



GLÓRIA MARIA PADOVANI EDERLI

**CLASSIFICAÇÃO DE ÁREAS INFESTADAS COM NEMATOIDES
BASEADA EM ALGORITMOS DE APRENDIZADO DE MÁQUINA
UTILIZANDO IMAGENS HIPERESPECTRAIS DE ALTA RESOLUÇÃO
ESPACIAL**



Presidente Prudente - SP

2023

GLÓRIA MARIA PADOVANI EDERLI

**CLASSIFICAÇÃO DE ÁREAS INFESTADAS COM NEMATOIDES
BASEADA EM ALGORITMOS DE APRENDIZADO DE MÁQUINA
UTILIZANDO IMAGENS HIPERESPECTRAIS DE ALTA RESOLUÇÃO
ESPACIAL**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Ciências Cartográficas da Faculdade de Ciência e Tecnologia – UNESP – Universidade Estadual “Júlio de Mesquita Filho” – Campus de Presidente Prudente – SP, como parte dos requisitos para obtenção do título de Mestre em Ciências Cartográficas.

Orientador: Prof. Dr. Aluir Porfírio Dal Poz

Coorientadores: Prof. Dr. Nilton Nobuhiro Imai e
Prof. Dr. Adilson Berveglieri

Presidente Prudente - SP

2023

E22c

Ederli, Glória Maria Padovani

Classificação de áreas infestadas com nematoides baseada em algoritmos de aprendizado de máquina utilizando imagens hiperespectrais de alta resolução espacial / Glória Maria Padovani Ederli. -- , 2024

88 p. : il., tabs., fotos, mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Farmacêuticas, Araraquara,

Orientador: Aluir Porfírio Dal Poz

Coorientador: Nilton Nobuhiro Imai

1. Ciências exatas. 2. Sensoriamento remoto. 3. Agricultura de precisão. I.
Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências
Farmacêuticas, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

O IMPACTO ESPERADO

A pesquisa propõe o uso de sensoriamento remoto para detectar nematoides em plantações de soja, reduzindo custos e aumentando eficiência. O benefício inclui manejo mais preciso, preservação da produtividade agrícola e potencial impacto econômico ao evitar perdas.

THE EXPECTED IMPACT

The research proposes the use of remote sensing to detect nematodes in soybean fields, reducing costs and increasing efficiency. The benefits include more precise management, preservation of agricultural productivity, and potential economic impact by preventing losses.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: CLASSIFICAÇÃO DE ÁREAS INFESTADAS COM NEMATOIDES BASEADA EM ALGORITMOS DE APRENDIZADO DE MÁQUINA UTILIZANDO IMAGENS HIPERESPECTRAIS DE ALTA RESOLUÇÃO ESPACIAL

AUTORA: GLÓRIA MARIA PADOVANI EDERLI

ORIENTADOR: ALUIR PORFIRIO DAL POZ

COORDENADOR: NILTON NOBUHIRO IMAI

COORDENADOR: ADILSON BERVEGLIERI

Aprovada como parte das exigências para obtenção do Título de Mestra em Ciências Cartográficas, área: Aquisição, Análise e Representação de Informações Espaciais pela Comissão Examinadora:



Documento assinado digitalmente

ALUIR PORFIRIO DAL POZ

Data: 30/01/2024 15:06:00-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. ALUIR PORFIRIO DAL POZ (Participação Presencial)

Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Profa. Dra. FERNANDA SAYURI YOSHINO WATANABE (Participação Virtual)

Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Prof. Dr. FÁBIO FERNANDO DE ARAÚJO (Participação Virtual)

Programa de Pós-graduação em Agronomia / Universidade do Oeste Paulista

Presidente Prudente, 19 de abril de 2023

DEDICATÓRIA

*À minha família por todo amor, apoio e
por proporcionarem momentos de
alegria. Aos meus amigos pelo
companheirismo.*

AGRADECIMENTOS

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido durante o desenvolvimento do projeto de pesquisa, por meio da bolsa de estudo (Processo 88887.482203/2020-00).

Ao Programa de Pós-Graduação em Ciências Cartográficas (PPGCC) e à Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) pela infraestrutura disponibilizada, fundamental para a realização da pesquisa de mestrado.

Aos meus coorientadores, Prof. Dr. Nilton Nobuhiro Imai e Prof. Dr. Adilson Berveglieri, expresso minha profunda gratidão pelos ensinamentos e constante apoio ao longo desta desse trabalho.

Ao Prof. Dr. Aluir Porfírio Dal Poz, sou grata pela orientação sólida e pelo incentivo ao longo dos últimos anos, que foram fundamentais para o meu crescimento acadêmico e profissional.

Aos amigos e colegas do PPGCC, em especial à turma do processo seletivo de 2020 e aos frequentadores do Laboratório de Fotogrametria e Sensoriamento Remoto, agradeço pelas conversas e contribuições. À Allana, Roberta, Vitória, João e Mariana, pela amizade.

Expresso minha gratidão a todos os professores do Departamento de Cartografia e do PPGCC, cujos ensinamentos durante os cursos de graduação e pós-graduação foram fundamentais para minha formação.

À banca de qualificação e da defesa final, agradeço pelas correções e sugestões construtivas, que foram essenciais para aprimorar a versão final desta dissertação.

Aos meus pais, Antonio e Sueli, por seu apoio, e aos meus irmãos, Jonas, Natana, Daniel, Rebeca e Maria Angélica, por estarem sempre presentes, oferecendo ajuda e suporte. Um agradecimento especial à minha sobrinha Helena, por proporcionar momentos de alegria.

