *et al.*, 2020).

Uma das principais vantagens da SVM é sua capacidade de lidar com problemas de alta dimensionalidade e com conjuntos de dados em que as classes não são linearmente separáveis. Nesses casos, a SVM utiliza funções de *kernel* para transformar o espaço de

entrada em um espaço de características de dimensão superior, onde a separação entre as classes se torna possível. Entre os *kernels* mais comuns estão o linear, o polinomial e o *Radial Basis Function* (RBF) (CORTES; VAPNIK, 1995).

Essa flexibilidade, aliada à boa capacidade de generalização, torna a SVM uma ferramenta eficaz para classificação espectral em estudos vegetacionais, como a identificação de espécies com assinaturas espectrais distintas (BHATTACHARYYA *et al.*, 2020).

2.9.5 Redes Neurais Artificiais (RNA)

As RNA são modelos computacionais inspirados no funcionamento do cérebro humano, capazes de aprender padrões complexos a partir de dados de entrada e produzir previsões ou classificações. Uma RNA é composta por unidades chamadas neurônios artificiais, organizadas em camadas: camada de entrada, camadas ocultas e camada de saída. Cada neurônio recebe sinais de entrada, aplica pesos e funções de ativação, e transmite um sinal para os neurônios subsequentes, permitindo que a rede capture relações não lineares entre variáveis (GOODFELLOW; BENGIO; COURVILLE, 2016).

O treinamento da RNA envolve ajustar os pesos sinápticos por meio de algoritmos de otimização, geralmente utilizando o método de retropropagação (*backpropagation*) aliado a uma função de custo que quantifica o erro entre a saída prevista e a saída desejada. Redes com múltiplas camadas ocultas são chamadas de redes neurais profundas (*deep neural networks*) e são particularmente eficazes em capturar padrões complexos em dados de alta dimensionalidade, como espectros hiperespectrais (LECUN; BENGIO; HINTON, 2015).

No contexto da detecção de plantas invasoras, RNAs têm se mostrado úteis por sua capacidade de modelar relações complexas entre variáveis espectrais e características morfológicas ou bioquímicas das plantas. Por exemplo, Azeem *et al.* (2020) demonstrou a eficácia de RNAs na estimativa da área foliar da planta invasora *Wedelia trilobata*, utilizando medidas não destrutivas de comprimento e largura das folhas. O modelo baseado em RNA apresentou melhor desempenho em comparação com métodos de regressão linear, com um coeficiente de determinação (R^2) de 0,96 e menores valores de *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE).

As RNAs podem ser aplicadas na modelagem de características morfológicas

de plantas invasoras, oferecendo subsídios para o desenvolvimento de estratégias de monitoramento e controle dessas espécies em contextos agrícolas (AZEEM *et al.*, 2020).

2.10 Técnicas de Validação e Balanceamento

A avaliação de modelos de classificação requer a aplicação de procedimentos de validação que permitam estimar seu desempenho de forma consistente e confiável. Além disso, quando os conjuntos de dados apresentam desbalanceamento entre classes, tornam-se necessárias estratégias adicionais para reduzir vieses e melhorar a capacidade preditiva dos algoritmos. Nesta seção são descritas as principais abordagens utilizadas neste trabalho, abrangendo técnicas de divisão dos dados para validação, bem como métodos de balanceamento aplicados aos conjuntos de treinamento.

2.10.1 Divisão *Hold-out* (Treinamento/Teste)

Uma das estratégias mais tradicionais de validação de modelos de aprendizado de máquina consiste na divisão dos dados em dois subconjuntos: treinamento e teste, conhecida como *hold-out*. Nessa abordagem, uma proporção dos dados, geralmente entre 60/40 e 80/20, é utilizada para o treinamento do modelo, enquanto o restante é reservado para avaliação do desempenho em dados não treinados (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Embora seja simples e amplamente empregada, essa estratégia pode levar a resultados instáveis em cenários de base de dados reduzida, pois a amostra de teste pode não ser suficientemente representativa. Para contornar essa limitação, abordagens como a validação cruzada *k-Fold* e o método *Leave-One-Out Cross Validation* passaram a ser frequentemente utilizadas em contextos científicos (KUHN; JOHNSON, 2013).

2.10.2 Validação Cruzada *k-Fold*

A validação cruzada *k-Fold* é uma técnica amplamente utilizada para avaliação de modelos de aprendizado de máquina, conforme descrito por Arlot e Celisse (2010). Nessa abordagem, o conjunto de dados é dividido em k subconjuntos aproximadamente do mesmo tamanho, denominados *folds*. O modelo é então treinado e testado k vezes,

sendo que, a cada iteração, um dos *folds* é utilizado como conjunto de teste, enquanto os $k-1$ *folds* restantes são usados para o treinamento.

Esse processo permite que cada amostra do conjunto de dados seja utilizada tanto para treinamento quanto para validação, proporcionando uma avaliação mais robusta e menos suscetível a variações na divisão dos dados. A média dos resultados obtidos em cada uma das k iterações é então calculada, oferecendo uma estimativa mais confiável do desempenho geral do modelo.

2.10.3 *Leave-One-Out Cross Validation (LOOCV)*

O *Leave-One-Out Cross Validation (LOOCV)* é uma técnica de validação cruzada amplamente utilizada em aprendizado de máquina e modelagem preditiva. Essa abordagem pertence à família de validação cruzada k -fold, mas com a particularidade de que $k = n$, sendo n o número total de amostras do conjunto de dados (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; NAJAFABADI; HESAMI; ESKANDARI, 2023).

No LOOCV, o modelo é treinado $n - 1$ vezes utilizando todas as amostras, exceto uma, que é reservada para teste. Esse processo é repetido n vezes, de modo que cada amostra do conjunto de dados seja utilizada exatamente uma vez como conjunto de teste e $n - 1$ vezes como parte do conjunto de treinamento. Essa metodologia permite avaliar o desempenho do modelo de maneira eficiente todo o conjunto de dados disponível (HAUG *et al.*, 2014).

Entre as principais vantagens do LOOCV, destaca-se o aproveitamento máximo da informação contida no conjunto de dados, o que é particularmente útil em situações com número reduzido de amostras. Além disso, o LOOCV fornece uma estimativa quase não enviesada do erro de generalização do modelo, permitindo uma avaliação detalhada de seu comportamento frente a diferentes instâncias (NAJAFABADI; HESAMI; ESKANDARI, 2023).

Entretanto, a técnica apresenta desvantagens relacionadas à complexidade computacional. Cada iteração exige o re-treinamento do modelo, o que pode tornar o LOOCV inviável para conjuntos de dados muito grandes ou modelos com custo elevado de treinamento. Por essa razão, em situações de grande volume de dados, métodos como k -fold cross validation com $k < n$ podem ser mais apropriados (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Além de sua aplicação teórica, o LOOCV tem sido amplamente documentado

em estudos de aprendizado de máquina, fornecendo uma base sólida para validação de modelos preditivos em diferentes contextos (HAUG *et al.*, 2014).

2.10.4 *Oversampling*

O *Oversampling* é uma técnica amplamente utilizada para lidar com o desbalanceamento de classes em problemas de classificação. Esse desbalanceamento ocorre quando há uma disparidade significativa na quantidade de amostras entre as classes, o que pode comprometer o desempenho dos modelos, sobretudo no reconhecimento da classe minoritária. Ao aplicar técnicas de *Oversampling*, busca-se aumentar artificialmente o número de amostras da classe menos representada, gerando novas instâncias que contribuam para o equilíbrio do conjunto de dados e a melhoria do aprendizado do modelo.

Entre as técnicas de *Oversampling*, Mixup estudado por (ZHANG *et al.*, 2017) e ADASYN estudado por (HE *et al.*, 2008) são frequentemente utilizadas para lidar com desequilíbrios amostrais em conjuntos de dados, cada uma com abordagens distintas de geração de amostras sintéticas.

2.10.4.1 *Mixup*

Considerando o tamanho reduzido do conjunto de dados, composto por apenas 46 amostras (ou 42 após a remoção de outliers), a técnica do *Mixup* pode ser utilizada para ampliar artificialmente o conjunto de treinamento. Conjuntos pequenos apresentam maior risco de *overfitting*, pois o modelo pode ajustar-se excessivamente às características específicas das amostras disponíveis, dificultando a generalização (SMUCNY *et al.*, 2022).

O *Mixup* cria instâncias sintéticas adicionais por meio da interpolação linear entre pares de espectros, aumentando a diversidade do conjunto de dados sem necessidade de coleta de novas amostras. Essa técnica suaviza as fronteiras de decisão do modelo, fornecendo mais exemplos para o aprendizado e reduzindo a probabilidade de sobreajuste (VALICHARLA *et al.*, 2024).

A interpolação no *Mixup* é controlada pelo parâmetro α , que define a distribuição Beta usada para combinar os espectros. Valores menores de α , como 0,2, geram instâncias próximas das amostras reais, preservando características espectrais relevantes,

ao mesmo tempo em que introduzem variações suficientes para melhorar o aprendizado do modelo (VALICHARLA *et al.*, 2024).

Além disso, a utilização do *Mixup* oferece vantagens adicionais, como a regularização natural do modelo. Em problemas de classificação baseados em dados espectrais, é comum a presença de variabilidades geradas por ruídos instrumentais ou ambientais. Ao suavizar essas variações por meio da interpolação entre amostras, o *Mixup* pode reduzir o impacto de ruídos e aumentar a capacidade de generalização do modelo quando aplicado a novos dados (SMUCNY *et al.*, 2022).

2.10.4.2 Adaptive Synthetic Sampling Approach (ADASYN)

O *Adaptive Synthetic Sampling Approach* (ADASYN) é uma técnica de *Oversampling* desenvolvida para lidar com conjuntos de dados desbalanceados em tarefas de classificação. A ideia central do ADASYN é gerar novas amostras sintéticas da classe minoritária, priorizando regiões do espaço de características onde essa classe é mais difícil de ser classificada. Dessa forma, o método concentra a criação de instâncias sintéticas nas áreas em que o modelo tem maior probabilidade de cometer erros, equilibrando o conjunto de treinamento de forma adaptativa (HE *et al.*, 2008).

Diferentemente de outras técnicas de *Oversampling*, como o *SMOTE* (*Synthetic Minority Over-sampling Technique*), que gera novas amostras de maneira uniforme ao longo do espaço da classe minoritária, o ADASYN ajusta a quantidade de amostras sintéticas geradas de acordo com a densidade local de amostras minoritárias. Essa abordagem permite que o modelo aprenda padrões mais representativos das regiões críticas do conjunto de dados, melhorando a capacidade de generalização e reduzindo o viés em favor da classe majoritária (HE *et al.*, 2008).

O ADASYN é particularmente útil em cenários de classificação em que a classe minoritária apresenta baixa representatividade, contribuindo para um treinamento mais equilibrado dos modelos e para a melhoria de métricas de desempenho como sensibilidade e acurácia. A técnica tem sido aplicada em diferentes contextos de aprendizado de máquina, incluindo detecção de padrões em dados desbalanceados e análise de fenótipos agrícolas com imagens obtidas por um drone (SARKAR, 2023; SALEEM *et al.*, 2023).

2.10.5 Métricas de Avaliação de Modelos

A avaliação do desempenho de algoritmos de classificação é fundamental para aferir a confiabilidade dos resultados obtidos. Para isso, utiliza-se a matriz de confusão, que relaciona as previsões realizadas pelo modelo com as classes verdadeiras dos dados. A partir dessa matriz, é possível derivar diferentes métricas de avaliação, comumente aplicadas em problemas de classificação binária (POWERS, 2020).

A acurácia corresponde à proporção de acertos do modelo em relação ao total de observações. A precisão (ou valor preditivo positivo) mede a proporção de amostras corretamente classificadas como positivas entre todas aquelas preditas como positivas. A sensibilidade (ou *recall*) representa a capacidade do modelo de identificar corretamente os casos positivos, enquanto a especificidade mede a capacidade de identificar corretamente os casos negativos. A medida-F1 é a média harmônica entre precisão e sensibilidade, sendo recomendada em situações de desbalanceamento entre classes (HICKS *et al.*, 2022).

Além dessas métricas, a análise dos erros também é relevante. Os falsos positivos (FP) correspondem às amostras incorretamente classificadas como pertencentes à classe positiva, enquanto os falsos negativos (FN) correspondem às amostras que pertenciam à classe positiva, mas foram classificadas como negativas. Esses indicadores permitem compreender melhor o tipo de erro cometido pelo modelo e suas implicações práticas (RAINIO *et al.*, 2024).

Embora tais métricas sejam tradicionalmente definidas para problemas binários, elas podem ser adaptadas para cenários triclasse. No presente estudo, adotou-se uma abordagem simplificada, agrupando as classes Annoni e Misto em uma única categoria (Não Nativo) e comparando-a com a classe Nativo, de modo a possibilitar o cálculo de falsos positivos e falsos negativos em um contexto multiclasse (POWERS, 2020).

2.11 Revisão Sistemática da Literatura

A revisão sistemática da literatura foi realizada por meio da ferramenta *Parsifal*². Foram utilizadas as seguintes expressões de busca:

("spectral analysis"OR "spectral reflectance") AND ("weeds"OR "invasive plants"OR "capim-annoni"OR "Eragrostis plana") AND ("detection"OR "control"OR

²Disponível em: <<https://parsif.al/>>. Acesso em: maio 2025.

"strategies") AND NOT("disease")

Essa expressão utilizou operadores booleanos para ampliar e refinar os resultados. Os termos "*spectral analysis*" e "*spectral reflectance*" foram combinados com o operador **OR**, permitindo recuperar publicações que empregassem diferentes expressões para descrever a análise espectral. O operador **AND** restringiu os resultados à intersecção entre os conjuntos, de modo a selecionar apenas trabalhos que tratassem simultaneamente da análise espectral e de plantas daninhas (*weeds*), espécies invasoras (*invasive plants*) ou, especificamente, o capim-annoni (*Eragrostis plana Nees*). Para abarcar tanto estudos voltados à identificação (*detection*) quanto às formas de manejo (*control* ou *strategies*), esses termos foram incluídos na string. Por fim, o operador **NOT** foi empregado para excluir trabalhos cujo foco estivesse em doenças de plantas (*disease*), uma vez que a maioria dos resultados iniciais se referia a esse tema, e não à identificação de espécies invasoras, que constitui o objetivo desta pesquisa.

As bases de dados selecionadas para a busca foram: *ScienceDirect*, *Springer Link*, *UFRGS*, *Embrapa*, *MDPI* e *IEEE Digital Library*. A escolha dessas bases se justifica pela sua relevância e complementaridade no escopo da pesquisa. A *ScienceDirect* e a *Springer Link* foram incluídas por reunirem periódicos de grande impacto internacional em áreas como ciências agrárias, ambientais e de sensoriamento remoto. A base da *Embrapa* foi considerada por ser referência nacional em estudos aplicados à agricultura e manejo de pastagens, incluindo espécies invasoras no Brasil. A *MDPI* foi selecionada pelo amplo acesso aberto e pela publicação de periódicos especializados em sensoriamento remoto, ecologia e uso de tecnologias para manejo de plantas. Por fim, a *IEEE Digital Library* foi incorporada por concentrar pesquisas na interface entre ciência da computação e engenharia, especialmente no uso de aprendizado de máquina e análise espectral, técnicas centrais para este estudo. Essas bases foram priorizadas em detrimento de outras por fornecerem melhor alinhamento ao objetivo da pesquisa, maior representatividade das áreas temáticas envolvidas e maior acesso ao texto completo dos artigos.

Além das bases de dados mencionadas, também foram consultados repositórios institucionais da *Universidade Federal do Rio Grande do Sul* (UFRGS). A inclusão dessa fonte se justifica pela tradição da instituição em pesquisas sobre o capim-annoni, espécie invasora de grande relevância para o contexto agropecuário do Sul do Brasil. Esse repositório reúne dissertações, teses e artigos científicos produzidos por grupos de pesquisa especializados, permitindo acesso a trabalhos que, muitas vezes, não estão indexados nas bases internacionais, mas que possuem elevada pertinência temática para o

presente estudo.

Para assegurar a relevância dos estudos incluídos na revisão sistemática e a consistência da análise, foram definidos critérios de inclusão e de exclusão.

Os critérios de inclusão determinaram quais trabalhos seriam considerados pertinentes ao escopo desta pesquisa, enquanto os de exclusão estabeleceram restrições para evitar resultados irrelevantes ou redundantes. Esses critérios estão listados a seguir:

Critérios de inclusão:

- Estudos que abordem a identificação ou detecção de plantas invasoras ou ervas daninhas por meio de análise espectral;
- Publicações que tratem especificamente do capim-annoni (*Eragrostis plana* Nees);
- Trabalhos que utilizem dados hiperespectrais para diferenciar plantas invasoras ou ervas daninhas de espécies nativas.