Agradeço a todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

A detecção de nematoides em plantações é fundamental para evitar prejuízos na produtividade, entretanto, o diagnóstico preciso da distribuição populacional requer uma grande quantidade de amostras de solo e de raiz em todo o campo, o que implica em um custo elevado. Nesse contexto, as técnicas de sensoriamento remoto podem auxiliar no manejo e controle dos nematoides, especialmente as que envolvem Veículos Aéreos Não Tripulados (VANTs). Este estudo propõe o uso de imagens hiperespectrais obtidas por câmera embarcada em VANT e técnicas de fotogrametria e sensoriamento remoto, para detectar e classificar as áreas parasitadas por nematoides na cultura de soja, de acordo com a intensidade de infestação a partir de algoritmos de aprendizado de máquina juntamente com amostras de solo coletadas em campo para validação. Foram testados os algoritmos de aprendizado de máquina *Random Forest*, *Support Vector Machine* e *K-Nearest Neighbors*, além do índice desenvolvido que relaciona os fitonematoides de uma região com os nematoides de vida livre (IFVL). Foram calculadas as matrizes de confusão e as métricas estatísticas sensibilidade, precisão, F_1 -score, RMSE e acurácia balanceada para avaliar os modelos gerados. Os resultados indicam que o melhor modelo foi obtido com o algoritmo *Random Forest*, e que a melhor divisão da intensidade de infestação de nematoides foi representada por cinco classes calculadas a partir do IFVL que alcançou uma acurácia balanceada de 96,5%. Também foi avaliado o melhor período fenológico para a coleta dos dados, que apresentou ser no fim do período reprodutivo, quando a diferença espectral entre as plantas saudáveis e parasitadas com fitonematoides aumenta, em que a região entre 771 nm e 790 nm demonstrou grande importância para detectar as áreas parasitadas. Portanto, este estudo conclui que as imagens hiperespectrais podem ser uma alternativa para o manejo e controle de fitonematoides em culturas de soja, reduzindo custos e aumentando a eficiência na detecção de áreas infectadas.

Palavras-chave: Agricultura de precisão, Nematoides, Aprendizado de Máquina, Imagens hiperespectrais.

Abstract

The detection of nematodes in plantations is fundamental to avoid losses in productivity, however, the precise diagnosis of population distribution requires a large amount of soil and root samples throughout the field, which implies a high cost. In this context, remote sensing techniques can support in nematode management and control, especially those involving Unmanned Aerial Vehicles (UAVs). This study proposes the use of hyperspectral images obtained by a camera onboard UAVs and remote sensing and photogrammetry techniques to detect and classify areas parasitized by nematodes in soybean crops according to the intensity of infestation using machine learning algorithms along with soil samples collected in the field for validation. Machine learning algorithms including Random Forest, Support Vector Machine, and K-Nearest Neighbors, as well as the developed index that relates plant parasitic nematodes to free-living nematodes (IFVL), were tested. Confusion matrix and statistical metrics including , sensitivity, precision, F1-score, RMSE, and balanced accuracy were calculated to evaluate the generated models. Results indicate that the best model was obtained using the Random Forest algorithm and that the best division of nematode infestation intensity was represented by five classes calculated from the IFVL, which achieved a balanced accuracy of 96.5%. The best phenological period for data collection was also evaluated, which was found to be at the end of the reproductive period when the spectral difference between healthy and nematode-infested plants increases, with the region between 771 nm and 790 nm demonstrating great importance in detecting parasitized areas. Therefore, this study concludes that hyperspectral images can be an alternative for nematode management and control in soybean crops, reducing costs and increasing efficiency in detecting infected areas.

Keywords: Precision agriculture, Nematodes, Machine learning, Hyperspectral images.

Lista de tabelas

Tabela - 1 Bandas espectrais do Sentinel-2.	30
Tabela - 2 Especificações técnicas da câmera hiperespectral Rikola.	31
Tabela - 3 Comprimentos de onda central das bandas utilizadas da câmera Rikola.	56
Tabela - 4 Matrizes de confusão normalizadas.	73
Tabela - 5 RMSE.	74
Tabela - 6 Sensibilidade, precisão e F_1 para 2 classes.	75
Tabela - 7 Sensibilidade, precisão e F_1 para 5 classes.	75
Tabela - 8 Sensibilidade, precisão e F_1 para 10 classes.	76
Tabela - 9 Acurácia balanceada.	77
Tabela - 10 Valores numérico de cada classe.	80

Lista de ilustrações

Figura - 1 Escala fenológica da soja.	26
Figura - 2 Raízes da soja: (a) plantas saudáveis e (b) com nematoides das lesões radiculares.	28
Figura - 3 Soja atacada pelo nematoide de galhas: (a) lavoura e (b) raízes.	28
Figura - 4 Câmera hiperespectral Rikola.	31
Figura - 5 Intervalo espectral dos sensores da câmara Rikola.	32
Figura - 6 Princípio do Interferômetro de Fabry-Perot (FPI).	33
Figura - 7 Estrutura interna da planta.	34
Figura - 8 Exemplo do funcionamento do algoritmo <i>Random Forest</i>	39
Figura - 9 Área de estudo localizada no Município de Caiuá.	44
Figura - 10 Divisão dos blocos da fazenda experimental.	45
Figura - 11 Grupo de parcelas utilizadas para a análise das respostas espectrais a partir de imagens da missão Sentinel-2.	46

Figura - 12 Grupo de parcelas utilizadas para extrair informações para treinamento dos algoritmos.	47
Figura - 13 (A) Distribuições dos pontos de amostras nematológicas de uma parcela e (B) área considerada por ponto de amostra.	48
Figura - 14 Estatísticas descritivas das amostras utilizadas como referência.	49
Figura - 15 Distribuição dos pontos das amostras nematológicas utilizadas para análise da resposta espectral.	51
Figura - 16 Distribuição dos pontos de controle e de verificação para fototriangulação das imagens.	52
Figura - 17 Exemplos dos alvos de isopor no ortomosaico. (A) Parcela, (B) e (C) alvos de isopor.	54
Figura - 18 VANT modelo X800 (GEO).	54
Figura - 19 Alvos de referência no processo de calibração por linha empírica.	55
Figura - 20 Espectrorradiômetro FieldSpec® Handheld.	56
Figura - 21 Distribuição dos pontos das amostras nematológicas.	58
Figura - 22 Área considerada para cálculo da média das respostas espectrais.	58
Figura - 23 Exemplo de seleção de conjunto de pixels aleatórios para cálculo da média das respostas espectrais: (A) considerando 50 pixels dentro da área, (B) considerando 75 e (C) considerando 95.	59
Figura - 24 Média das respostas espectrais das classes da data 1 (Dezembro). ...	64
Figura - 25 Média das respostas espectrais das classes da data 2 (Janeiro).	64
Figura - 26 Média das respostas espectrais das classes da data 3 (Fevereiro).	65
Figura - 27 Gráfico das médias das respostas espectrais em dezembro de 2021. ...	65
Figura - 28 Gráfico das médias das respostas espectrais em janeiro de 2022.	66
Figura - 29 Gráfico das médias das respostas espectrais em fevereiro de 2022.	66
Figura - 30 Gráfico da diferença entre as respostas espectrais da data de janeiro e dezembro.	67
Figura - 31 Gráfico da diferença entre as respostas espectrais da data de fevereiro e janeiro.	68
Figura - 32 Ortomosaico gerado a partir das imagens adquiridas pelo DJI Phantom 4.	70

Figura - 33 Exemplos de ortomosaicos gerados a partir das imagens adquiridas pelo sensor hiperespectral.	71
Figura - 34 Gráfico das médias das respostas espectrais para o sensor hiperespectral.	71
Figura - 35 Classificação das variáveis por ordem de importância.	79
Figura - 36 Resultado utilizando o classificador RF e 5 classes.	81

Lista de abreviaturas e siglas

CART	<i>Classification and Regression Trees</i>
CEPEA	Centro de Estudos Avançados em Economia Aplicada
CMOS	<i>Complementary Metal Oxide Semiconductor</i>
Colab	Google Colaboratory
EVA	Etileno Vinil Acetato
ESA	<i>European Space Agency</i>
FPI	<i>Fabry-Perot Interferometer</i>
FSLP	<i>Feature Selection via Linear Programming</i>
FWHM	<i>Full Width at Half Maximum</i>
GNSS	<i>Global Navigation Satellite System</i>
GSD	<i>Ground Sampling Distance</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
ICE	<i>Interactive Correlation Environment</i>
IFVL	Índice fitonematoides e nematoides de vida livre
KNN	<i>K-Nearest Neighbors</i>
MMQ	Método dos Mínimos Quadrados
MSI	<i>MultiSpectral Instrument</i>
ND	Número Digital
NRI	<i>Normalized Ratio Indices</i>
ONU	Organização das Nações Unidas
PIB	Produto Interno Bruto
PPK	<i>Post Processed Kinematic</i>
QGIS	Quantum GIS
RBF	<i>Radial Basis Function</i>

REM	Radiação Eletromagnética
RF	<i>Random Forest</i>
RTK	<i>Real Time Kinematic</i>
SBN	Sociedade Brasileira de Nematologia
SIG	Sistemas de Informação Geográfica
SVM	<i>Support Vector Machine</i>
SWIR	<i>Short-Wave Infrared</i>
TPU	<i>Tensor Processing Unit</i>
UNOESTE	Universidade do Oeste Paulista
USDA	<i>United States Department of Agriculture</i>
VANT	Veículos Aéreos Não Tripulados

Sumário

1	INTRODUÇÃO	17
1.1	Objetivos	20
1.1.1	Geral	20
1.1.2	Específicos	20
1.2	Justificativa	21
1.3	Justificativa	23
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Fenologia da soja	25
2.2	Nematoides na soja	27
2.3	Missão Sentinel-2	29
2.4	Câmera hiperespectral Rikola	30
2.4.1	FPI	32
2.5	Interação da radiação eletromagnética com a vegetação	33
2.6	Calibração Radiométrica	35
2.7	Aprendizado de máquina na agricultura de precisão	37
2.7.1	Random Forest	37
2.7.2	Support Vector Machine	41
2.7.3	K-Nearest Neighbors	42
3	MATERIAL E MÉTODOS	44
3.1	Materiais	44
3.1.1	Área de estudo e dados	44
3.1.2	Dados de Referência	47
3.1.3	Equipamentos e aplicativos	49
3.2	Metodologia	50
3.2.1	Avaliação das respostas espectrais na cultura da soja com diferentes níveis de infestação por fitonematoides nos estádios fenológicos	50
3.2.2	Levantamento Geodésico	52