Critérios de exclusão:

- Estudos publicados até 2014, exceto se abordarem capim-annoni;
- Estudos que enfoquem exclusivamente a detecção de doenças em plantas por meio de análise espectral;
- Trabalhos que não abordem espectrometria, identificação de plantas invasoras ou ervas daninhas, ou que realizem a detecção exclusivamente por imagens;
- Publicações repetidas, versões preliminares de artigos, ou múltiplos trabalhos de mesma autoria com conteúdo redundante;
- Artigos redigidos em idiomas diferentes do inglês, português ou espanhol.

2.12 Trabalhos Correlatos

Durante a revisão sistemática da literatura foram identificados 596 trabalhos, nos repositórios mencionados na seção 2.11, que atenderam aos termos de busca. Após a aplicação dos critérios de inclusão e exclusão, 255 trabalhos foram rejeitados por não atenderem ao escopo definido. Em seguida, foram eliminados 119 trabalhos duplicados.

Os estudos remanescentes passaram por uma análise de qualidade, na qual foram avaliados de acordo com critérios previamente estabelecidos (clareza metodológica, relevância para o tema, consistência dos resultados e originalidade). Cada trabalho era avaliado com notas de 0 a 2 para cada pergunta, sendo 0 - Não atende, 1 - Atende

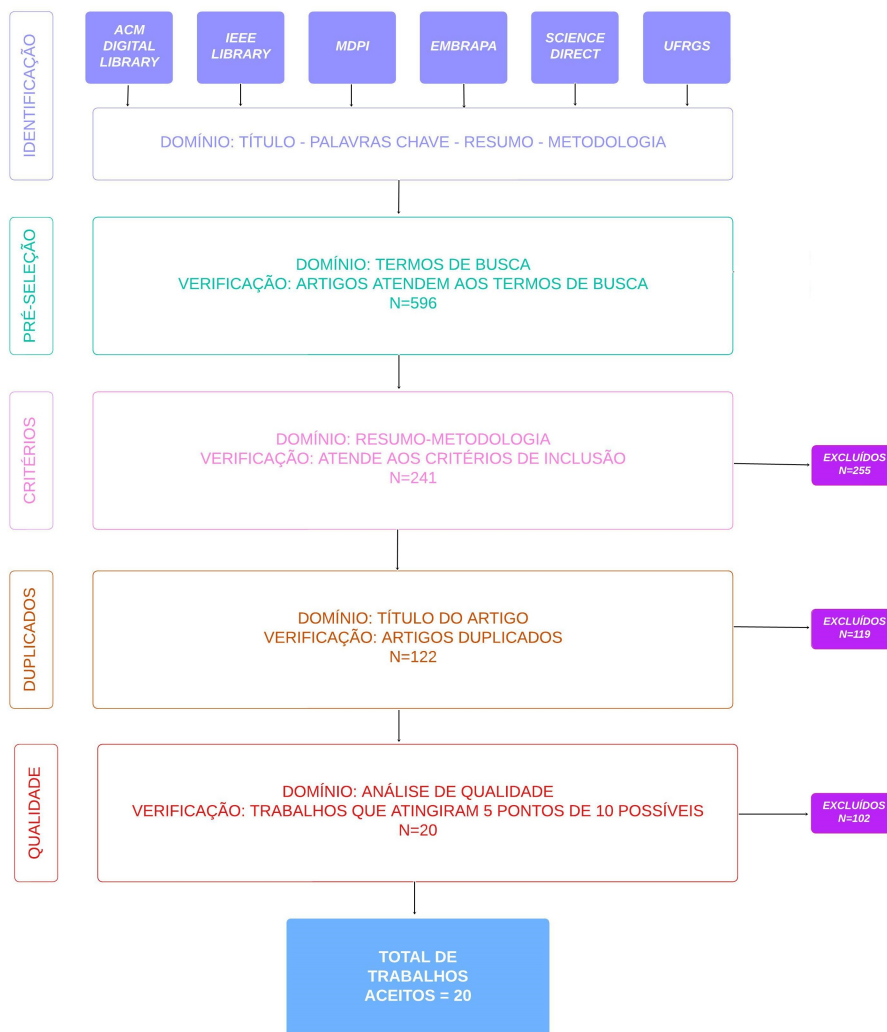
parcialmente, 2 Atende plenamente. Apenas aqueles que atingiram a pontuação mínima de 5 em uma escala de 0 a 10 foram considerados. A tabela 1 apresenta os critérios de qualidade utilizados nesse processo:

Ao final desse processo, 20 trabalhos foram aceitos e incluídos na revisão sistemática. A Figura 3 apresenta um resumo do processo de seleção.

Tabela 1 – Critérios de qualidade utilizados na seleção dos trabalhos na revisão sistemática.

| Questões de Avaliação | Pontuação (0–2) |
|--|------------------------|
| A validação dos resultados é realizada por um método apropriado? | 0–2 |
| A técnica apresentada obteve resultados satisfatórios? | 0–2 |
| A técnica utilizada pode se aplicar ao problema de pesquisa do capim-annoni? | 0–2 |
| O artigo aborda o problema do capim-annoni? | 0–2 |
| Este trabalho define o problema que pretende resolver? | 0–2 |
| Total máximo possível | 10 |
| Pontuação mínima para inclusão | 5 |

Figura 3 – Fluxo da Revisão Bibliográfica.



Fonte: Próprio autor (2024).

O estudo de Akbarzadeh, Ahderom e Alameh (2019) utiliza a análise da reflectância espectral para a diferenciação entre plantas na agricultura de precisão. Em função das exigências para a detecção de imagens de espectro visível para a identificação em tempo real, os veículos precisam ser usados em velocidades baixas para a correta detecção. A proposta foi de usar a reflectância espectral, mediante a coleta dos dados e a aplicação de Redes Neurais Convolucionais (RNC) para a identificação das ervas daninhas e/ou culturas em tempo real usando uma Unidade de Processamento Gráfico (*Graphics Processing Unit* - GPU) convencional de baixo peso. Os dados para detecção espectral foram obtidos através de uma Unidade de Discriminação de Plantas (*Plant Discrimination Unit* – PDU) desenvolvida pelo *Electron Science Research Institute* (ESRI) da Austrália, onde os dados foram coletados. A análise foi feita sobre canola e rabanete, pois o

rabanete selvagem é uma erva daninha dominante encontrada na maioria das fazendas de grãos da Austrália ocidental. Os resultados obtidos pelo trabalho confirmaram a melhora na velocidade de identificação das plantas em tempo real com os dados de reflectância espectral se comparados a análise com imagens de câmeras *Red Green Blue* (RGB) (aproximadamente 10 vezes mais rápidos). Segundo os autores, isso se deve ao tamanho dos dados coletados que é menor em relação ao tamanho dos dados coletados por câmera RGB. Embora a resolução dos dados coletados pela PDU seja muito inferior se comparado a imagem da câmera, os resultados apresentados na detecção foram muito próximos, fato que se deve às informações adicionais obtidas, pois a análise espectral coleta dados além do espectro visível até o infravermelho. Como a clorofila das plantas analisadas possui níveis de absorção distintos, são obtidos dados adicionais que facilitam a classificação com a PDU.

A pesquisa realizada por Chisholm e Marthick (2008) ocorreu em zonas úmidas altamente infestadas por ervas daninhas na Austrália, sendo baseada em um inventário de mapeamento para o projeto de Zonas Úmidas de *Illawarra*. Coletas de dados foram realizadas em novembro e março, utilizando um espectrorradiômetro ASD-FR para padronizar os espectros de reflectância com *Spectralon*. As amostras representavam espectros com reflectância entre 350 e 2500 nm em intervalos de amostragem de 2 nm, medidos tanto em laboratório quanto em campo. A suavização de dados espectrais e as análises foram conduzidas com o *Spectral Analyser*, através de *scripts* customizados no SPlus, que é um software de estatística e análise de dados amplamente utilizado para análise de dados complexos. A comparação entre as duas espécies, arbusto bitou (*Chrysanthemoides monilifera*) e lantana (*Lantana camara* L), revelou diferenças espectrais distintas na escala foliar, com maior reflectância *Visible* (Visível – VIS) e *Near Infrared* (Infravermelho Próximo – NIR) para o arbusto bitou, mas menor reflectância no *Short-Wave Infrared* (Infravermelho de Ondas Curtas – SWIR) em comparação com a lantana. O estudo concluiu sugerindo análises adicionais para observar a reflectância espectral sazonal, visando avaliar possíveis impactos da floração.

O estudo de Holmes *et al.* (2019) abordou a utilização de reflectância espectral de infravermelho para identificação de ervas daninhas em pastagens na Nova Zelândia. Os objetivos do estudo incluíam caracterizar a reflectância espectral proximal do infravermelho de três espécies comuns de ervas daninhas que crescem na área do experimento, em três diferentes tipos de solo, e determinar a extensão e adequação de diferentes métodos de análise estatística quando aplicados às medições de reflectância

espectral no infravermelho próximo para identificar espécies vegetais específicas. O artigo introduz a inclusão da composição do solo como um parâmetro físico-químico adicional. A razão para tal adição é que o solo é um fator primário na absorção de nutrientes e, portanto, na composição química da planta. O estudo relatado foi realizado no condado de Franklin, *South Auckland*, Nova Zelândia. As medições espectrais foram obtidas usando uma câmera *hiperespectral Pika NIR* com faixa de 900 nm a 1700 nm. As medições foram feitas aproximadamente ao do meio-dia solar para reduzir distorções que poderiam ocorrer por sombreamentos. Os espectros médios da referência branca de *Politetrafluoroetileno* (PTFE) foram usados para converter os dados da imagem de radiância média em refletância média, atendendo assim às variações na saída espectral de luz natural. O estudo concluiu não ser possível usar um pequeno número de canais de comprimento de onda para a identificação devido a *Signal-to-Noise Ratio* (Relação Sinal-Ruído – SNR). Foi utilizada a análise discriminante reforçada, o que permite uma discriminação viável entre diferentes espécies de plantas sob condições de SNR muito baixas. A conclusão foi de que as imagens hiperespectrais de infravermelho próximo com a análise discriminante reforçada têm o potencial para ser usada para a detecção de ervas daninhas típicas das pastagens na Nova Zelândia.

O trabalho conduzido por Amziane *et al.* (2021) investigou o efeito da luminosidade em imagens multiespectrais capturadas para a detecção de ervas daninhas em ambientes externos. O experimento foi realizado em uma estufa com luz natural (*greenhouse under skylight*) durante os meses de maio e junho de 2019, em diferentes horários do dia. O estudo utilizou uma câmera multiespectral Snapscan, que captura imagens em comprimentos de onda entre 475 nm e 902 nm. A Snapscan adquire uma sequência de imagens para compor uma imagem multiespectral, e o estudo explorou a correlação espectral, ou seja, a sobreposição significativa entre canais adjacentes. Os autores destacam que essa correlação é influenciada por variações na luminosidade do ambiente. Os resultados confirmaram que mudanças na luminosidade afetam os níveis de radiância capturados pelas imagens. Além disso, o modelo desenvolvido pelos autores pode ser ajustado para analisar imagens obtidas por outras câmeras, como a HySpex VNIR-1800 ou a V-EOS Bragg-gratin. A pesquisa sugere que a precisão e a robustez dos sistemas de reconhecimento de ervas daninhas em relação às variações de iluminação podem ser aprimoradas utilizando características de refletância. Embora o local do experimento esteja descrito, o artigo não detalha a abordagem metodológica utilizada para a classificação das plantas, ao contrário de outros trabalhos analisados, o

que limita a comparação direta com estudos que aplicam técnicas específicas de detecção ou aprendizado de máquina.

O trabalho de Shen, Tang e Li (2021) foi realizado no condado de *Knox*, no *Tennessee*, com o objetivo de usar análise espectral para identificar áreas infestadas por *Pueraria montana* (Kudzu), uma espécie invasora que se espalha rapidamente no sudeste dos Estados Unidos. As condições de precipitação na região estão dentro do intervalo adequado para sustentar um crescimento saudável da vegetação em muitas áreas. As imagens utilizadas no estudo foram obtidas pelo satélite Sentinel-2, que possui *Multispectral Instrument Sensors* (Sensores de Instrumentos Multiespectrais – MSI) a bordo. Esses dados fornecem informações espectrais detalhadas nas faixas VIS, NIR e SWIR, que são essenciais para a detecção e mapeamento precisos de plantas invasoras. O Sentinel-2 também oferece bandas estreitas da borda vermelha e comprimentos de onda do infravermelho, que são sensíveis à vegetação.

Para o estudo, foram utilizados produtos de Nível 2A, que são imagens de reflectância corrigidas atmosféricamente e geometricamente. Esses dados foram acessados em 12 de janeiro de 2021. O método utilizado na pesquisa foi baseado na fenologia e na separação espectral, seguindo três etapas sequenciais. Embora o Kudzu apresente características espectrais semelhantes às de árvores e pastagens próximas, ele pode ser diferenciado utilizando a banda NIR e as bandas da borda vermelha. Essa diferenciação é mais evidente no outono e início do inverno. Com base nisso, o *Normalized Difference Vegetation Index* (Índice de Vegetação da Diferença Normalizada – NDVI) foi utilizado, pois era esperado que o verde do Kudzu fosse mais intenso no final do outono em comparação com a primavera (SHEN; TANG; LI, 2021).

A classificação foi agrupada em duas categorias: kudzu e não-kudzu (incluindo pastagens, árvores e outros tipos de vegetação). A abordagem de classificação baseada em fenologia obteve resultados um pouco melhores, com mais de 50% das previsões corretas, produzindo mapas mais precisos. Segundo os autores, o método proposto pode ser aprimorado através de uma melhor definição dos parâmetros de separação espectral e da coleta de características fenológicas. Além disso, a combinação ideal de bandas espectrais para a separação pode ajudar a reduzir a redundância de informações e melhorar a precisão e eficácia da análise espectral. A conclusão do estudo foi que a análise espectral de separação é uma abordagem eficaz para identificar plantas invasoras, alcançando índices de *Jaccard* e *Kappa* de 0,858, 0,907, 0,789 e 0,725, respectivamente, para os mapas de presença de Kudzu no condado de *Knox* (SHEN; TANG; LI, 2021).

O estudo realizado por Mallmann *et al.* (2023) aborda a espectroscopia em nível de folhas para distinguir espécies invasoras das arbóreas nativas da mata atlântica no sul do Brasil. O estudo foi realizado no Parque Estadual Quarta Colônia que protege uma área de 1.847 ha de floresta secundária no sul do Brasil dentro do domínio da Mata Atlântica. As espécies invasoras, alvo de identificação neste estudo, são passa japonesa (*Hovenia dulcis Thunb.*) e Goiaba (*Psidium guajava L.*), e duas espécies nativas co-ocorrentes, nomeadamente chicole-cavalo (*Luehea divaricata Mart. & Zucc.*) e Goiaba Morango (*Psidium Cattleianum Sabine*). A análise se centrou em 12 características funcionais das espécies nativas e invasoras. As concentrações de clorofila a, clorofila b e carotenóides foram quantificadas utilizando um espectrofotômetro (Espectrofotômetro VM5, Celm E-205D, Bel Engineering, *Monza, Itália*) com base em sua absorbância em 663, 645 e 470 nm. A espectroscopia de campo foi realizada utilizando um espectrorradiômetro portátil (ASD FieldSpec®, Malvern Panalytical, *Malvern, Reino Unido*) na faixa de 300–1200 nm, com resolução espectral de 3 nm. Porém, a faixa espectral de 400–900 nm foi delimitada para este trabalho devido ao ruído observado no equipamento. O estudo foi realizado em 3 etapas: análise das características funcionais das árvores nativas e invasoras utilizando características foliares medidas em campo; análise de refletância espectral; e análise integrada de características funcionais foliares e refletância espectral. A Análise Discriminante Linear (*Linear Discriminant Analysis - LDA*) foi realizada para distinguir os espectros em dois subconjuntos: treino e teste. O estudo concluiu que a LDA pode ser usada para discriminar com precisão *Psidium guajava* e *Hovenia dulcis* de *Psidium Cattleianum* e *Luehea divaricata* e que a maior separação entre as espécies ocorre na região do Visível, mais especificamente nas regiões borda do vermelho (705nm) e verde (553nm).