3.2.3 Levantamento aéreo e processamento de imagens digitais	53
3.2.4 Índice fitonematoides e nematoides de vida livre	57
3.2.5 Extração de informações para treinamento de algoritmo de aprendizado de máquina	57
3.2.6 Algoritmos de aprendizado de máquina	59
3.2.7 Variáveis de importância	61
3.2.8 Métricas de avaliação dos classificadores	62
4 RESULTADOS E DISCUSSÕES	64
4.1 Análise das respostas espectrais na cultura da soja com diferentes níveis de infestação por fitonematoides nos estádios fenológicos utilizando imagens do sensor MSI/Sentinel-2	64
4.2 Análise das respostas espectrais com diferentes níveis de infestação por fitonematoides utilizando imagens do sensor hiperespectral	69
4.3 Métricas de avaliação dos classificadores	72
4.3.1 Matriz de confusão	72
4.3.2 Sensibilidade, precisão, F_1 score, RMSE e Acurácia balanceada	74
4.3.3 Variáveis por ordem de importância	78
4.4 Mapa temático	80
5 CONCLUSÕES E RECOMENDAÇÕES	82
REFERÊNCIAS	85

1 Introdução

A soja é a principal oleaginosa cultivada em todo o mundo. Durante o período de 2000 a 2020, o Brasil ocupou a posição de segundo maior produtor e exportador de soja. No entanto, a partir de 2020, o país conquistou o primeiro lugar nesse ranking, registrando uma produção de 126 milhões de toneladas e exportando 84 milhões de toneladas. O Brasil é responsável por aproximadamente 50% do comércio mundial de soja, de acordo com dados da Embrapa (ARAGAO; CONTINI, 2020).

Como ocorre em qualquer cultura agrícola, vários fatores podem impactar negativamente o processo de crescimento e produção da soja. Dentre esses fatores, destaca-se a presença de nematoides, especificamente o grupo dos nematoides parasitas de plantas, também conhecidos como fitonematoides. Esses microrganismos do solo possuem a capacidade de parasitar o sistema radicular das plantas ao longo de todo o ciclo de cultivo. Os fitonematoides apresentam baixa mobilidade e interferem na absorção adequada de nutrientes pelas plantas (PINHEIRO, 2022), comprometendo sua saúde ao longo de um período prolongado. Essa condição persiste mesmo após a colheita e pode afetar negativamente as safras subsequentes.

No Brasil, os fitonematoides mais prejudiciais à cultura são os formadores de galhas (*Meloidogyne spp.*), o nematoide de cisto (*Heterodera glycines*), o nematoide das lesões radiculares (*Pratylenchus brachyurus*) e o nematoide reniforme (*Rotylenchulus reniformis*) (DIAS et al., 2010). No oeste do Estado de São Paulo, onde predominam solos arenosos, os fitonematoides de maior predominância são os causadores das lesões radiculares (*Pratylenchus brachyurus*), que são considerados endoparasitas migratórios, e o nematoide das galhas (*Meloidogyne*), um endoparasita sedentário.

Além do grupo dos fitonematoides, outro grupo de nematoides que deve ser levado em consideração quando se trata de cultivares de soja é o grupo de nematoides de vida livre. Esse grupo, ao contrario dos fitonematoides, são organismos benéficos, não afetam a produção vegetal, mas estão diretamente ligados aos processos relevantes no solo, como reciclagem de carbono, compostos nitrogenados e de nutrientes minerais, com efeitos positivos para a fertilidade dos solos e produtividade das culturas agrícolas (FERRAZ; BROWN, 2016). Também são comumente utilizados como indicadores de perturbação ou qualidade do solo (BONGERS; FERRIS, 1999).

O sintoma mais comum que os fitonematoides causam são as reboleiras, que podem ser confundidas com sintomas de outros problemas recorrentes nos cultivos, como mancha de calcário, compactação do solo e deficiência nutricional, e assim podem passar despercebidos pelos produtores. Tradicionalmente, o diagnóstico de extensas áreas infestadas por nematoides e a determinação dos níveis de gravidade nessas áreas têm sido onerosos e nem

sempre viáveis do ponto de vista técnico. Uma estimativa precisa do potencial de danos requer a análise laboratorial de um grande número de amostras de solo para identificar os locais mais críticos em uma área infestada. As perdas causadas pelos fitonematoides, o alto custo da análise laboratorial de amostras de solo e a necessidade de controle eficaz exigem o desenvolvimento de métodos específicos de detecção e quantificação de danos em cada local (MARTINS; GALO; VIEIRA, 2017).

Dessa forma, uma alternativa que tem ganhado destaque nos últimos anos são as abordagens de agricultura de precisão que utilizam técnicas provenientes do sensoriamento remoto. Um princípio fundamental dessa abordagem é que a presença, distribuição e intensidade de um fator específico que reduz a produtividade em um campo devem ser identificáveis (MELAKEBERHAN, 2002). A reflectância do dossel da vegetação por meio de sensores remotos multiespectrais e hiperespectrais, tanto de longa distância quanto de curta distância, tem sido investigada para a detecção de diferentes nematoides e patógenos fúngicos presentes no solo (HEATH et al., 2000; LAUDIEN; BARETH; DOLUSCHITZ, 2003; LAWRENCE et al., 2004; WHEELER; KAUFMAN, 2003).

Essas técnicas permitem analisar as características espectrais das plantas e identificar possíveis infestações ou doenças que afetam o seu desempenho produtivo. As imagens multiespectrais e hiperespectrais são ferramentas eficientes na agricultura devido ao comportamento espectral das plantas, que permite detectar variações em suas características físicas (MORIYA, 2015). Essa capacidade de detecção de variações espectrais é especialmente útil na identificação de pragas e doenças, uma vez que as plantas infectadas apresentam comportamentos espectrais distintos das plantas saudáveis (RIEDEL; BLACKMER, 1999; ZHANG et al., 2003; MARIN et al., 2019). Além disso, de acordo com Hillnhütter et al. (2011), a indução de sintomas nas folhas, a formação de grupos (reboleiras) pelo parasita e sua limitada capacidade de mobilidade no solo tornam sua detecção facilitada por meio do sensoriamento remoto, permitindo a aplicação de técnicas de agricultura de precisão.

Os métodos de classificação de dados obtidos por sensores ópticos, baseados em aprendizado de máquina, têm recebido destaque devido à sua eficiência. Segundo Lary et al. (2016), os algoritmos baseados em aprendizado de máquina têm a capacidade de reconhecer padrões a partir de conjuntos de amostras de treinamento e generalizar a informação extraída para toda a imagem. Esses algoritmos não são limitados apenas a dados espectrais em seu processamento, podendo incluir outras fontes de informações relevantes para a classificação. O objetivo principal desses algoritmos é construir limites de decisão que possibilitem a maior separação entre os objetos, buscando minimizar os erros de classificação (VAPNIK, 1999).

Várias abordagens têm sido avaliadas para a detecção e classificação de nematoides em diferentes culturas, como em Martins (2013) em que avaliaram o potencial das imagens

multiespectrais RapidEye na detecção e discriminação de áreas de cana-de-açúcar infestadas por nematoides usando uma abordagem de classificação não supervisionada. Hillnhütter et al. (2011) utilizou dados hiperespectrais para detectar o estresse causado pelo nematoide do cisto na cultura de beterraba. Em Arantes et al. (2021) são utilizadas imagens aéreas hiperespectrais para a detecção espectral de nematoides na soja durante o estágio de floração, utilizando regressões lineares simples e múltiplas.

O uso dos dados provenientes de sensoriamento remoto combinado com algoritmos de aprendizado de máquina representam uma abordagem promissora para a detecção e classificação de nematoides. Estudos recentes, como os de Xavier et al. (2019) e Marra (2020), têm se dedicado a investigar o potencial das imagens multiespectrais do sensor MSI/Sentinel-2 para classificar nematoides na cultura do café utilizando redes neurais como técnica de aprendizado de máquina, esses estudos demonstraram eficiência na classificação dos nematoides. Porém, é importante ressaltar que no conjunto de estudos focados na detecção e classificação de fitonematoides a presença de nematoides de vida livre não são consideradas. Para uma compreensão completa das interações que ocorrem na área estudada, é fundamental considerar a presença e a influência dos nematoides de vida livre, além dos fitonematoides.

Para um mapeamento mais detalhado de plantações, é necessário o uso de imagens com maior resolução espacial, obtidas por levantamentos fotogramétricos realizados em dias e horários específicos. Os Veículos Aéreos Não Tripulados (VANTs), combinados com sensores espectrais e novos algoritmos, têm possibilitado diversas aplicações de imagem digital para gerar produtos cartográficos de alta precisão em grandes áreas (COLOMINA; MOLINA, 2014). As câmeras utilizadas nos levantamentos fotogramétricos com VANT devem ser leves e de porte pequeno para viabilizar o levantamento aéreo de maneira eficiente. Por exemplo, a câmera hiperespectral Rikola, adotada neste trabalho, possui um peso aproximado de 700 gramas, o que permite seu transporte por um VANT.

Neste contexto, nesta pesquisa levantou-se a hipótese de que imagens hiperespectrais obtidas por câmara embarcada em VANT e técnicas baseadas na fotogrametria e no sensoriamento remoto podem ser utilizadas para classificar as áreas parasitadas por nematoides no cultivo de soja de acordo com a intensidade de infestação de fitonematoides e levando em consideração a presença de nematoides de vida livre na área. Sendo assim, o trabalho visa avaliar o potencial das imagens hiperespectrais de alta resolução espacial na classificação de acordo com a intensidade de infestação de nematoides na cultura de soja, utilizando os algoritmos baseados em aprendizado de máquina *Random Forest*, *Support Vector Machine* e *K-Nearest Neighbors* e validar os resultados por meio de amostras nematológicas coletadas em campo.

1.1 Objetivos

1.1.1 Geral

Analisar o potencial das imagens hiperespectrais para a classificação da intensidade de infestação de nematoides na cultura de soja, utilizando uma abordagem de aprendizado de máquina.

1.1.2 Específicos

Os objetivos específicos do projeto são:

1. Analisar a partir das imagens obtidas pelo sensor MSI/Sentinel-2 qual o período fenológico as plantas parasitadas apresentam maior diferença espectral em relação as plantas sadias;
2. Realizar a calibração radiométrica das imagens hiperespectrais obtidas por câmera embarcada em VANT e estimar o fator de reflectância espectral;
3. Treinar, validar e comparar classificadores baseados em técnicas de aprendizado de máquina, a fim de determinar o algoritmo mais eficaz na classificação de áreas parasitadas por nematoides na plantação de soja;
4. Avaliar o potencial das imagens hiperespectrais para a classificação de áreas parasitadas por nematoides no cultivo de soja, de acordo com a intensidade de infestação e baseado no melhor algoritmo;
5. Estimar a acurácia dos mapas temáticos produzidos, levando em consideração as amostras coletadas em campo e analisadas em laboratório, a fim de avaliar a eficácia da abordagem proposta.

1.2 Justificativa

De acordo com o Censo Agropecuário do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2017, o Brasil utiliza cerca de 33,6% da sua extensão territorial para a produção agrícola. A soja é uma das principais culturas do país e ocupa aproximadamente 9% do território nacional, conforme dados do IBGE de 2020.

Em relação ao Produto Interno Bruto (PIB) do agronegócio brasileiro, o Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2020) aponta que houve um crescimento de 5,2% em 2020, representando cerca de 26,6% do PIB total brasileiro. Essas informações evidenciam a importância do agronegócio para a economia do país.