O estudo conduzido por Iqbal *et al.* (2021) buscava identificar as assinaturas espectrais de espécies invasoras e nativas em duas áreas protegidas no Paquistão. A área de estudo abrange duas florestas arbustivas do distrito de *Jhelum*, na província de *Punjab*, Paquistão. A Floresta da Reserva *Lehri* cobre uma área de 4.843,3 ha, enquanto a Floresta da Reserva *Jindi* se estende por 2.163,1 ha. Os espectros de refletância superficial das folhas foram coletados usando um *ASD FieldSpec* portátil. Este espectrorradiômetro mede a radiação eletromagnética na faixa de 325nm a 1075nm. Foram feitas várias leituras cerca de 2 a 5 polegadas acima da superfície das folhas. Cada espectro folhar foi amostrado 3 vezes. As medições foram feitas entre as 9h e as 15h em condições com sol entre outubro de 2018 e outubro de 2019. As estatísticas descritivas foram calculadas

depois da extração dos dados com o software *View Spec Pro* da ASD. O software estatístico R foi usado para identificar as melhores combinações de banda para separar as espécies nativas das invasivas. O índice de água (*Water Index* - WI) foi o mais efetivo para separar entre as espécies invasoras e as nativas. *L. leucocephala* foi corretamente separada de 10 espécies de plantas e *J. adhatoda* foi separada de 8 espécies de plantas. O estudo constatou que o sensoriamento remoto hiperespectral é eficaz em monitorar e discriminar invasões de plantas em uma variedade de habitats, com base em suas propriedades espectrais. O objetivo do estudo era verificar o potencial da espectroscopia de campo para a separação das espécies nativas e invasoras. Diferentes bandas e comprimentos de onda foram identificados usando seus dados de reflectância pelo potencial de diferenciar as espécies das plantas. Os melhores comprimentos de onda para identificar *L. leucocephala* foram de 900nm e 970nm, com uma separabilidade espectral chegando a 100%. *P. hysterophorus* foi melhor identificada com ondas de 670nm e 830nm. As descobertas do estudo foram consideradas promissoras e confirmaram o potencial da aplicação das técnicas estatísticas junto a dados hiperespectrais para identificação de espécies invasoras na área alvo.

O trabalho de Neves *et al.* (2019) abordou a utilização de sensor multiespectral para mapeamento de áreas infestadas por capim-annoni. O experimento foi conduzido em uma área de 31 ha, divididos em quatro poteiros (divisões), e estes divididos em 3 zonas, com pastos nativos e diferentes níveis de infestação. O monitoramento era feito mensalmente e eram coletadas amostras das 12 zonas. A amostra era então pesada e separada nos seguintes componentes: fração verde das espécies nativas, fração verde de capim-annoni e material morto. Esta imagem foi gerada pelo mosaico de fotografias aéreas, obtidas pelo sensor multiespectral RedEdge-M® da empresa *MicaSense*. Este sensor foi montado em um *Remotely Piloted Aircraft* (RPA) de asa fixa. A imagem multiespectral foi cedida pela empresa Agronatura e o sensor possuía 5 bandas espectrais (475nm, 560nm, 668nm, 717nm e 840nm). Foram realizadas várias análises estatísticas sobre os dados obtidos. O modelo ajustado apresentou alta significância ($p < 0.05$), com apenas uma variável explicativa, a banda 4 (717nm). Neste trabalho, foi possível estimar a infestação de capim-annoni somente com esta banda espectral.

O trabalho conduzido por Mouta *et al.* (2023) visava a identificação de espécies invasoras ao longo do rio Cávado, no noroeste de Portugal. Na etapa 1, os dados foram coletados com um Veículo Aéreo Não Tripulado (VANT), sincronizados com as imagens do *Sentinel-2*. Na etapa 2, as séries temporais do *Sentinel-2* foram processadas com

o algoritmo DSen2 para padronizar a resolução espectral. A espécie alvo do estudo, o aguapé (*E. crassipes*), é uma planta aquática invasora listada no Regulamento da União Europeia de 22 de outubro de 2014. Foi utilizado o pacote *biomod2* (versão 3.5.1) implementado em R. Os resultados sugerem menor dispersão do aguapé durante o inverno, crescimento rápido na primavera com pico no verão e diminuição lenta e progressiva no outono. Os dados multiespectrais do *Sentinel-2* demonstraram resolução espacial e temporal suficientes para detectar e mapear a distribuição do aguapé em ecossistemas dinâmicos de água doce. A expansão das espécies aquáticas de plantas exóticas invasoras (*Aquatic Invasive Alien Plant Species - AIAPS*) foi monitorada de forma eficiente, evidenciando que a utilização de informações espectrais sensíveis a um AIAPS altamente dinâmico, como o aguapé, permite processos de classificação semiautomáticos com resultados temporal e espacialmente precisos, especialmente quando aplicados em modelos de fusão de dados.

Os trabalhos relacionados nesta seção forneceram conhecimentos necessários para a realização deste projeto, e demonstraram que a análise de dados espectrais aliada ao aprendizado de máquina produzem resultados positivos para a diferenciação entre plantas invasoras ou ervas daninhas de plantas nativas ou culturas. Os resultados confirmam que o uso destas técnicas pode permitir a detecção das plantas invasoras permitindo as ações necessárias a conter seus avanços sobre as plantas nativas. A seguir é apresentada a Tabela 2 que resume os estudos selecionados que exploram análise espectral para identificação de plantas:

Tabela 2 – Trabalhos correlatos identificados na revisão bibliográfica.

| Trabalho | Equipamento | Bandas | Análise | Acurácia |
|-------------------------------------|-------------------------------|----------------|----------------------------------|----------------------------------|
| Akbarzadeh, Ahderom e Alameh (2019) | PDU | Não Disponível | RNC inception | 85,00% |
| Amziane <i>et al.</i> (2021) | Snapscan | 475nm - 901nm | LGBM/QDA | 86,00% |
| Chisholm e Marthick (2008) | PDA | 350nm - 2500nm | SPlus/Spectral Analyser | Não Disponível |
| Holmes <i>et al.</i> (2019) | Pika NIR hyperspectral camera | 900nm - 1700nm | Análise Discriminante / Anova | Não Disponível |
| Iqbal <i>et al.</i> (2021) | ASD FieldSpec handheld2 | 325nm - 1075nm | JM Distance Analysis / Anova | Não Disponível |
| Mallmann <i>et al.</i> (2023) | ASD FieldSpec | 300nm - 1200nm | LDA | 95,00% |
| Mouta <i>et al.</i> (2023) | Sentinel-2 satellite sensor | 443nm - 2190nm | RF, GLM, FDA | Não Disponível |
| Neves <i>et al.</i> (2019) | multispectral RedEdge-M | 475nm - 840nm | NDVI / NDRE | R ² ajustado = 0,9341 |
| Shen, Tang e Li (2021) | Sentinel-2 satellite | 443nm - 2190nm | spectral unmixing / phenological | 90,00% |

Fonte: Próprio autor (2024).

3 METODOLOGIA

Este capítulo apresenta os passos envolvidos na realização desta pesquisa, descrevendo a caracterização e delimitação do problema, o processo de revisão sistemática da literatura, a coleta e análise dos dados, riscos e dificuldades possíveis durante a investigação, os resultados esperados e, finalmente, o processo de desenvolvimento da solução de *software*.

3.1 Caracterização da Pesquisa

De acordo com os critérios propostos por Gil (2008), esta pesquisa é classificada como:

- Quanto à natureza, caracteriza-se como aplicada, pois busca a geração de conhecimento voltado à resolução de um problema real, no caso a identificação do capim-annoni em áreas de pastagem.
- Quanto aos objetivos, enquadra-se como explicativa, pois procura analisar, compreender e explicar os fatores subjacentes ao fenômeno estudado, fornecendo base para investigações futuras.
- Quanto à abordagem, é quantitativa, uma vez que foram gerados dados numéricos analisados por métodos estatísticos e computacionais, possibilitando a extração de informações que conduzem a conclusões sobre o problema.
- Quanto aos procedimentos técnicos, classifica-se como experimental, por empregar métodos destinados à identificação das frequências espectrais mais relevantes para a diferenciação do capim-annoni.

3.2 Delimitação do Problema

Na delimitação do problema, destaca-se a questão central da viabilidade da análise espectral para diferenciar o capim-annoni das plantas nativas nos campos de pastagem nativa do bioma Pampa, utilizando os dados específicos coletados na área de estudo. A coleta desses dados ocorreu exclusivamente no bioma Pampa, em campos de pastagem nativa, restringindo assim o escopo da investigação. Esta limitação geográfica fornece um contexto preciso para avaliar a eficácia da análise espectral na identificação desta invasora

em meio à vegetação nativa.

Adicionalmente, a delimitação do estudo foi estabelecida a partir da interação com a Embrapa Pecuária Sul, instituição de referência na pesquisa agropecuária da região. A investigação está alinhada a diferentes eixos temáticos da unidade de pesquisa, incluindo sistemas de produção, sustentabilidade, forrageiras, pecuária familiar e inovação tecnológica, o que reforça a relevância científica e prática deste trabalho.

3.3 Delineamento Experimental

Os dados utilizados no experimento foram obtidos sem a adoção de um delineamento experimental padronizado. O processo de amostragem teve como objetivo principal abranger a maior diversidade possível de espécies vegetais presentes no ambiente investigado, priorizando a representatividade florística em detrimento de critérios estritos de distribuição espacial ou temporal.

3.3.1 Coleta de Dados Espectrais

Os dados espectrais foram coletados com o espectrorradiômetro ASD FieldSpec® 4 Standard-Res da empresa Malvern Panalytical. Este equipamento permite a coleta uniforme dos dados entre 350 nm e 2500 nm, abrangendo os espectros Visível (VIS), Infravermelho Próximo (NIR) e Infravermelho de Ondas Curtas (SWIR), com intervalos de 1 nm, resultando em 2151 bandas espectrais por amostra. A coleta das imagens espectrais foi realizada em abril de 2016 pela Embrapa Pecuária Sul (Figura 4), contemplando amostras com campo nativo e várias espécies de plantas, além de outras com alta presença de capim-annoni. Foram coletadas 46 amostras das seguintes espécies:

- *Cynodon dactylon* (bermudinha)
- *Eragrostis plana* Nees (capim-annoni)
- *Eryngium horridum* (caraguatá)
- *Axonopus*
- *Trifolium repens* L. (trevo branco)
- *Cyperus difformis* (junquinho)
- *Lolium multiflorum* (azevém)

- *Baccharis trimera* (carqueja)
- *Paspalum palci*
- *Paspalum dilatatum*
- *Eupatorium buniifolium* (chirca)
- *Senecio brasiliensis* (maria mole)
- *Vúlpia*
- *Paspalum notatum*



Figura 4 – Localização da área de coleta de dados espectrais na Embrapa Pecuária Sul, Bagé – RS.

Os dados foram exportados em um arquivo de texto (.txt) com as informações de cada observação e os dados obtidos nos comprimentos de onda entre 350 nm e 2500 nm.

Figura 5 – Espectrorradiômetro ASD FieldSpec® 4 Standard-Res utilizado na coleta dos dados espectrais.



Fonte: Malvern Panalytical. ¹.

3.4 Riscos e Dificuldades

No decorrer deste estudo, algumas dificuldades e limitações do conjunto de dados se mostraram relevantes para a interpretação dos resultados. A primeira delas refere-se ao número reduzido de amostras disponíveis (42 no total), o que afeta diretamente a capacidade dos modelos de aprendizado de máquina de generalizar padrões espectrais. Além disso, houve um desequilíbrio entre as classes, já que apenas 5 amostras apresentavam alta concentração de capim-annoni, enquanto as demais 37 continham mistura de diferentes espécies. Esse desbalanceamento tende a prejudicar principalmente modelos sensíveis à proporção entre classes, aumentando a probabilidade de erros na classificação.

Outro fator relevante diz respeito à heterogeneidade das amostras. Como grande parte delas continha múltiplas espécies sem identificação individualizada, não foi possível atribuir uma assinatura espectral exclusiva a cada espécie presente. Essa limitação reduz a precisão do processo de distinção entre vegetação nativa e invasora, uma vez que a resposta espectral registrada representa um conjunto misto de plantas. Tal condição aumenta o risco de confusão espectral, especialmente em regiões onde diferentes espécies

¹<https://p3.aprimocdn.net/malvernpanalytical/e6b38d1e-9dfa-462c-bffd-ad9f010f0ae1/636990436883146007PN_Original%20file.jpg?quality=90&width=2000>

apresentam reflectâncias semelhantes.

Com relação à inclinação das touceiras, a literatura aponta que o ângulo das folhas e do dossel pode afetar a forma das curvas de reflectância de acordo com estudo de Farinha *et al.* (2023). Entretanto, no conjunto utilizado neste estudo, todas as coletas foram realizadas com o sensor posicionado perpendicularmente ao alvo (aproximadamente 90°), o que minimiza esse tipo de interferência. Não houve, porém, acesso às fotografias originais da campanha de campo que permitissem verificar visualmente a composição exata de cada amostra ou confirmar a identificação das espécies presentes. Embora registros fotográficos existam na instituição responsável pela coleta, eles não estavam disponíveis para consulta durante o desenvolvimento deste trabalho.

Assim, as principais limitações observadas dizem respeito à quantidade reduzida de amostras, ao desbalanceamento entre classes e à mistura de espécies dentro de uma mesma amostra — fatores que, em conjunto, aumentam a complexidade do problema e podem ter contribuído para a variabilidade nos desempenhos obtidos pelos modelos analisados.

3.4.1 Análise de Dados de Espectrometria

Após a coleta dos dados espectrais, procedeu-se à verificação da qualidade das medições, etapa que incluiu a remoção de ruídos e a padronização dos dados para garantir comparabilidade entre as amostras. A remoção de ruídos seguiu o critério estabelecido a partir da Figura 11 apresentada por Macedo (2024), cuja descrição identifica explicitamente as faixas espectrais com comportamento instável. Segundo o autor, as setas pontilhadas destacam absorções no VIS em torno de 475 nm e 680 nm, enquanto as setas pretas indicam bandas de absorção de água localizadas aproximadamente em 970 nm, 1450 nm e 1940 nm; adicionalmente, as setas cinzas assinalam absorções centralizadas em 1190 nm e 1750 nm. Essas regiões são tradicionalmente reconhecidas por apresentarem baixa razão sinal-ruído ou forte interferência hídrica, motivo pelo qual todas foram removidas do conjunto de dados. Nenhuma técnica adicional de suavização foi aplicada; trabalhou-se exclusivamente com as bandas remanescentes após o descarte dessas faixas instáveis, assegurando que as variações espectrais analisadas refletissem diferenças reais entre as espécies avaliadas.

Na sequência, realizou-se uma análise estatística preliminar com o objetivo de investigar se alguns comprimentos de onda apresentavam indícios de separabilidade

entre as três classes definidas neste estudo: **Nativo**, **Annoni** e **Misto**. Em um primeiro momento, aplicou-se o teste de Shapiro–Wilk para verificar a normalidade das distribuições espectrais em cada banda. Como os dados não atenderam ao pressuposto de normalidade, utilizou-se o teste não paramétrico de Kruskal–Wallis para comparar as medianas entre os grupos. A hipótese nula adotada foi a de que as distribuições espectrais das três classes são equivalentes, sendo avaliada ao nível de confiança de 95% (ou seja, com $\alpha = 0,05$). Ressalta-se que essa análise teve caráter exclusivamente exploratório, visando apenas verificar a existência de indícios preliminares de separabilidade entre os grupos. Os resultados dessa etapa são apresentados e discutidos no Capítulo 4.

Após a análise estatística, exploraram-se técnicas de aprendizado de máquina para melhorar a precisão na identificação das plantas. Utilizaram-se algoritmos supervisionados como SVM, RPart, *Random Forest* e RNA, devido à pequena quantidade de amostras que inviabilizava o uso das RNC, que são capazes de classificar as amostras com base em suas assinaturas espectrais. Foi utilizado no conjunto de dados completo o classificador baseado em ângulo espectral (SAM). Esses modelos foram treinados com os dados rotulados, permitindo a criação de um sistema, do qual se espera que possa distinguir as espécies invasoras das nativas. Essas abordagens combinadas visaram construir um modelo robusto para a identificação das espécies vegetais, com foco na diferenciação do capim-annoni e outras plantas, utilizando as assinaturas espectrais coletadas (IQBAL *et al.*, 2021).