O conhecimento sobre a produção de culturas agrícolas que possuem grande importância no cenário econômico brasileiro é muito relevante para o país, principalmente para prever safras e definir estratégias adequadas para sua economia agrícola, com o propósito de resolver o problema de escassez ou do excesso de determinado produto (IPPOLITI; EPIPHANIO; SHIMABUKURO, 1998). A evolução dos sensores ópticos eletrônicos, técnicas fotogramétricas, sistemas de posicionamento global e tecnologias em geoprocessamento contribuem com muitas áreas de estudos, entre elas está a agricultura.

Segundo informações da Sociedade Brasileira de Nematologia (SBN), os nematoides causam prejuízos para o agronegócio brasileiro. De acordo com a instituição, considerando apenas a produção de soja, essa perda é de aproximadamente R\$16,2 bilhões ao ano. Os nematoides são um dos principais problemas fitossanitários da sojicultura no Brasil, uma vez que esses parasitas podem inviabilizar algumas áreas de cultivo de soja (OTOBONI, 2018). Embora sejam habitantes naturais do solo, algumas espécies de nematoides podem causar danos aos cultivos e prejuízo aos agricultores, pois o parasitismo dos nematoides provocam o crescimento deficiente da planta, o que causa a perda da produtividade. Esse tipo de praga parasita o solo e a raiz da planta dificultando sua identificação nas plantações. A incidência do parasita ocasiona o crescimento deficiente da planta e, dependendo da intensidade de infestação, pode resultar em uma perda significativa da produção.

Os nematoides são considerados pragas de baixa mobilidade no campo, deste modo, as safras podem demorar para apresentar prejuízos econômicos significativos o que pode causar a descoberta tardia da infestação. Deste modo, a utilização de técnicas baseadas na em informações provenientes do sensoriamento remoto podem contribuir para a detecção da presença dos parasitas nas plantações mais cedo.

De acordo com Plant (2001), a agricultura de precisão é compreendida pelo uso da tecnologia de informação em todo ciclo da produção agrícola, um processo em prol da agricultura sustentável. Schaer, Strauch e Knickel (2002) dizem que a agricultura de precisão é como um conjunto de tecnologias capaz de auxiliar o produtor rural a identificar as estratégias a serem adotadas para aumentar a eficiência no gerenciamento da agricultura,

maximizando a rentabilidade das colheitas, tornando o agronegócio mais competitivo, possibilitando a otimização do uso dos recursos edafoclimáticos e minimizando, assim, os insumos utilizados no sistema de produção de um local. Ainda, de acordo com Inamasu et al. (2011), a agricultura de precisão pode auxiliar com a racionalização da produção agropecuária e florestal, podendo ajudar na diminuição de perdas e no impacto ambiental, aumentando a renda dos produtores.

Ainda são escassos os estudos sobre detecção e classificação de fitonematoides em culturas de soja, apesar da agricultura exigir um monitoramento frequente devido à sua dinamicidade. O Sensoriamento Remoto e a Fotogrametria podem fornecer informações precisas para os produtores agrícolas (SEELAN et al., 2003). No estudo de Arantes et al. (2021), imagens coletadas por VANT foram utilizadas para detecção espectral de nematoides na soja durante o estágio de florescimento, onde sensores multiespectrais foram necessários para estimar o nematoide das lesões radiculares (*Pratylenchus brachyurus*). Vários processos baseados em sensores foram testados, tanto em plataformas terrestres quanto aéreas, para detectar o estresse causado por nematoides (HEATH et al., 2000; NUTTER et al., 2002; HILLNHÜTTER et al., 2012). O estudo de Martins et al. (2019) descreve o uso de imagens de sensoriamento remoto para identificar e mapear a infecção por nematoides de galha na cultura de café. Os autores coletaram dados espectrais de reflectância em diferentes comprimentos de onda e utilizaram algoritmos de aprendizado de máquina para distinguir entre áreas infectadas e não infectadas. Os resultados mostraram que a abordagem proposta pode ser eficaz na detecção precoce da infecção por nematoides de galha na cultura de café. No entanto, ainda não foram encontrados estudos que avaliem o nível de infestação de fitonematoides na soja considerando a quantidade de nematoides de vida livre na área, o que poderia ajudar no equilíbrio da comunidade de nematoides no solo.

As novas tecnologias na área da Agricultura de Precisão otimizam e automatizam os procedimentos envolvidos na agricultura, contribuindo para o desenvolvimento agrônômico, trazendo benefícios tanto para a produção agrícola quanto para o meio ambiente (ZHANG; WANG; WANG, 2002). Assim, essa tecnologia está relacionada a vários dos 17 objetivos da Organização das Nações Unidas (ONU) (Organização das Nações Unidas (ONU), acesso em 26 de fevereiro de 2023), como a redução da fome e da pobreza, auxiliando no aumento da produtividade e da eficiência da produção agrícola, o que pode ajudar na disponibilidade de alimentos e preços mais acessíveis para a população. Além disso, contribui para o desenvolvimento sustentável, ajudando a reduzir a utilização de insumos e protegendo o meio ambiente, reduzindo a emissão de gases de efeito estufa e minimizando a contaminação do solo e da água.

A Agricultura de Precisão também está ligada à inovação e tecnologia, já que é uma técnica inovadora que utiliza tecnologias avançadas, como sensores, drones e sistemas de

informação geográfica (SIG). Além disso, pode contribuir para a ação climática, reduzindo as emissões de gases do efeito estufa e aumentando a capacidade de adaptação das culturas às mudanças climáticas.

A pesquisa encontra-se no âmbito de interesse da comunidade científica brasileira da área de agricultura de precisão, fotogrametria e sensoriamento remoto, bem como no contexto do agronegócio brasileiro. Além disso, o projeto está alinhado com diversos objetivos de desenvolvimento sustentável estabelecidos pela ONU.