3.5 Ambiente de Desenvolvimento

Os experimentos foram conduzidos utilizando o RStudio versão *desktop*, instalado em um notebook HP Elitebook 840 G5 equipado com um processador Intel Core i5-8350u, 32GB de memória RAM DDR4 e placa gráfica *Intel UHD Graphics 620*.

O RStudio é um *Integrated Development Environment* (Ambiente de Desenvolvimento Integrado – IDE) para a linguagem de programação R. Ele fornece uma interface gráfica e ferramentas poderosas para escrever códigos, executar análises estatísticas, criar visualizações de dados e desenvolver aplicações baseadas em R. A linguagem R possui vários pacotes para o desenvolvimento de atividades específicas.

3.6 Processo de desenvolvimento

O processamento dos dados espectrais foi realizado no *software* RStudio, em ambiente R (R Core Team, 2023). Após a etapa inicial de tratamento, que consistiu na remoção de ruídos e padronização das medidas espectrais, os dados foram organizados e preparados para análise.

Em seguida, os espectros foram divididos em subconjuntos de acordo com seus comprimentos de onda, sendo cada subconjunto exportado em formato `xlsx` por meio do *software* Microsoft Excel. Essa organização permitiu estruturar diferentes bases de dados para as etapas posteriores de modelagem e comparação.

Foram gerados três arquivos principais:

1. um arquivo com todos os 1.688 comprimentos de onda (com ruídos removidos conforme estudo de (MACEDO, 2024));
2. um arquivo contendo 30 comprimentos de onda selecionados a partir da importância das variáveis pelo algoritmo *Random Forest* aplicado sobre o conjunto de 1688 bandas;
3. um arquivo com 28 comprimentos de onda indicados previamente na literatura por Thenkabail, Lyon e Huete (2018b).

Esses arquivos serviram de base para os experimentos subsequentes com os diferentes algoritmos de classificação.

As bibliotecas utilizadas no ambiente R e suas respectivas funcionalidades estão apresentadas na Tabela 3.

No *software* RStudio, foi conduzido o teste de Shapiro–Wilk apresentado no estudo de Shapiro e Wilk (1965) foi utilizado para avaliar se os dados coletados neste estudo seguem uma distribuição normal. Os resultados obtidos, tanto para o arquivo com 1.688 comprimentos de onda quanto para os 28 comprimentos de onda destacados por Thenkabail, Lyon e Huete (2018b), indicaram que os dados não seguem uma distribuição normal, pois o valor de p foi menor que $0,05 (p < 0.05)$.

Diante da violação da suposição de normalidade, utilizou-se o teste de Kruskal–Wallis apresentado por Kruskal e Wallis (1952) para verificar a existência de diferenças estatisticamente significativas entre as categorias dos dados. Esse teste é apropriado para a comparação de mais de duas amostras independentes que não seguem distribuição normal.

Tabela 3 – Bibliotecas R Utilizadas nos Experimentos

| Biblioteca | Funcionalidade |
|---|--|
| Análise Estatística | |
| car | Análise de variância (ANOVA) e regressão |
| DescTools | Ferramentas estatísticas diversas |
| psych | Ferramentas para análise psicológica |
| rstatix | Testes estatísticos robustos |
| RVAideMemoire | Testes estatísticos multivariados |
| Aprendizado de Máquina e Modelagem | |
| caret | Ferramentas para machine learning |
| e1071 | Métodos de classificação e regressão |
| partykit | Construção de árvores de decisão e florestas |
| randomForest | Algoritmo de Random Forest |
| ROCR | Curvas ROC e medidas de performance |
| rpart | Árvores de regressão e classificação |
| tree | Modelos de árvores de decisão |
| Geração de Tabelas em LaTeX | |
| kableExtra | Criação de tabelas LaTeX aprimoradas |
| xtable | Conversão de objetos R para tabelas LaTeX |
| Manipulação e Transformação de Dados | |
| dplyr | Manipulação eficiente de dados |
| readxl | Leitura de arquivos Excel |
| tidyr | Transformação de dados entre formatos longos e largos |
| writexl | Exportação de dados para Excel |
| Técnicas de Balanceamento de Dados | |
| smotefamily | Implementação do método ADASYN para balanceamento de classes |
| Validação e Síntese de Dados | |
| groupdata | Divisão de dados para validação cruzada (LOOCV) |
| mixup | Técnica de aumento de dados baseada em interpolação |
| synthpop | Síntese de dados (geração de dados sintéticos) |

Fonte: Elaborado pelo autor (2025).

A etapa seguinte compreende a realização de diversos testes e criação de modelos para a classificação dos dados, objetivando a criação de um modelo que possa realizar a classificação com a maior acurácia em relação aos dados coletados. Dentre os algoritmos testados para a classificação dos dados estão o RPart e o *Random Forest*.

3.6.1 Justificativa dos Modelos Utilizados

Neste estudo, optou-se pela utilização do algoritmo RPart na tarefa de classificação das amostras espectrais. A escolha foi motivada principalmente pela

interpretabilidade do modelo gerado, uma vez que as árvores de decisão permitem a visualização clara das regras que guiaram a separação das classes. Essa característica é particularmente relevante em estudos ambientais e agrícolas, nos quais muitas vezes é necessário compreender os limiares e as relações quantitativas entre as variáveis que influenciam os processos analisados. Além disso, o RPart é reconhecido por sua eficiência computacional e por lidar bem com variáveis contínuas e dados ruidosos, como os espectros de reflectância. Conforme discutido por Hastie, Tibshirani e Friedman (2009), esse algoritmo é apropriado para problemas com múltiplas variáveis e apresenta desempenho robusto mesmo em conjuntos de dados pequenos. De forma complementar, Friedl e Brodley (1997) demonstraram que árvores de decisão aplicadas ao sensoriamento remoto oferecem desempenho competitivo em comparação com métodos tradicionais, destacando-se por sua capacidade de lidar com dados não paramétricos e pela facilidade de interpretação dos resultados. Assim, o uso do RPart neste estudo justifica-se pela necessidade de um modelo interpretável, que permita compreender como as diferentes bandas espectrais contribuem para a distinção das classes analisadas.

O algoritmo *Random Forest* foi empregado neste estudo devido à sua robustez e alta acurácia em tarefas de classificação com dados espectrais, conforme demonstrado por Moazzam *et al.* (2019). Essa técnica constrói múltiplas árvores de decisão a partir de subconjuntos aleatórios dos dados e das variáveis, combinando os resultados por meio de votação majoritária, o que reduz o risco de *overfitting* e melhora a generalização do modelo. O estudo de Ok, Akar e Güngör (2012) apontou que o *Random Forest* supera métodos tradicionais, como o de Máxima Verossimilhança e o SVM, em termos de precisão na classificação de culturas agrícolas utilizando imagens de sensoriamento remoto. Além disso, conforme destacado por Belgiu e Dragut (2016), sua capacidade de lidar com dados de alta dimensionalidade e fornecer medidas de importância das variáveis torna-o particularmente adequado para identificar os comprimentos de onda mais relevantes na distinção entre espécies vegetais. Dessa forma, sua adoção neste estudo foi motivada pela necessidade de um modelo robusto, capaz de explorar relações complexas entre as bandas espectrais e de reduzir o impacto de ruídos e correlações entre variáveis.

O modelo de SVM foi selecionado por sua eficácia em lidar com conjuntos de dados de alta dimensionalidade e tamanho reduzido, características comuns em estudos com dados espectrais. O SVM busca encontrar o hiperplano ótimo que separa as classes, maximizando a margem entre elas, entendida como a distância entre o hiperplano

e as amostras mais próximas de cada classe (os chamados vetores de suporte). A maximização dessa distância aumenta a capacidade de generalização do modelo, o que o torna adequado para problemas com variabilidade sutil entre as classes. Revisões recentes destacam que o SVM apresenta desempenho competitivo em aplicações de agricultura de precisão, especialmente quando comparado a outros modelos de aprendizado de máquina, conforme estudado por Abouzahir, Sadik e Sabir (2018). Além disso, de acordo com Pal (2008), sua capacidade de generalização e robustez frente a dados ruidosos o tornam uma escolha apropriada para a classificação de culturas agrícolas utilizando imagens de sensoriamento remoto. Assim, sua inclusão neste estudo visa avaliar o desempenho de um modelo amplamente consolidado na literatura para cenários de alta dimensionalidade e amostras limitadas.

As RNA foram exploradas neste trabalho devido à sua capacidade de modelar relações complexas e não lineares presentes nos dados espectrais. Embora arquiteturas mais profundas, como as Redes Neurais Convolucionais (RNC), tenham demonstrado eficácia na detecção de plantas daninhas, conforme estudo de Mohanty, Hughes e Salathé (2016), a limitação no tamanho do conjunto de dados neste estudo levou à escolha de RNA mais simples, que requerem menos dados para um treinamento eficaz. Além disso, estudos aplicados ao bioma Cerrado demonstraram que RNA podem prever com precisão variáveis ambientais, como a biomassa acima do solo, utilizando índices de vegetação derivados de imagens de satélite, conforme apresentado por Oliveira *et al.* (2024). Assim, sua adoção neste estudo justifica-se pela capacidade de capturar padrões não lineares nas respostas espectrais, complementando os demais modelos e permitindo avaliar o potencial de abordagens baseadas em aprendizado profundo com conjuntos de dados reduzidos.

Os modelos foram posteriormente avaliados com base em métricas como acurácia, precisão, sensibilidade, especificidade e medida-F1, utilizando validação cruzada do tipo LOOCV, conforme detalhado na próxima seção.

Cabe destacar que, durante a etapa de revisão bibliográfica e redação, utilizou-se o ChatGPT como ferramenta de apoio para a síntese de informações já disponíveis na literatura e para o aprimoramento da clareza e da coesão textual, sem substituir a interpretação científica nem a análise estatística realizadas pelo autor.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, são descritos os procedimentos adotados para o tratamento dos dados espectrais, a aplicação de métodos estatísticos e de aprendizado de máquina, bem como a análise e interpretação dos resultados obtidos. Inicialmente, é apresentado o processo de limpeza e organização dos dados, seguido pela aplicação de testes estatísticos para seleção de variáveis relevantes.

4.1 Tratamento dos Dados

Foram coletados inicialmente 2.151 comprimentos de onda ou bandas espectrais. Durante o processo de pré-processamento, os comprimentos de onda que apresentaram ruídos foram removidos conforme o estudo de Macedo (2024), resultando em 1.688 comprimentos de onda válidos.

A Figura 6 apresenta os espectros médios das amostras depois da remoção das bandas com ruído, destacando as regiões afetadas pela absorção atmosférica de vapor d'água (em torno de 1,4 μm e 1,9 μm), que foram excluídas do conjunto final de dados. Essas regiões correspondem às janelas espectrais em que a reflectância apresenta valores fora da faixa física esperada (0–1), conforme descrito por Thenkabail (2018).

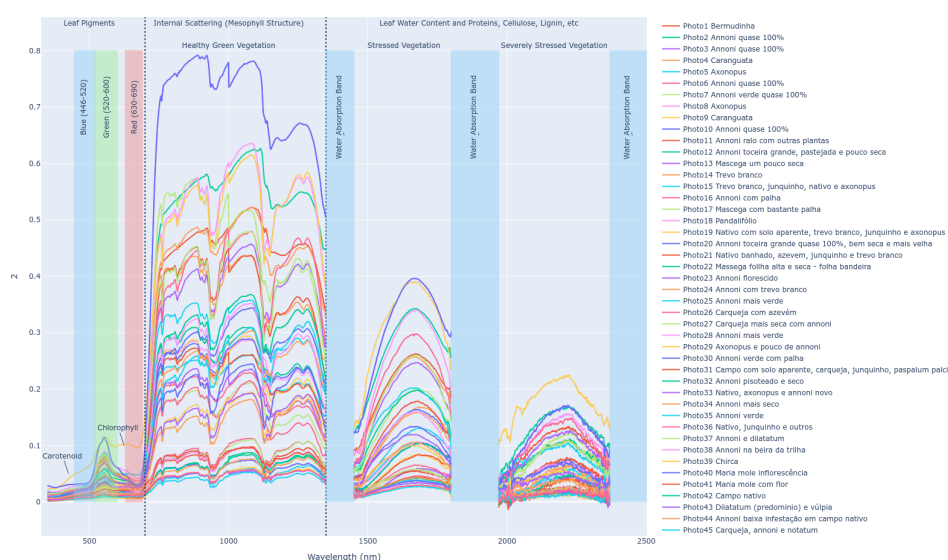


Figura 6 – Espectros representativos das amostras antes e após o pré-processamento. As áreas sombreadas indicam as regiões de forte absorção atmosférica pela umidade, removidas do conjunto de dados conforme Thenkabail (2018) e Macedo (2024).

Adicionalmente, das 46 amostras originais, 4 foram identificadas como valores discrepantes (*outliers*) a partir de análises exploratórias realizadas no RStudio, sendo então excluídas da base de dados. Após essa análise 42 amostras permaneceram para as análises subsequentes.

4.2 Análise Estatística

Com a base de dados ajustada, Realizou-se uma análise estatística preliminar para verificar se existiam comprimentos de onda que, isoladamente, apresentavam capacidade de distinguir as classes **Annoni**, **Nativo** e **Misto**.

Inicialmente, aplicou-se o teste de Shapiro–Wilk para avaliar a normalidade das distribuições espectrais em cada comprimento de onda. Os resultados indicaram violação do pressuposto de normalidade em todas as bandas analisadas, o que corrobora achados frequentes na literatura sobre espectroscopia, em que dados de reflectância tendem a apresentar comportamento não paramétrico.

Diante disso, utilizou-se o teste de Kruskal–Wallis como alternativa não paramétrica à ANOVA, avaliando a hipótese nula de que as distribuições espectrais dos três grupos são equivalentes, ao nível de confiança de 95%. O teste não rejeitou a hipótese nula em nenhuma das bandas avaliadas, indicando ausência de diferenças estatisticamente significativas entre as classes ao nível adotado.

Em razão da ausência de significância no teste global, não se procedeu à aplicação de testes pós-hoc, uma vez que tais comparações múltiplas somente são justificadas quando há rejeição da hipótese nula no teste principal. Assim, os resultados desta etapa reforçam que, do ponto de vista univariado, as bandas individuais não apresentam separabilidade estatística suficiente para distinguir as classes, motivando a adoção de métodos multivariados e algoritmos de aprendizado de máquina na etapa seguinte.

Os valores completos obtidos nos testes de Shapiro–Wilk e Kruskal–Wallis encontram-se disponíveis no Apêndice A, onde constam as estatísticas e valores de p para cada um dos 1.688 comprimentos de onda analisados.

4.3 Seleção de Atributos

Realizou-se uma análise estatística preliminar para verificar se existiam comprimentos de onda que, isoladamente, apresentavam capacidade de distinguir as classes **Annoni**, **Nativo** e **Misto**. Como não foram observadas diferenças estatisticamente significativas entre as classes, aplicou-se o algoritmo *Random Forest* para identificar os comprimentos de onda mais relevantes, resultando em um conjunto reduzido de 30 bandas espectrais selecionadas com base na importância das variáveis.

Foi considerado ainda, um terceiro conjunto de dados, baseado em 28 comprimentos de onda previamente sugeridos por Thenkabail, Lyon e Huete (2018b) em estudos aplicados à agricultura. Essa seleção baseia-se em comprimentos de onda indicados como relevantes em aplicações de sensoriamento remoto vegetal, permitindo uma comparação entre métodos baseados em aprendizado de máquina e conhecimento prévio da literatura especializada.

Foram considerados na modelagem os seguintes conjuntos de dados:

1. Conjunto completo: 1.688 comprimentos de onda e 42 amostras, após remoção de ruídos e *outliers*;
2. Conjunto reduzido via *Random Forest*: 30 comprimentos de onda mais relevantes segundo a importância das variáveis no modelo;
3. Conjunto baseado na literatura: 28 comprimentos de onda sugeridos por Thenkabail, Lyon e Huete (2018b).



Os arquivos correspondentes a esses três conjuntos de dados encontram-se anexados ao PDF desta dissertação.

A comparação entre os resultados obtidos pelos conjuntos definidos permite avaliar a eficácia da seleção automática de atributos frente a métodos consagrados da literatura, além de contribuir para a identificação de bandas espectrais comuns entre os métodos. Essa análise comparativa será retomada na Seção 4.4.3, durante a interpretação dos resultados obtidos pelos modelos classificatórios.

4.4 Modelagem e Avaliação dos Classificadores

Nesta Seção são apresentados os algoritmos de classificação utilizados, os procedimentos de validação adotados e os resultados obtidos com os diferentes conjuntos de dados.

4.4.1 Algoritmos Classificatórios Utilizados

Foram utilizados quatro algoritmos de classificação: RPart, Random Forest (RF), SVM e Rede Neural Artificial (RNA). O RPart foi selecionado por sua alta interpretabilidade, aspecto amplamente destacado na literatura como um diferencial em aplicações de sensoriamento remoto e agricultura de precisão. O RF foi empregado em razão de sua robustez, capacidade de generalização e uso recorrente em estudos correlatos que envolvem dados espectrais com alta dimensionalidade. O SVM foi incluído devido ao seu desempenho consistente em conjuntos de dados complexos e com poucas amostras, enquanto as RNA foram exploradas como abordagem adicional baseada em aprendizado profundo.

O algoritmo RF foi aplicado inicialmente com o objetivo de realizar a seleção automática de variáveis espectrais, reduzindo o número de bandas a partir de sua importância relativa para a discriminação das classes. Essa escolha metodológica buscou identificar um subconjunto otimizado de comprimentos de onda que concentrasse as informações mais relevantes, evitando redundâncias e diminuindo o risco de sobreajuste nos modelos subsequentes. No entanto, neste trabalho, o número de trinta bandas foi definido de forma arbitrária, tendo em vista que o conjunto original continha 1.688 bandas. Optou-se por não utilizar vinte e oito bandas, como na literatura, a fim de evitar confusões quanto à origem dos conjuntos e manter uma distinção clara entre o subconjunto obtido neste estudo e o proposto por outros autores. Assim, o número de trinta bandas foi adotado como um contraponto adequado para a análise comparativa, permitindo avaliar a coerência entre os resultados obtidos e aqueles já descritos na literatura.

A adoção do conjunto de trinta bandas também permitiu investigar se uma redução controlada de variáveis poderia manter ou até melhorar o desempenho dos modelos. Embora fosse esperado que um subconjunto menor pudesse reduzir o desempenho, observou-se comportamento inverso, possivelmente em razão da eliminação de bandas

redundantes ou com contribuições negativas, isto é, que adicionam ruído ou sobreposição espectral. Esse resultado reforça o potencial do RF não apenas como classificador, mas também como ferramenta de seleção de variáveis relevantes, contribuindo para a interpretabilidade e a eficiência do processo de classificação.

Durante a aplicação do RPart, foram geradas imagens das árvores de decisão resultantes, as quais fornecem uma representação gráfica da lógica aprendida pelo modelo. Essas estruturas foram utilizadas posteriormente na análise dos resultados, auxiliando na interpretação dos padrões de separação espectral entre as classes.

4.4.2 Validação dos Modelos e Tratamento de Desbalanceamento

Inicialmente, tentou-se adotar a divisão clássica dos dados em conjuntos de treinamento e teste nas proporções de 70/30 e 60/40, conforme descrito na Seção 2.10.1, com o objetivo de simular um cenário de validação tradicional. No entanto, devido ao número reduzido de amostras disponíveis (42 após a exclusão de outliers), observou-se elevada variabilidade nos resultados, indicando instabilidade nos modelos e possível viés estatístico.

Diante disso, optou-se por estratégias de validação mais robustas e apropriadas para bases de dados com número reduzido de instâncias, detalhadas na Seção 2.10:

- Validação Cruzada Estratificada k-Fold ($k = 10$) (Seção 2.10.2): garante que a proporção entre as classes seja mantida em cada partição de treinamento e teste, contribuindo para maior equilíbrio e confiabilidade das métricas.
- Validação Leave-One-Out (LOOCV) (Seção 2.10.3): O LOOCV utiliza uma única amostra como teste e o restante como treino em cada iteração, maximizando o aproveitamento dos dados. Esse procedimento gera estimativas de erro com baixo viés, porém com maior variabilidade, pois cada iteração depende de uma única observação de teste.

Além dos métodos de validação, técnicas de *Oversampling* foram aplicadas para lidar com o desbalanceamento entre classes. As seguintes técnicas foram utilizadas:

- ADASYN: gera amostras sintéticas para as classes minoritárias com base na densidade local, concentrando-se nas regiões de difícil classificação (Seção 2.10.4.2).
- *Mixup*: cria novas amostras a partir de combinações convexas entre pares de

instâncias de classes diferentes, promovendo regularização e melhor generalização dos modelos (Seção 2.10.4.1).

Essas abordagens foram aplicadas aos três conjuntos de dados definidos anteriormente (Seção 4.3), garantindo um cenário comparativo justo para avaliação dos classificadores.

4.4.3 Resultados dos Modelos de Classificação

Nesta subseção, são apresentados os resultados de classificação obtidos com os conjuntos previamente definidos (Seção 4.3). Para os conjuntos reduzidos (28 e 30 bandas), foram testados quatro algoritmos de classificação sob três condições: sem *Oversampling*, com Mixup e com ADASYN. No conjunto completo (1688 bandas), foram aplicados RPart, RF, SVM e SAM e não foi utilizado oversampling.

4.4.3.1 Desempenho com o conjunto completo (1.688 bandas)

Observando os resultados apresentados na Tabela 4, nota-se que, apesar de alguns modelos apresentarem valores relativamente altos de acurácia como o SVM no cenário binário (0,81) e os modelos RPart e RF no cenário binário (ambos com 0,73), as demais métricas de desempenho revelam limitações importantes na capacidade preditiva dos classificadores. A precisão, a sensibilidade e a medida-F1, que são métricas mais sensíveis ao desbalanceamento das classes, permanecem em níveis baixos em todos os casos, não ultrapassando 0,60 mesmo nos melhores cenários, indicando que os modelos tendem a acertar as instâncias da classe majoritária (que possui mais amostras), inflando a acurácia de forma artificial, enquanto falham ao identificar corretamente as amostras da classe de interesse (amostras com capim-annoni), especialmente nos contextos com mais de duas classes. Para ilustrar de forma mais concreta essas limitações, a Tabela 5 apresenta a matriz de confusão do modelo SVM no cenário binário. No que se refere às análises realizadas com o método SAM (linhas finais da Tabela 4), observa-se que o SAM não foi capaz de classificar adequadamente as amostras, apresentando desempenho inferior em comparação com os modelos de aprendizado de máquina supervisionados. Os percentuais de Falsos Positivos (FP) e Falsos Negativos (FN) indicam a fragilidade dos resultados. No cenário binário, observou-se uma taxa elevada de FP (30%) e FN (10%), evidenciando tanto a tendência a gerar falsos positivos quanto a falha na identificação

Tabela 4 – Resultados dos modelos utilizando 1688 comprimentos de onda e análises com SAM

| Modelo | Classes | Acur. | Prec. | Sens. | F1 | FP | FN |
|---------------|----------------|--------------|--------------|--------------|-----------|-----------|-----------|
| RPart | 2 | 0.73 | 0.33 | 0.40 | 0.36 | 0.15 | 0.12 |
| RF | 2 | 0.73 | 0.33 | 0.40 | 0.36 | 0.15 | 0.12 |
| SVM | 2 | 0.81 | 0.50 | 0.60 | 0.55 | 0.12 | 0.08 |
| RPart | 3 | 0.42 | 0.33 | 0.29 | 0.31 | 0.24 | 0.29 |
| RF | 3 | 0.47 | 0.27 | 0.23 | 0.25 | 0.19 | 0.24 |
| SVM | 3 | 0.54 | 0.50 | 0.44 | 0.47 | 0.19 | 0.24 |
| SAM | 2 | 0.60 | 0.25 | 0.50 | 0.33 | 0.30 | 0.10 |
| SAM | 3 | 0.38 | 0.33 | 0.50 | 0.40 | 0.12 | 0.06 |

Fonte: Próprio autor (2025).

Acur. = Acurácia; Prec. = Precisão; Sens. = Sensibilidade; F1 = Medida-F1; FP = Falsos Positivos; FN = Falsos Negativos.

Tabela 5 – Matriz de confusão do modelo SVM no conjunto binário com 1688 bandas.

| | Predito Annoni | Predito Nativo |
|--------------------|-----------------------|-----------------------|
| Real Annoni | 3 | 2 |
| Real Nativo | 3 | 18 |

Fonte: Próprio autor (2025).

Kernel: radial; Oversampling: nenhum.

correta das amostras positivas. O cenário triclasse apresentou FP de 12% e FN de 6%.

Nesse contexto, para o cálculo de FP e FN, as classes Annoni e Misto foram consideradas como uma única categoria “não-Nativo”. Assim, FP correspondem às amostras de Nativo classificadas como Annoni ou Misto, e FN às amostras de Annoni/Misto classificadas como Nativo. Confusões entre Annoni e Misto não são contabilizadas nesse cálculo, o que explica a aparente melhora das métricas. As métricas de acurácia (0,60 e 0,38), precisão, sensibilidade e medida-F1 foram sistematicamente inferiores às obtidas pelos demais modelos, especialmente quando comparadas ao SVM no cenário binário (acurácia de 0,81 e F1 de 0,55). O desempenho relativamente superior do SVM nesse cenário pode ser explicado por características próprias do algoritmo, que é particularmente adequado para dados de alta dimensionalidade, como os espectros com 1.688 bandas utilizados neste estudo. O SVM busca maximizar a margem entre as classes no espaço das características, o que reduz fortemente o impacto de ruídos e de variáveis altamente correlacionadas — uma condição comum em dados espectrais, nos quais bandas adjacentes frequentemente carregam informações redundantes. Além disso, o uso de funções de kernel permite ao SVM modelar fronteiras de decisão não

lineares sem necessidade de parametrizar explicitamente a relação entre as variáveis, o que favorece a separação entre assinaturas espectrais sutis. Esses fatores explicam por que, mesmo apresentando uma medida-F1 moderada (0,55), o SVM atingiu acurácia superior aos demais métodos e mostrou maior estabilidade na distinção entre as classes. Em contraste, métodos baseados em distâncias, como o SAM, e métodos altamente sensíveis ao particionamento dos dados, como árvores de decisão individuais, tendem a sofrer mais com o ruído, redundância e colinearidade inerentes ao conjunto de bandas completo, enquanto o SVM mantém desempenho robusto nesse tipo de configuração.

O baixo desempenho do SAM pode ser explicado por características inerentes ao próprio método. Por ser um classificador baseado exclusivamente na comparação do ângulo entre vetores espectrais, o SAM não utiliza mecanismos de aprendizado capazes de modelar relações não lineares ou ponderar individualmente a contribuição das bandas. Assim, em conjuntos com alta dimensionalidade, como o de 1.688 bandas, pequenas flutuações de ruído, variações instrumentais, diferenças de iluminação ou alterações fisiológicas entre amostras tendem a afetar fortemente o ângulo espectral, reduzindo a separabilidade entre as classes. Além disso, o SAM assume que a forma geral da assinatura espectral é suficiente para distinguir os alvos, o que não se confirmou neste estudo, já que as assinaturas do capim-annoni, vegetação mista e nativa apresentam trechos com grande similaridade espectral. Em contraste, os modelos supervisionados conseguem explorar padrões sutis distribuídos ao longo do espectro, reduzindo o impacto do ruído e da alta correlação entre bandas. Esses fatores explicam por que o SAM, embora útil na exploração inicial ou na seleção de bandas, mostrou-se ineficaz como método de classificação neste conjunto específico de dados. Ressalta-se que não foi aplicado *Oversampling* ao conjunto completo quando avaliado pelo SAM. Essa decisão fundamenta-se nas propriedades do próprio método, amplamente descritas na literatura. O SAM é um classificador baseado exclusivamente na comparação angular entre vetores espectrais, sendo altamente sensível à forma e à proporcionalidade das assinaturas segundo estudo de Kruse *et al.* (1993). Técnicas de *Oversampling*, como o SMOTE e suas variantes, geram observações sintéticas por interpolação entre amostras reais. Embora adequadas para modelos supervisionados que aprendem fronteiras de decisão, essas técnicas podem modificar artificialmente a geometria espectral dos vetores, alterando a inclinação relativa das bandas e, conseqüentemente, o ângulo espectral calculado pelo SAM. Como o método depende diretamente da preservação da forma original das assinaturas, o uso de amostras sintéticas poderia induzir padrões inexistentes na vegetação

real, comprometendo a interpretação física dos resultados.

Adicionalmente, trabalhos que aplicam o SAM enfatizam seu uso em dados originais, sem manipulações que alterem a curvatura espectral, justamente porque o método é mais indicado para avaliar similaridade espectral bruta e não para compensar desequilíbrios amostrais como estudado por Jensen (2015), Thenkabail, Lyon e Huete (2018b). Portanto, optou-se por avaliar o SAM apenas no conjunto completo, sem técnicas de balanceamento, utilizando-o como referência inicial de similaridade espectral. Entretanto, os resultados obtidos demonstraram desempenho insuficiente, razão pela qual o SAM não foi empregado para derivar subconjuntos de bandas, sendo mantido apenas como análise comparativa.

4.4.3.2 Desempenho com o conjunto reduzido via *Random Forest* (30 bandas)

A Tabela 6 apresenta os resultados obtidos pelos quatro algoritmos de classificação (RPart, *Random Forest*, SVM e RNA) aplicados ao conjunto de dados com 30 comprimentos de onda selecionados com base na importância de variáveis do *Random Forest*. Os modelos foram avaliados em três cenários: sem técnica de *Oversampling*, com aplicação de Mixup, e com ADASYN. São reportadas as métricas de Falsos Positivos (FP), Falsos Negativos (FN), Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), acurácia, precisão, sensibilidade e medida-F1.

Observa-se que, em diversos casos, os modelos apresentaram elevada acurácia, mas com desempenho insatisfatório nas demais métricas. Essa discrepância evidencia um problema comum em conjuntos desbalanceados: a acurácia pode ser inflada pelas predições corretas da classe majoritária, mascarando a incapacidade do modelo em identificar a classe de interesse. Em casos extremos, como observado no SVM e na RNA sem aplicação de *Oversampling*, as métricas de precisão, sensibilidade e F1 foram nulas, indicando que nenhuma amostra da classe positiva foi corretamente identificada.

Valores nulos ou muito baixos nas métricas também foram registrados em outros cenários, mesmo com a aplicação de técnicas de *Oversampling*, evidenciando a dificuldade dos modelos em aprender padrões representativos da classe minoritária. Esses resultados indicam que o baixo desempenho está relacionado principalmente à quantidade limitada de amostras e ao desbalanceamento entre as classes, mais do que à seleção automática de bandas.

Entre os modelos avaliados, o RPart com ADASYN destacou-se com desempenho mais equilibrado, apresentando precisão de 0,667, sensibilidade de 0,800 e medida-F1 de

0,727, além de manter o número de Falsos Positivos igual a zero. Este último aspecto é particularmente relevante no contexto de aplicação, pois predições equivocadas que indicam capim-annoni onde não há implicam no uso indevido de herbicidas em espécies nativas ou cultivadas, o que deve ser evitado a todo custo.

Já o RF apresentou desempenho mediano, mesmo com as bandas que ele próprio selecionou, o que levanta questionamentos sobre a eficácia da seleção baseada apenas em importância estatística. As RNA e o SVM mostraram sensibilidade acentuada às técnicas de *Oversampling*, com oscilações significativas nas métricas.