1.3 Justificativa

Segundo informações da Sociedade Brasileira de Nematologia (SBN), os nematoides causam prejuízos para o agronegócio brasileiro. De acordo com a instituição, considerando apenas a produção de soja, essa perda é de aproximadamente R\$16,2 bilhões ao ano. Os nematoides são um dos principais problemas fitossanitários da sojicultura no Brasil, uma vez que esses parasitas podem inviabilizar algumas áreas de cultivo de soja (OTOBONI, 2018). Embora sejam habitantes naturais do solo, algumas espécies de nematoides podem causar danos aos cultivos e prejuízo aos agricultores, pois o parasitismo dos nematoides provocam o crescimento deficiente da planta, o que causa a perda da produtividade. Esse tipo de praga parasita o solo e a raiz da planta dificultando sua identificação nas plantações. A incidência do parasita ocasiona o crescimento deficiente da planta e, dependendo da intensidade de infestação, pode resultar em uma perda significativa da produção.

Os nematoides são considerados pragas de baixa mobilidade no campo, deste modo, as safras podem demorar para apresentar prejuízos econômicos significativos o que pode causar a descoberta tardia da infestação. Deste modo, a utilização de técnicas baseadas na em informações provenientes do sensoriamento remoto podem contribuir para a detecção da presença dos parasitas nas plantações mais cedo.

De acordo com Plant (2001), a agricultura de precisão é compreendida pelo uso da tecnologia de informação em todo ciclo da produção agrícola, um processo em prol da agricultura sustentável. Schaer, Strauch e Knickel (2002) dizem que a agricultura de precisão é como um conjunto de tecnologias capaz de auxiliar o produtor rural a identificar as estratégias a serem adotadas para aumentar a eficiência no gerenciamento da agricultura, maximizando a rentabilidade das colheitas, tornando o agronegócio mais competitivo, possibilitando a otimização do uso dos recursos edafoclimáticos e minimizando, assim, os insumos utilizados no sistema de produção de um local. Ainda, de acordo com Inamasu et al. (2011), a agricultura de precisão pode auxiliar com a racionalização da produção agropecuária e florestal, podendo ajudar na diminuição de perdas e no impacto ambiental, aumentando a renda dos produtores.

Ainda são escassos os estudos sobre detecção e classificação de fitonematoides em

culturas de soja, apesar da agricultura exigir um monitoramento frequente devido à sua dinamicidade. O Sensoriamento Remoto e a Fotogrametria podem fornecer informações precisas para os produtores agrícolas (SEELAN et al., 2003). No estudo de Arantes et al. (2021), imagens coletadas por VANT foram utilizadas para detecção espectral de nematoides na soja durante o estágio de florescimento, onde sensores multiespectrais foram necessários para estimar o nematoide das lesões radiculares (*Pratylenchus brachyurus*). Vários processos baseados em sensores foram testados, tanto em plataformas terrestres quanto aéreas, para detectar o estresse causado por nematoides (HEATH et al., 2000; NUTTER et al., 2002; HILLNHÜTTER et al., 2012). O estudo de Martins et al. (2019) descreve o uso de imagens de sensoriamento remoto para identificar e mapear a infecção por nematoides de galha na cultura de café. Os autores coletaram dados espectrais de reflectância em diferentes comprimentos de onda e utilizaram algoritmos de aprendizado de máquina para distinguir entre áreas infectadas e não infectadas. Os resultados mostraram que a abordagem proposta pode ser eficaz na detecção precoce da infecção por nematoides de galha na cultura de café. No entanto, ainda não foram encontrados estudos que avaliem o nível de infestação de fitonematoides na soja considerando a quantidade de nematoides de vida livre na área, o que poderia ajudar no equilíbrio da comunidade de nematoides no solo.

As novas tecnologias na área da Agricultura de Precisão otimizam e automatizam os procedimentos envolvidos na agricultura, contribuindo para o desenvolvimento agrônomo, trazendo benefícios tanto para a produção agrícola quanto para o meio ambiente (ZHANG; WANG; WANG, 2002). Assim, essa tecnologia está relacionada a vários dos 17 objetivos da Organização das Nações Unidas (ONU) (Organização das Nações Unidas (ONU), acesso em 26 de fevereiro de 2023), como a redução da fome e da pobreza, auxiliando no aumento da produtividade e da eficiência da produção agrícola, o que pode ajudar na disponibilidade de alimentos e preços mais acessíveis para a população. Além disso, contribui para o desenvolvimento sustentável, ajudando a reduzir a utilização de insumos e protegendo o meio ambiente, reduzindo a emissão de gases de efeito estufa e minimizando a contaminação do solo e da água.

A Agricultura de Precisão também está ligada à inovação e tecnologia, já que é uma técnica inovadora que utiliza tecnologias avançadas, como sensores, drones e sistemas de informação geográfica (SIG). Além disso, pode contribuir para a ação climática, reduzindo as emissões de gases do efeito estufa e aumentando a capacidade de adaptação das culturas às mudanças climáticas.

A pesquisa encontra-se no âmbito de interesse da comunidade científica brasileira da área de agricultura de precisão, fotogrametria e sensoriamento remoto, bem como no contexto do agronegócio brasileiro. Além disso, o projeto está alinhado com diversos objetivos de desenvolvimento sustentável estabelecidos pela ONU.