Tabela 6 – Métricas de desempenho dos modelos de classificação do conjunto com 30 comprimentos de onda obtidos do Random Forest

| Modelo | Classes | Oversampling | Acur. | Prec. | Sens. | F1 | FP | FN |
|--------|---------|--------------|-------|-------|-------|-------|-------|-------|
| rpart | 2 | adasyn | 0.885 | 0.667 | 0.800 | 0.727 | 0.077 | 0.038 |
| rf | 2 | adasyn | 0.808 | 0.500 | 0.600 | 0.545 | 0.115 | 0.077 |
| svm | 2 | adasyn | 0.808 | 0.000 | 0.000 | 0.000 | 0.000 | 0.192 |
| rna | 2 | adasyn | 0.769 | 0.000 | 0.000 | 0.000 | 0.038 | 0.192 |
| rpart | 2 | mixup | 0.692 | 0.200 | 0.200 | 0.200 | 0.154 | 0.154 |
| rf | 2 | mixup | 0.731 | 0.250 | 0.200 | 0.222 | 0.115 | 0.154 |
| svm | 2 | mixup | 0.808 | 0.000 | 0.000 | 0.000 | 0.000 | 0.192 |
| rna | 2 | mixup | 0.808 | 0.500 | 0.200 | 0.286 | 0.038 | 0.154 |
| rpart | 2 | nenhum | 0.846 | 0.600 | 0.600 | 0.600 | 0.077 | 0.077 |
| rf | 2 | nenhum | 0.808 | 0.500 | 0.600 | 0.545 | 0.115 | 0.077 |
| svm | 2 | nenhum | 0.808 | 0.000 | 0.000 | 0.000 | 0.000 | 0.192 |
| rna | 2 | nenhum | 0.808 | 0.000 | 0.000 | 0.000 | 0.000 | 0.192 |
| rpart | 3 | adasyn | 0.595 | 0.550 | 0.630 | 0.565 | 0.143 | 0.119 |
| rf | 3 | adasyn | 0.571 | 0.472 | 0.466 | 0.468 | 0.119 | 0.143 |
| svm | 3 | adasyn | 0.190 | 0.107 | 0.127 | 0.116 | 0.310 | 0.405 |
| rna | 3 | adasyn | 0.524 | 0.434 | 0.430 | 0.430 | 0.143 | 0.190 |
| rpart | 3 | mixup | 0.429 | 0.321 | 0.320 | 0.320 | 0.238 | 0.214 |
| rf | 3 | mixup | 0.548 | 0.474 | 0.451 | 0.456 | 0.143 | 0.190 |
| svm | 3 | mixup | 0.476 | 0.304 | 0.347 | 0.324 | 0.167 | 0.286 |
| rna | 3 | mixup | 0.405 | 0.269 | 0.290 | 0.279 | 0.190 | 0.262 |
| rpart | 3 | nenhum | 0.571 | 0.523 | 0.609 | 0.539 | 0.143 | 0.143 |
| rf | 3 | nenhum | 0.595 | 0.480 | 0.482 | 0.481 | 0.095 | 0.119 |
| svm | 3 | nenhum | 0.190 | 0.107 | 0.127 | 0.116 | 0.310 | 0.405 |
| rna | 3 | nenhum | 0.595 | 0.489 | 0.482 | 0.481 | 0.095 | 0.190 |

Fonte: Elaborado pelo autor(2025). Valores arredondados a três casas decimais.

Acur. = Acurácia; Prec. = Precisão; Sens. = Sensibilidade; F1 = Medida-F1; FP = Falsos Positivos; FN = Falsos Negativos.

Em síntese, os resultados indicam que a seleção automática de bandas, embora útil para reduzir dimensionalidade, não conseguiu garantir uma melhora no desempenho dos classificadores. A capacidade dos modelos de identificar corretamente a classe de

interesse depende tanto da qualidade das variáveis quanto do equilíbrio entre as classes, exigindo uma abordagem cuidadosa na construção e avaliação dos modelos.

Tabela 7 – Matriz de confusão do modelo RPart no cenário binário com ADASYN no conjunto de 30 bandas (cenário binário).

| Classe Real / Predita | Annoni | Nativo |
|------------------------------|---------------|---------------|
| Annoni | 4 | 1 |
| Nativo | 2 | 19 |

Fonte: Próprio autor (2025).

Modelo RPart cenário:binário configurado com cp=0.001, minsplit=2 e minbucket=1.

4.4.3.3 Desempenho com o conjunto baseado na literatura (28 bandas)

A Tabela 8 apresenta os resultados obtidos pelos quatro algoritmos de classificação (RPart, RF, SVM e RNA) sob três condições distintas de balanceamento do conjunto de dados: sem *Oversampling*, com Mixup e com ADASYN. Para cada modelo, são reportadas as métricas de Falsos Positivos (FP), Falsos Negativos (FN), acurácia, precisão, sensibilidade e medida-F1, a fim de avaliar de forma abrangente o desempenho em um cenário com classes desbalanceadas; para ilustrar de forma mais concreta esses resultados, a Tabela 7 apresenta a matriz de confusão do modelo RPart no cenário binário utilizando o método ADASYN.

Tabela 8 – Resultados dos modelos de classificação sob diferentes técnicas de oversampling no conjunto com 28 bandas da literatura.

| Modelo | Oversampling | Acur. | Prec. | Sens. | F1 | FP | FN |
|---------------|---------------------|--------------|--------------|--------------|-----------|-----------|-----------|
| Rpart | Nenhum | 0.77 | 0.43 | 0.60 | 0.50 | 0.15 | 0.08 |
| RF | Nenhum | 0.69 | 0.00 | 0.00 | 0.00 | 0.12 | 0.19 |
| SVM | Nenhum | 0.81 | 0.00 | 0.00 | 0.00 | 0.00 | 0.19 |
| RNA | Nenhum | 0.81 | 0.50 | 0.60 | 0.55 | 0.12 | 0.08 |
| Rpart | Mixup | 0.77 | 0.43 | 0.60 | 0.50 | 0.15 | 0.08 |
| RF | Mixup | 0.77 | 0.40 | 0.40 | 0.40 | 0.12 | 0.12 |
| SVM | Mixup | 0.69 | 0.38 | 1.00 | 0.56 | 0.31 | 0.00 |
| RNA | Mixup | 0.73 | 0.38 | 0.60 | 0.46 | 0.19 | 0.08 |
| Rpart | ADASYN | 0.77 | 0.43 | 0.60 | 0.50 | 0.15 | 0.08 |
| RF | ADASYN | 0.77 | 0.40 | 0.40 | 0.40 | 0.12 | 0.12 |
| SVM | ADASYN | 0.73 | 0.42 | 1.00 | 0.59 | 0.27 | 0.00 |
| RNA | ADASYN | 0.73 | 0.38 | 0.60 | 0.46 | 0.19 | 0.08 |

Fonte: Próprio Autor (2025). Valores arredondados a três casas decimais.

Acur. = Acurácia; Prec. = Precisão; Sens. = Sensibilidade; F1 = Medida-F1; FP = Falsos Positivos; FN = Falsos Negativos.

Observa-se que, no cenário sem *Oversampling*, as RNA otimizadas para acurácia apresentaram o melhor desempenho geral, com acurácia de 0,81 e medida-F1 de 0,55, superando os demais modelos. No entanto, técnicas de *Oversampling* trouxeram ganhos importantes, principalmente no caso do SVM, cujo medida-F1 aumentou de 0,00 para 0,56 com Mixup e para 0,59 com ADASYN, indicando uma melhora significativa na capacidade do modelo em identificar corretamente a minoria.

O algoritmo RPart demonstrou consistência entre os cenários, com desempenho estável e interpretável, como discutido na revisão da literatura. Já o RF, apesar de amplamente citado em trabalhos correlatos, teve desempenho inferior no cenário sem balanceamento, evidenciando a importância do tratamento de desbalanceamento neste conjunto. A Tabela 9 apresenta as diferenças de desempenho observadas no RPart, refletindo a sensibilidade do modelo à função de custo adotada no treinamento, ou seja, à forma como os erros das previsões são quantificados e penalizados durante o ajuste dos pesos do modelo.

Tabela 9 – Matriz de confusão do modelo RPart no conjunto de 28 bandas (cenário binário).

| Predito / Real | Annoni | Nativo |
|----------------|--------|--------|
| Annoni | 3 | 2 |
| Nativo | 4 | 17 |

Fonte: Próprio autor (2025).

Configuração do modelo RPart: cp = 0.001, minsplit = 2, minbucket = 1; método de balanceamento: ADASYN.

Esses resultados reforçam a necessidade de estratégias de balanceamento de dados e mostram como diferentes algoritmos respondem a tais técnicas, justificando a escolha dos métodos aplicados nesta pesquisa.

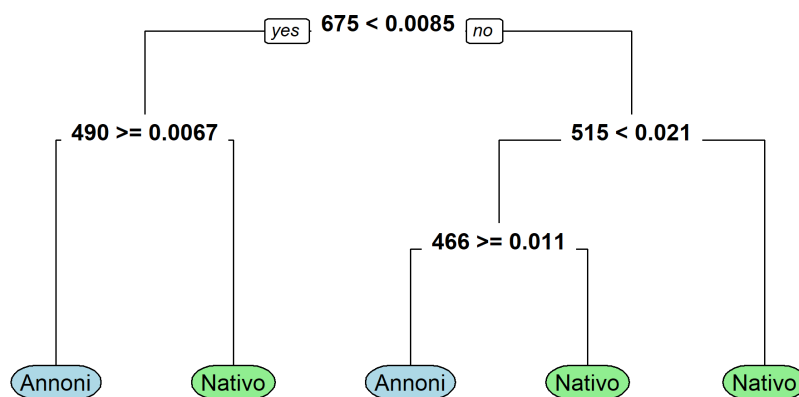
4.4.3.4 Análise das Árvores de Decisão

As árvores de decisão foram geradas a partir dos três conjuntos de dados espectrais definidos na Seção 4.3. Para cada conjunto, foram escolhidas duas árvores, correspondentes aos cenários binário e triclasse, totalizando seis árvores que foram escolhidas dentre as árvores geradas nas iterações do LOOCV que gerou 26 árvores para cada conjunto no cenário binário e 42 árvores para cada conjunto no cenário triclasse. A escolha ilustrar foram as árvores que acertaram a previsão. As folhas das árvores indicam

as classes (*Annoni*, *Misto* e *Nativo*), enquanto os ramos internos apresentam as bandas e os respectivos pontos de decisão.

No conjunto de 28 bandas da literatura, no cenário binário, foram empregadas as bandas 466, 490, 515 e 675 nm, todas no visível, associadas à absorção pela clorofila (Figura 7).

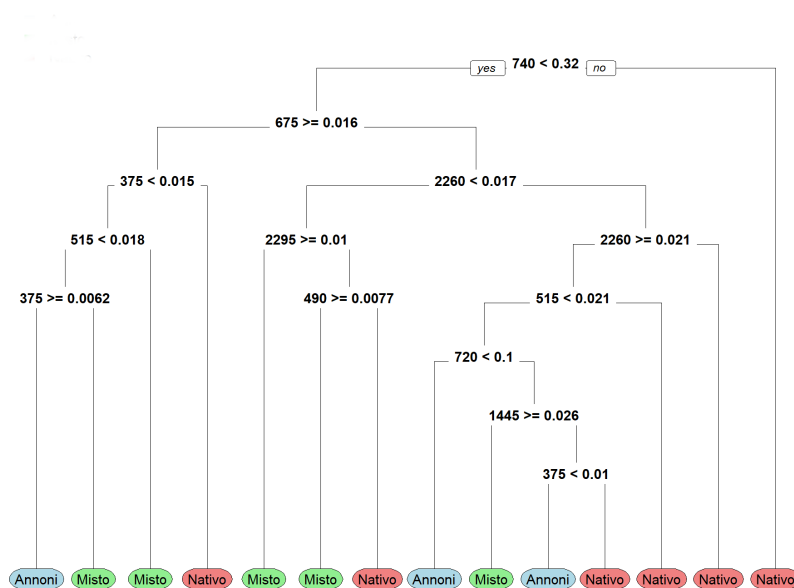
Figura 7 – Árvore de decisão gerada com o conjunto de 28 bandas da literatura, cenário binário.



Fonte: Elaborado pelo autor (2025).

No cenário triclasse, além das citadas anteriormente, a árvore utilizou também as bandas 375, 720, 740, 1445, 2260 e 2295 nm, abrangendo o SWIR, relacionadas à absorção de água e a compostos estruturais, como lignina e celulose (Figura 8) (THENKABAIL; LYON; HUETE, 2018b).

Figura 8 – Árvore de decisão gerada com o conjunto de 28 bandas da literatura, cenário triclasse.

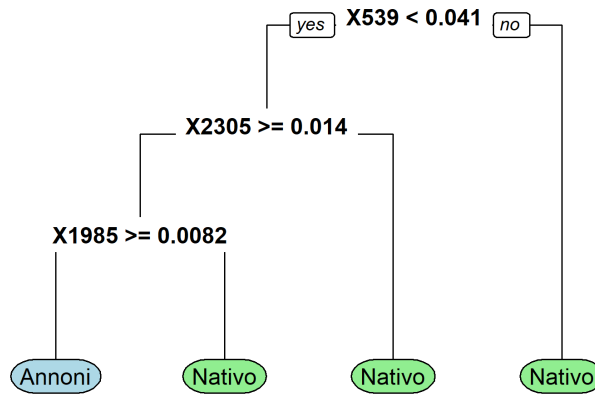


Fonte: Elaborado pelo autor (2025).

Nas árvores construídas com as 30 bandas selecionadas pelo *Random Forest*, o cenário binário envolveu os comprimentos de onda 539, 1985 e 2305 nm, distribuídos entre o VIS e o SWIR (Figura 9).

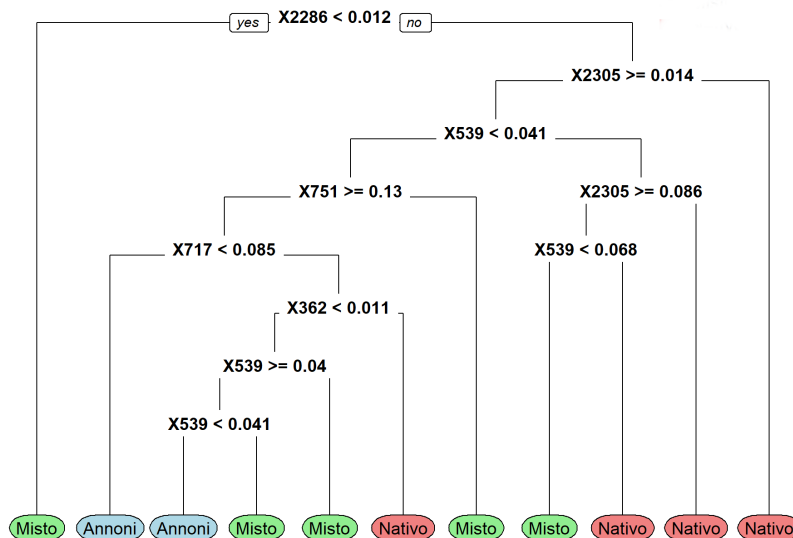
Já no cenário triclasse, as bandas decisórias incluíram 362, 539, 717, 751, 2286 e 2305 nm, abrangendo regiões do VIS, bordas do vermelho, NIR e SWIR, associadas à variação nos pigmentos fotossintéticos, ao teor de clorofila e ao conteúdo hídrico da vegetação (Figura 10) (THENKABAIL; LYON; HUETE, 2018c; THENKABAIL; LYON; HUETE, 2018a).

Figura 9 – Árvore de decisão gerada com as 30 bandas selecionadas pelo *Random Forest*, cenário binário.



Fonte: Elaborado pelo autor (2025).

Figura 10 – Árvore de decisão gerada com as 30 bandas selecionadas pelo *Random Forest*, cenário triclasse.

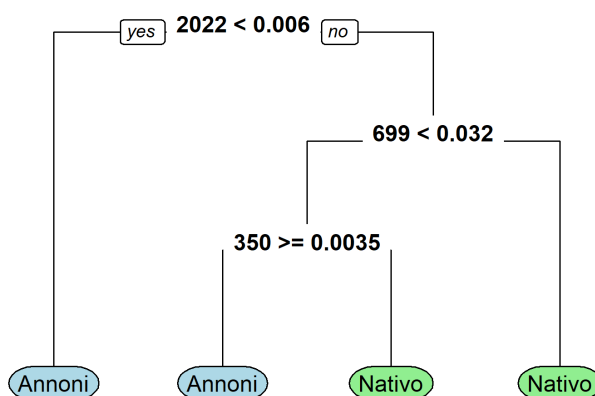


Fonte: Elaborado pelo autor (2025).

No conjunto completo de 1.688 bandas, o cenário binário envolveu os comprimentos de onda 350, 699 e 2022 nm, abrangendo o VIS, bordas do vermelho e

o SWIR (Figura 11).

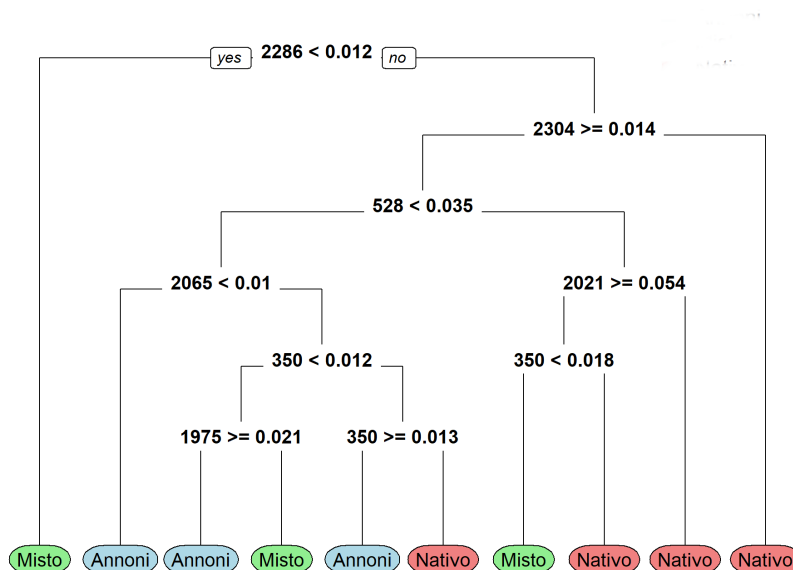
Figura 11 – Árvore de decisão gerada com o conjunto completo de 1688 bandas, cenário binário.



Fonte: Elaborado pelo autor (2025).

No cenário triclasse, a árvore utilizou as bandas 350, 528, 1975, 2065, 2286 e 2304 nm, concentradas principalmente no SWIR, além de pontos no visível (Figura 12). Essas regiões estão fortemente relacionadas à absorção por água e a compostos estruturais, demonstrando que a separação entre classes depende de regiões espectrais além do visível (THENKABAIL; LYON; HUETE, 2018b; THENKABAIL; LYON; HUETE, 2018a).

Figura 12 – Árvore de decisão gerada com o conjunto completo de 1688 bandas, cenário triclasse.



Fonte: Elaborado pelo autor (2025).

Em síntese, a classificação do capim-annoni depende de comprimentos de onda distribuídos em diferentes regiões do espectro. A recorrência de bandas no VIS e na borda do vermelho não apenas reflete a influência dos pigmentos fotossintéticos, mas também indica que essa espécie apresenta diferenças sutis na composição e organização da clorofila quando comparada à vegetação nativa. Essas diferenças tornam-se evidentes principalmente na região do *red edge*, onde pequenas variações na estrutura celular e no estado fisiológico das folhas alteram de forma significativa a inclinação da curva espectral. Isso sugere que o capim-annoni possui respostas espectrais que refletem seu comportamento ecológico agressivo, maior rusticidade e adaptação a ambientes empobrecidos, características já relatadas na literatura. Da mesma forma, a seleção consistente de bandas no SWIR indica que o conteúdo hídrico e a estrutura interna das folhas desempenham papel importante na separação das classes. Como o SWIR é sensível à absorção por água e à compactação das camadas foliares, sua presença recorrente entre as bandas mais relevantes sugere que o capim-annoni apresenta diferenças anatômicas e fisiológicas em relação ao campo nativo, como variações na espessura do mesófilo ou maior resiliência hídrica. Esse padrão corrobora o destacado por Thenkabail, Lyon e Huete (2018b), Thenkabail, Lyon e Huete (2018c), Thenkabail, Lyon e Huete (2018a), que apontam o SWIR como fundamental na distinção de espécies devido ao seu vínculo

com propriedades bioquímicas e estruturais. O fato de essas regiões espectrais se repetirem em todas as árvores reforça que os modelos capturaram características inerentes da espécie e não padrões artificiais ou instáveis. Assim, os resultados não apenas confirmam a relevância das faixas já reconhecidas na literatura, mas também evidenciam que a análise hiperespectral é capaz de distinguir o capim-annoni a partir de alterações fisiológicas e estruturais próprias da espécie, oferecendo uma base sólida para aplicações em monitoramento de pastagens e mapeamento de invasoras no Bioma Pampa.

4.4.3.5 Análise das abordagens de classificação

A análise das tabelas de desempenho dos modelos revela diferenças entre os algoritmos e conjuntos de bandas avaliados. Para o conjunto completo com 1688 bandas (Tabela 4), SAM apresentou o pior desempenho geral, tanto para duas quanto para três classes, evidenciado por menores valores de acurácia, sensibilidade e medida-F1. SVM apresentou o melhor desempenho, superando RPart e RF, que apresentaram desempenho intermediário.

No conjunto reduzido com 30 bandas selecionadas via Random Forest (Tabela 6), o uso de técnicas de *Oversampling* (ADASYN e Mixup) impactou de forma distinta os modelos. RPart e RF mostraram aumento na sensibilidade e na medida-F1 com ADASYN, especialmente para duas classes, enquanto SVM e RNA apresentaram desempenho limitado em algumas métricas, possivelmente devido à menor capacidade de generalização para classes minoritárias. Para três classes, a medida-F1 foi reduzida, evidenciando novamente a dificuldade de classificar o cenário intermediário.

No conjunto com 28 bandas da literatura (Tabela 8), o padrão de desempenho seguiu tendência semelhante: RPart manteve consistência moderada entre os diferentes métodos de *Oversampling*, RF apresentou baixo desempenho em cenários de duas classes sem *Oversampling*, e SVM e RNA mostraram variação na sensibilidade e medida-F1 dependendo da técnica aplicada.

De maneira geral, os resultados indicam que:

- Árvores de decisão (RPart e RF) apresentaram desempenho estável, mas sua sensibilidade à distribuição das classes decorre da forma como esses modelos particionam o espaço de decisão. Como as divisões são baseadas em limiares definidos a partir da frequência das classes, desbalanceamentos podem levar a partições enviesadas, privilegiando a classe majoritária. Ainda assim, o

Random Forest demonstrou maior robustez ao ruído e à alta dimensionalidade por combinar múltiplas árvores e reduzir variâncias, o que explica seu desempenho consistentemente superior ao do RPart.

- SVM e RNA apresentaram maior variabilidade, pois ambos dependem diretamente da representação matemática das fronteiras entre as classes. No SVM, a posição dos vetores de suporte é fortemente influenciada pelo número de amostras de cada classe e pela presença de ruído, tornando o modelo mais sensível a desbalanceamentos e à necessidade de *Oversampling*. Já a RNA, por aprender padrões não lineares a partir de um conjunto de treino reduzido, tende a sofrer tanto com sobreajuste quanto com instabilidade quando o número de amostras por classe é limitado, justificando o desempenho variável observado.
- SAM apresentou desempenho inferior porque seu mecanismo de classificação baseia-se exclusivamente na comparação angular entre vetores espectrais, sem explorar diferenças sutis distribuídas ao longo das bandas. Em conjuntos de alta dimensionalidade, como o de 1.688 bandas, pequenas variações instrumentais, ruído espectral e alta correlação entre bandas reduzem a capacidade do SAM de separar adequadamente as classes. Isso explica, em especial, a dificuldade do método em reduzir falsos negativos, pois assinaturas espectralmente semelhantes tendem a gerar ângulos pequenos, mesmo quando pertencem a classes distintas.

Essas observações reforçam que a escolha do modelo deve considerar não apenas a acurácia geral, mas também medidas complementares como precisão, sensibilidade, medida-F1, FP e FN, além da disponibilidade de bandas espectrais e estratégias de balanceamento de dados.

A relação entre os conjuntos de bandas reduzidos (28 bandas da literatura e 30 bandas do RF) revelou uma correspondência espectral relevante. Conforme apresentado na Tabela 10, sete bandas do conjunto de 28 apresentam forte correlação (distância espectral inferior a 10 nm) com vinte e duas bandas do conjunto de 30. Essa sobreposição ocorre principalmente nas regiões do verde (aproximadamente entre 520 e 550 nm), do vermelho e borda do vermelho (700–740 nm) e do infravermelho de ondas curtas (próximas a 2290 nm), faixas reconhecidas por sua sensibilidade à estrutura foliar, ao teor de clorofila e à presença de compostos orgânicos.

Essas características sugerem que ambos os conjuntos possivelmente retêm informações espectralmente relevantes para a discriminação do capim-annoni, o que pode explicar o desempenho semelhante observado entre os modelos treinados com 28

e 30 bandas. Essa hipótese, no entanto, deve ser confirmada em estudos futuros, com ampliação da base de dados e validação em diferentes condições de campo. Ainda assim, destaca-se que o conjunto de 30 bandas obtido via Random Forest corrobora as regiões espectrais apontadas na literatura, demonstrando a capacidade do método em identificar variáveis potencialmente representativas para a caracterização da espécie.

Tabela 10 – Bandas com forte correlação (distância até 10 nm) entre os conjuntos analisados.

| Conjunto | Bandas com forte correlação |
|---|---|
| 28 bandas | |
| 375, 466, 490, 515, 520, 525, 550, 575, 675, 682, 700, 720, 740, 845, 915, 975, 1100, 1215, 1245, 1316, 1445, 1518, 1725, 2035, 2173, 2260, 2295, 2359 | 515, 520, 525, 550, 720, 740, 2295 |
| 30 bandas | |
| 362, 517, 522, 528, 530, 531, 534, 535, 537, 539, 540, 542, 543, 546, 549, 553, 555, 556, 558, 559, 717, 730, 751, 760, 814, 986, 1985, 2286, 2304, 2305 | 517, 522, 528, 530, 531, 534, 535, 540, 542, 543, 546, 549, 553, 555, 556, 558, 559, 717, 730, 2286, 2304, 2305 |

Fonte: elaboração própria (2025). Nota: as bandas em negrito na coluna da esquerda indicam as bandas dos conjuntos de 28 e 30 bandas que apresentam forte correlação (até 10 nm) com as bandas da literatura. Essa relação evidencia sobreposição espectral entre as faixas analisadas.

5 CONCLUSÕES

O presente estudo avaliou o uso de classificadores supervisionados na discriminação de pastagens infestadas por capim-annoni, empregando dados espectrais em diferentes cenários. Foram testados três conjuntos de variáveis (28, 30 e 1.688 bandas espectrais), em cenários binário e triclasse, e aplicadas técnicas de balanceamento para reduzir o efeito do desbalanceamento entre as classes.

A análise das árvores de decisão mostrou que o desempenho variou conforme o conjunto de bandas e o cenário de classificação. Nos conjuntos de 28 e 30 bandas, em cenário binário, observaram-se valores mais consistentes para métricas associadas à identificação correta da classe alvo (capim-annoni), como sensibilidade, precisão e medida-F1, além de menores taxas de falsos negativos. No conjunto com 1.688 bandas, os resultados foram inferiores em comparação, com maior ocorrência de erros na detecção dessa classe.

No cenário triclasse, todos os conjuntos apresentaram redução no desempenho, com aumento do número de falsos positivos e falsos negativos. Esse resultado pode estar relacionado à maior heterogeneidade da classe “Misto”, formada por amostras de capim-annoni junto com espécies nativas, dificultando a separabilidade espectral.

A interpretação das árvores permitiu identificar bandas espectrais relevantes em diferentes regiões do espectro. Entre as mais frequentes destacaram-se comprimentos de onda no visível (466, 490, 515, 539, 615, 675, 717, 720 e 740 nm), no infravermelho próximo e no infravermelho de ondas curtas (1445, 1975, 1985, 2021, 2022, 2065, 2260, 2286, 2304 e 2305 nm). A recorrência dessas bandas em diferentes cenários sugere sua relevância na distinção do capim-annoni em relação às demais coberturas vegetais.

A recorrência de determinadas bandas espectrais em diferentes cenários sugere que essas faixas estão associadas a propriedades químicas e fisiológicas específicas da vegetação. Por exemplo, bandas no visível, como 466, 490, 515 e 539 nm, estão relacionadas à absorção de clorofila e à pigmentação foliar, refletindo a saúde e a densidade da biomassa do capim-annoni. No vermelho e borda do vermelho (675, 717, 720 e 740 nm), observa-se forte sensibilidade à clorofila e à estrutura interna das folhas, enquanto bandas no infravermelho próximo e de ondas curtas (1445, 1975, 1985, 2021, 2022, 2065, 2260, 2286, 2304 e 2305 nm) estão ligadas à absorção de água, conteúdo de celulose e lignina. Essa associação sugere que as bandas identificadas pelo Random Forest e presentes em ambos os conjuntos (28 e 30 bandas) capturam características

espectrais representativas do capim-annoni, fornecendo suporte físico-químico para a classificação obtida e indicando potenciais áreas de investigação futura sobre os mecanismos bioquímicos subjacentes às respostas espectrais observadas.

Um aspecto importante a destacar é que, no conjunto de 30 bandas selecionadas pelo classificador Random Forest, 22 delas (73%) apresentam distância menor que 10 nm em relação à sete das 28 bandas utilizadas em trabalhos de referência da literatura. Essa proximidade espectral indica alta correlação entre bandas adjacentes e reforça a consistência dos resultados obtidos, corroborando e confirmando a importância das regiões espectrais já identificadas por Thenkabail, Lyon e Huete (2018b).

Ao considerar a amplitude espectral empregada (1.688 bandas), este estudo amplia a contribuição em relação a trabalhos anteriores, como o de Neves *et al.* (2019), que utilizaram apenas cinco bandas derivadas de índices de vegetação. A análise comparativa evidencia que o presente estudo não apenas avalia conjuntos espectrais mais amplos, mas também permitiu identificar possíveis bandas relevantes de forma consistente, reforçando a robustez da detecção do capim-annoni em diferentes cenários.

Em síntese, verificou-se que o desempenho dos classificadores é influenciado tanto pela escolha do conjunto de variáveis quanto pela definição das classes. O uso de métricas além da acurácia mostrou-se fundamental para evidenciar limitações e potenciais na identificação do capim-annoni, fornecendo subsídios práticos para futuras aplicações em sensoriamento remoto no monitoramento e manejo dessa espécie invasora.

REFERÊNCIAS

- ABOUZAHIR, S.; SADIK, M.; SABIR, E. Enhanced approach for weeds species detection using machine vision. In: **2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)**. IEEE, 2018. p. 1–6. Disponível em: <https://doi.org/10.1109/ICECOCS.2018.8610505>.
- AKBARZADEH, S.; AHDEROM, S.; ALAMEH, K. Application of spectral reflectance for increasing plant discrimination speed in precision agriculture. In: **2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT, IoT and AI (HONET-ICT)**. IEEE, 2019. p. 140–142. Disponível em: <https://doi.org/10.1109/HONET.2019.8907994>.
- AMZIANE, A. *et al.* Reflectance estimation from multispectral linescan acquisitions under varying illumination—application to outdoor weed identification. **Sensors**, MDPI, v. 21, n. 11, p. 3601, maio 2021. ISSN 1424-8220. Disponível em: <https://doi.org/10.3390/s21113601>.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, Institute of Mathematical Statistics, v. 4, p. 40–79, janeiro 2010. ISSN 1935-7516. Disponível em: <https://doi.org/10.1214/09-SS054>.
- ASNER, G. P. Hyperspectral remote sensing of canopy chemistry, physiology, and biodiversity in tropical rainforests. In: _____. **Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests**. CRC Press, 2008. p. 261–296. ISBN 9781420053432. Disponível em: <https://doi.org/10.1201/9781420053432.ch12>.
- AZEEM, A. *et al.* Artificial neural networking to estimate the leaf area for invasive plant *Wedelia trilobata*. **Nordic Journal of Botany**, Wiley, v. 38, n. 6, p. e02768, junho 2020. ISSN 1756-1051. Disponível em: <https://doi.org/10.1111/njb.02768>.
- BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 114, p. 24–31, abril 2016. ISSN 0924-2716. Disponível em: <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- BHATTACHARYYA, A. *et al.* Spectral classification of vegetation using support vector machines: a review. **Remote Sensing Applications: Society and Environment**, Elsevier, v. 18, p. 100297, 2020.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer, 2006. v. 4. (Information Science and Statistics, 4). ISBN 978-0-387-31073-2. Disponível em: <https://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738>.
- CHISHOLM, L. A.; MARTHICK, J. Spectral and spatial variation at leaf and patch scale of invasive wetland weeds. In: **IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium**. Boston, USA: IEEE, 2008. v. 1, p. I-265–I-267.
- CHO, M. A.; MATHIEU, R.; DEBBA, P. Multiple endmember spectral-angle-mapper (sam) analysis improves discrimination of savanna tree species. In: **2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing**. IEEE, 2009. p. 1–4. Disponível em: <http://dx.doi.org/10.1109/WHISPERS.2009.5289031>.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995. Disponível em: <https://doi.org/10.1007/BF00994018>.

COSTA, B. M. d. Tipos de pastagens, sob o ponto de vista ecológico.

AGPTEA - Biblioteca de Zootecnia, 2003. Acesso em: 4 set. 2025. Disponível em:

<https://www.bibliotecaagptea.org.br/zootecnia/forragens/artigos/TIPOS%20DE%20PASTAGENS%20SOB>

EMBRAPA. **Capim-annoni-2: origem, morfologia, características,**

disseminação. Bagé, RS, 1993. Reunião Regional de Avaliação

de Pesquisa da Embrapa Pecuária Sul. Disponível em:

<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/225160/1/REUNIAOREGION.DEAVAL..pdf>.

FARINHA, C. M. *et al.* Identificação e mapeamento de áreas infestadas por

capim-annoni (*eragrostis plana*) através do uso de índice de vegetação no município

de dom pedrito/rs. **Biodiversidade Brasileira**, v. 13, n. 1, 2023. Disponível em:

<https://doi.org/10.37002/biodiversidadebrasileira.v13i1.2175>.

FRIEDL, M.; BRODLEY, C. Decision tree classification of land cover

from remotely sensed data. **Remote Sensing of Environment**, Elsevier,

v. 61, n. 3, p. 399–409, setembro 1997. ISSN 0034-4257. Disponível em:

[http://dx.doi.org/10.1016/s0034-4257\(97\)00049-7](http://dx.doi.org/10.1016/s0034-4257(97)00049-7).

FÁVERO, L. P.; BELFIORE, P.; SOUZA, R. de F. **Data Science, Analytics and Machine**

Learning with R. 1. ed. Academic Press, 2023. ISBN 9780128242711. Disponível em:

<https://www.elsevier.com/books/data-science-analytics-and-machine-learning-with-r/favero/978-0-12-8242>

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. [S.l.]: Atlas, 2008. ISBN

9788522451425.

GONZÁLEZ, J. D. M. **Análise da susceptibilidade à invasão do capim-annoni-2**

sobre áreas do bioma Pampa do município de Aceguá-RS. Dissertação (Dissertação

de Mestrado) — Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre,

RS, 2017. Disponível em: <http://hdl.handle.net/10183/170412>.

GONZÁLEZ, J. D. M.; FONSECA, E. L. da; PEREZ, N. B. Análise do potencial de

invasão do capim-annoni-2 sobre áreas do bioma pampa no município de aceguá-rs. In:

Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto (SBSR). Santos, SP,

Brasil: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. v. 18, p. 4197–4204.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed.

Cambridge, MA: MIT Press, 2016. ISBN 9780262035613. Disponível em:

<https://www.deeplearningbook.org/>.

GOULART, I. C. G. *et al.* Controle do capim-annoni (*eragrostis plana*) em pastagens

naturais utilizando herbicidas pré-emergentes e diferentes métodos de manejo de

vegetação. **Planta Daninha**, v. 27, n. 1, p. 181–190, março 2009. ISSN 0100-8358.

Disponível em: <https://doi.org/10.1590/S0100-83582009000100023>.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical**

Learning: Data Mining, Inference, and Prediction. 2. ed. New York: Springer,

2009. (Springer Series in Statistics). ISBN 9780387848587. Disponível em:

<https://doi.org/10.1007/978-0-387-84858-7>.

HAUG, S. *et al.* Plant classification system for crop/weed discrimination without segmentation. In: **Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)**. IEEE, 2014. p. 1142–1149. Disponível em: <https://doi.org/10.1109/WACV.2014.6835733>.

HE, H. *et al.* Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: **2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)**. IEEE, 2008. p. 1322–1328. Disponível em: <https://doi.org/10.1109/IJCNN.2008.4633969>.

HE, Y.; GUO, X. Hyperspectral vegetation indices and their relationships with environmental conditions in salt marsh ecosystems. **Journal of Coastal Research**, Coastal Education and Research Foundation, v. 26, n. 5, p. 850–856, 2010.

HICKS, S. A. *et al.* On evaluation metrics for medical applications of artificial intelligence. **Journal of Medical Imaging**, SPIE, v. 9, n. 4, p. 1–10, 2022.

HOLMES, W. S. *et al.* Proximal near-infrared spectral reflectance characterisation of weeds species in new zealand pasture. In: **2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)**. IEEE, 2019. p. 1–6. Disponível em: <https://doi.org/10.1109/I2MTC.2019.8826876>.

IQBAL, I. M. *et al.* Identifying the spectral signatures of invasive and native plant species in two protected areas of pakistan through field spectroscopy. **Remote Sensing**, MDPI AG, v. 13, n. 19, p. 4009, outubro 2021. ISSN 2072-4292. Disponível em: <http://dx.doi.org/10.3390/rs13194009>.

JENSEN, J. R. **Remote Sensing of the Environment: An Earth Resource Perspective**. 2. ed. Upper Saddle River, NJ: Pearson, 2015. ISBN 978-0134058160.

JONES, H. G.; VAUGHAN, R. A. **Remote Sensing of Vegetation: Principles, Techniques, and Applications**. New York, USA: Oxford University Press, 2010. ISBN 9780199207794.

KRUSE, F. A. *et al.* The spectral angle mapper (sam) algorithm. In: **Summaries of the 4th Annual JPL Airborne Geoscience Workshop**. [S.l.]: Jet Propulsion Laboratory, Pasadena, CA, 1993. v. 1, p. 1–4. JPL Publication 93-26.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, Taylor & Francis, v. 47, n. 260, p. 583–621, dezembro 1952. ISSN 0162-1459. Disponível em: <https://doi.org/10.1080/01621459.1952.10483441>.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York, NY: Springer, 2013. ISBN 978-1-4614-6849-3. Disponível em: <https://doi.org/10.1007/978-1-4614-6849-3>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, maio 2015. ISSN 0028-0836, 1476-4687. Disponível em: <https://doi.org/10.1038/nature14539>.

LIAKOS, K. G. *et al.* Machine learning in agriculture: A review. **Sensors**, MDPI, v. 18, n. 8, p. 2674, 2018. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/18/8/2674>.

MACEDO, W. K. d. N. **Espectroscopia Vis-NIR portátil para a caracterização e predição da qualidade de consumo de mangas Palmer produzidas no Vale do São Francisco**. Dissertação (Dissertação (Mestrado em Agronomia – Produção Vegetal)) — Universidade Federal do Vale do São Francisco, Petrolina, PE, 2024. Orientador: Sérgio Tonetto de Freitas (Embrapa Semiárido). Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1163324/1/Espectroscopia-vis-nir-portatil-para-a-caracterizacao-e-predicao-da-qualidade-de-consumo-de-mangas-Palmer-produzidas-no-Vale-do-Sao-Francisco..pdf>.

MACHADO, L. C. **Aspectos fisiológicos de gramíneas nativas do bioma Pampa em relação ao excesso de alumínio**. Dissertação (Mestrado) — Universidade Federal de Santa Maria, 2020. Dissertação (Mestrado em Agronomia). Disponível em: <https://repositorio.ufsm.br/handle/1/23188>.

MALLMANN, C. *et al.* Leaf-level field spectroscopy to discriminate invasive species (psidium guajava l. and hovenia dulcis thunb.) from native tree species in the southern brazilian atlantic forest. **Remote Sensing**, MDPI AG, v. 15, n. 3, p. 791, janeiro 2023. ISSN 2072-4292. Disponível em: <http://dx.doi.org/10.3390/rs15030791>.

MATEI, A. P.; FILIPPI, E. E. O bioma pampa e o desenvolvimento regional no rio grande do sul. In: PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL. **Anais do 6º Encontro de Economia Gaúcha**. Porto Alegre, RS, Brasil, 2012. Disponível em: <https://cdn.fee.tche.br/eeg/6/mesa8/O_Bioma_Pampa_e_o_Desenvolvimento_Regional_no_RS.pdf> (acesso em: 5 set. 2025).

MEDEIROS, R. B.; FOCHT, T. Invasão, prevenção, controle e utilização do capim-annoni-2 (*Eragrostis plana* Nees) no rio grande do sul, brasil. **Pesquisa Agropecuária Gaúcha**, v. 13, n. 1/2, p. 105–114, 2007. Disponível em: <<https://revistapag.agricultura.rs.gov.br/ojs/index.php/revistapag/article/view/259>> (acesso em: 5 set. 2025).

MEDEIROS, R. B. d. *et al.* Seed longevity of *Eragrostis plana* nees buried in natural grassland soil. **Revista Brasileira de Zootecnia**, Sociedade Brasileira de Zootecnia, v. 43, n. 11, p. 561–567, novembro 2014. ISSN 1516-3598. Disponível em: <http://dx.doi.org/10.1590/S1516-35982014001100001>.

MOAZZAM, S. I. *et al.* A review of application of deep learning for weeds and crops classification in agriculture. In: **2019 International Conference on Robotics and Automation in Industry (ICRAI)**. IEEE, 2019. p. 1–6. Disponível em: <http://dx.doi.org/10.1109/icrai47710.2019.8967350>.

MOHANTY, S. P.; HUGHES, D. P.; SALATHÉ, M. Using deep learning for image-based plant disease detection. **Frontiers in Plant Science**, Frontiers Media SA, v. 7, setembro 2016. ISSN 1664-462X. Disponível em: <http://dx.doi.org/10.3389/fpls.2016.01419>.

MOUTA, N. *et al.* Sentinel-2 time series and classifier fusion to map an aquatic invasive plant species along a river—the case of water-hyacinth. **Remote Sensing**,

MDPI AG, v. 15, n. 13, p. 3248, junho 2023. ISSN 2072-4292. Disponível em: <http://dx.doi.org/10.3390/rs15133248>.

NAJAFABADI, M. Y.; HESAMI, M.; ESKANDARI, M. Machine learning-assisted approaches in modernized plant breeding programs. **Genes**, v. 14, n. 4, p. 777, abril 2023. ISSN 2073-4425. PubMedID: 37107535. Disponível em: <https://www.mdpi.com/2073-4425/14/4/777>.

NEVES, M. C. *et al.* Avaliação de sensor multiespectral em mapeamentos da infestação de pastagens naturais pelo capim-annoni. In: **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto**. Santos, SP: INPE, 2019. v. 19, p. 96348. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1117339>.

OK, A. O.; AKAR, Ö.; GÜNGÖR, O. Evaluation of random forest method for agricultural crop classification. **European Journal of Remote Sensing**, Informa UK Limited, v. 45, n. 1, p. 421–432, janeiro 2012. ISSN 2279-7254. Disponível em: <http://dx.doi.org/10.5721/eujrs20124535>.

OKOYE, K.; HOSSEINI, S. **R Programming: Statistical Data Analysis in Research**. Springer Nature Singapore, 2024. XV, 309 p. ISBN 9789819733859. Disponível em: <http://dx.doi.org/10.1007/978-981-97-3385-9>.

OLIVEIRA, P. L. G. *et al.* Artificial neural network and remote sensing combined to predict the aboveground biomass in the cerrado biome. **Anais da Academia Brasileira de Ciências**, Academia Brasileira de Ciências, v. 96, n. 3, p. e20221041, 2024. ISSN 0001-3765. Disponível em: <http://dx.doi.org/10.1590/0001-3765202420221041>.

PAL, M. Ensemble of support vector machines for land cover classification. **International Journal of Remote Sensing**, Informa UK Limited, v. 29, n. 10, p. 3043–3049, abril 2008. ISSN 1366-5901. Disponível em: <https://doi.org/10.1080/01431160802007624>.

PEREZ, N. B. Método integrado de recuperação de pastagens - mirapasto: prevenção e manejo de espécies indesejáveis no pampa brasileiro. In: RODRÍGUEZ, P. A. B.; CIURLANTI, C. T.; CARRASCO-LETELIER, L. (Ed.). **Libro de Actas del Tercer Seminario Internacional: Restauración del Bioma Pampa (Argentina, Brasil y Uruguay)**. Montevideo, Uruguay: Facultades de Ciencias y de Agronomía de la Universidad de la República; Instituto Nacional de Investigación Agropecuaria, 2022. p. 57–63. Disponível em: <https://hdl.handle.net/20.500.12008/35206>.

POWERS, D. M. W. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. **arXiv preprint arXiv:2010.16061**, 2020. Disponível em: <https://doi.org/10.48550/arXiv.2010.16061>.

RAINIO, O. *et al.* Evaluation metrics and statistical tests for machine learning. **Scientific Reports**, Nature Publishing Group, v. 14, n. 1, p. 1–15, 2024.

SALEEM, M. *et al.* Predicting thalassemia using feature selection techniques: A comparative analysis. **Diagnostics**, v. 13, n. 22, 2023. ISSN 2075-4418. PubMed ID: 37998577. Disponível em: <https://www.mdpi.com/2075-4418/13/22/3441>.

SARKAR, S. **Quantifying Soybean Phenotypes Using UAV Imagery and Machine Learning, Deep Learning Methods**. Dissertação (Mestrado) — University of Missouri–Columbia, Columbia, MO, USA, 2023. Disponível em: https://www.researchgate.net/publication/373002513_Assessment_of_Soybean_Lodging_Using_UAV_Ima

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford University Press, v. 52, n. 3/4, p. 591–611, dezembro 1965. ISSN 0006-3444. Disponível em: <http://dx.doi.org/10.2307/2333709>.

SHEN, M.; TANG, M.; LI, Y. Phenology and spectral unmixing-based invasive kudzu mapping: A case study in knox county, tennessee. **Remote Sensing**, MDPI AG, v. 13, n. 22, p. 4551, novembro 2021. ISSN 2072-4292. Disponível em: <http://dx.doi.org/10.3390/rs13224551>.

SILVA, N. T. d. **Conservação e manutenção de espécies nativas do gênero Paspalum no Bioma Pampa**. 2023.

SMUCNY, J. *et al.* Data augmentation with mixup: Enhancing performance of a functional neuroimaging-based prognostic deep learning classifier in recent onset psychosis. **NeuroImage: Clinical**, Elsevier BV, v. 36, p. 103214, 2022. ISSN 2213-1582. Disponível em: <http://dx.doi.org/10.1016/j.nicl.2022.103214>.

THENKABAIL, P. S. (Ed.). **Remote Sensing Handbook - Three Volume Set**. CRC Press, 2018. ISBN 9781482282672. Disponível em: <http://dx.doi.org/10.1201/b19355>.

THENKABAIL, P. S.; LYON, J. G.; HUETE, A. **Advances in Hyperspectral Remote Sensing of Vegetation**. 2. ed. Boca Raton, FL: CRC Press, 2018. v. 3. ISBN 9781138031456.

THENKABAIL, P. S.; LYON, J. G.; HUETE, A. **Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation**. 2. ed. Boca Raton, FL: CRC Press, 2018. v. 1. ISBN 9781138031456.

THENKABAIL, P. S.; LYON, J. G.; HUETE, A. **Vegetation Index and Imaging Spectroscopy**. 2. ed. Boca Raton, FL: CRC Press, 2018. v. 2. ISBN 9781138031456.

USTIN, S. L. Remote sensing of canopy chemistry. **Proceedings of the National Academy of Sciences**, Proceedings of the National Academy of Sciences, v. 110, n. 3, p. 804–805, janeiro 2013. ISSN 1091-6490. Disponível em: <http://dx.doi.org/10.1073/pnas.1219393110>.

VALICHARLA, S. K. *et al.* Detection and multi-class classification of invasive knotweeds with drones and deep learning models. **Drones**, MDPI, v. 8, n. 7, p. 293, 2024.

VEIGAS, R. S. **Aplicação da detecção remota na monitorização de espécies exóticas invasoras em ambientes aquáticos**. Dissertação (Dissertação de Mestrado) — Universidade de Coimbra, julho 2023.

YANO, I. H. *et al.* Identification of weeds in sugarcane fields through images taken by uav and random forest classifier. **IFAC-PapersOnLine**, Elsevier BV, v. 49, n. 16, p. 415–420, 2016. ISSN 2405-8963. Disponível em: <http://dx.doi.org/10.1016/j.ifacol.2016.10.076>.

ZHANG, H. *et al.* mixup: Beyond empirical risk minimization. In: **International Conference on Learning Representations (ICLR)**. [S.l.: s.n.], 2017.

APÊNDICE A – TABELAS DAS ANÁLISES ESTATÍSTICAS E IMPORTÂNCIA DOS COMPRIMENTOS DE ONDA

As tabelas completas com os resultados das análises estatísticas e da importância dos comprimentos de onda estão disponíveis nos seguintes links:

- Resultados das Análises Estatísticas:

<https://docs.google.com/spreadsheets/d/18tEnanvb6uOCkpoICoWy-bhmf68Wdk7o/edit?usp=drive_link&oid=112870428854896110277&rtpof=true&sd=true>(=📄)

- Importância dos Comprimentos de Onda:

<https://docs.google.com/spreadsheets/d/1ctPIBBJv1i1jWudZXss3AILJOAkUpDiE/edit?usp=drive_link&oid=112870428854896110277&rtpof=true&sd=true>(=📄)

Esses arquivos contêm todas as variáveis analisadas e seus respectivos resultados completos, complementando as tabelas resumidas apresentadas no texto principal.