2 Fundamentação teórica

Nesta Seção, será apresentada a fundamentação teórica que servirá como apoio para o desenvolvimento deste trabalho. Sendo assim, na Subseção 2.1 será apresentado um resumo sobre fenologia da soja e na Subseção 2.2 será apresentado uma breve introdução sobre nematoides, especialmente sobre os nematoides das lesões radiculares (*Pratylenchus brachyurus*), os formadores de galhas (*Meloidogyne spp.*), *Helicotylenchus* e *Criconebella*. Em seguida, será apresentado um resumo sobre a Missão Sentinel-2 (Subseção 2.3) e sobre a Câmera hiperespectral Rikola (Subseção 2.4), que foram os sensores de aquisição de imagens para esta pesquisa. Na Subseção 2.5 é apresentada a interação da radiação eletromagnética com a vegetação, na Subseção 2.6 é descrita algumas teorias da calibração radiométrica e na Subseção 2.7 é apresentado um resumo sobre aprendizado de máquina na agricultura de precisão, com ênfase nos algoritmos *Random Forest*, *Support Vector Machine* e *K-Nearest Neighbors*.

2.1 Fenologia da soja

A Fenologia é o estudo dos fenômenos periódicos dos seres vivos e suas relações com as condições do ambiente e a correlação com os aspectos morfológicos. A Fenologia de uma planta representa o seu desenvolvimento ao longo de suas diferentes etapas (OLIVEIRA et al., 2018). Existem uma grande variedade de cultivares de soja com ciclos que podem variar entre 100 a 160 dias.

A soja faz parte do conjunto de atividades agrícolas com maior destaque no mercado mundial. Segundo o Departamento de Agricultura dos Estados Unidos (USDA- *United States Department of Agriculture*), a safra 2019/20 fez do Brasil o maior produtor de soja do mundo. Estudar a Fenologia da soja é uma ferramenta importante para compreender quando os sintomas de pragas ou parasitas podem estar evidenciados e mais perceptíveis nas imagens, contribuindo para a seleção da melhor época e do melhor intervalo do espectro eletromagnético para as tomadas das imagens.

A dinâmica do comportamento fenológico das culturas agrícolas é uma característica relevante a ser estudada quando se trata de extração de informações de interesse para agricultura de precisão, uma vez que evidência o perfil da cultura no decorrer do tempo (FORMAGGIO; SANCHES, 2017). A altura e a folhagem da planta variam ao longo do ciclo da plantação e influencia na resposta espectral registrada pelo sensor. Isto porque as respostas espectrais captadas pelos sensores acompanham as variações das quantidades de folhas e de recobrimento vegetal sobre a superfície do solo.

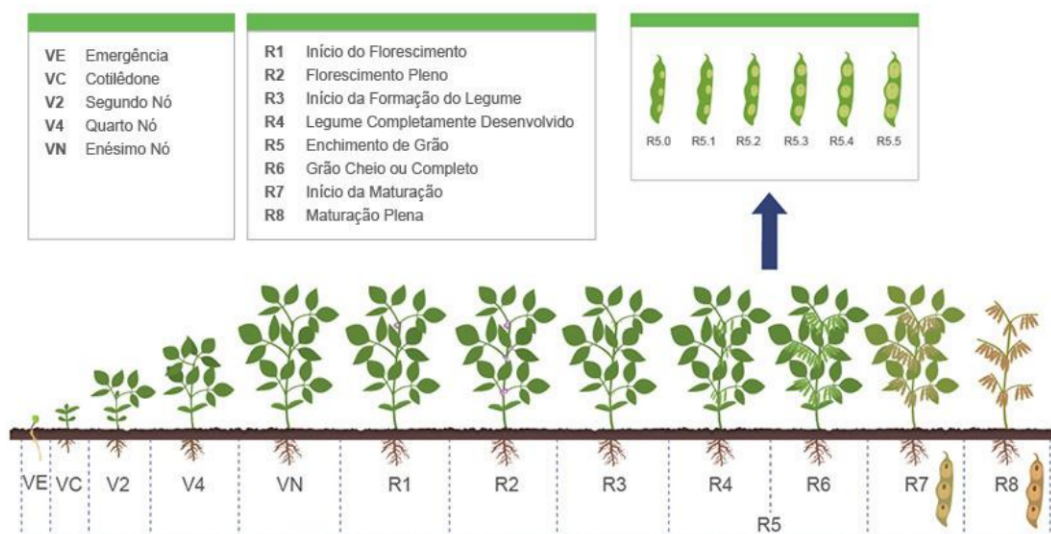
Os estádios fenológicos de desenvolvimento da planta de soja foram descritos

por Fehr e Caviness (1977); neste sistema, os estádios são divididos em vegetativos e reprodutivos. Os estádios vegetativos são representados pela letra V e os reprodutivos pela letra R, com exceção dos estádios VE (emergência) e VC (cotilédone). As letras são seguidas de um índice numérico que identificam os estádios específicos.

O estágio VE representa uma plântula recém emergida, ou seja, a emergência dos cotilédones. O estágio vegetativo VC representa o estágio em que os cotilédones se encontram completamente abertos e expandidos. A partir do VC, as subdivisões dos estádios vegetativos são sequencialmente numeradas, como V1, V2, V3, V4, V5, V6,...Vn, onde Vn é o último nó no topo da planta com a folha completamente desenvolvida (NEUMAIER et al., 2000).

De acordo com Neumaier et al. (2000), os estádios reprodutivos descrevem o período da cultura de florescimento-maturação, são denominados pela letra R seguida dos números de um até oito. Este estágio abrange quatro fases diferentes do desenvolvimento reprodutivo da planta, o florescimento (R1 e R2), desenvolvimento da vagem (R3 e R4), desenvolvimento do grão (R5 e R6) e maturação da planta (R7 e R8), conforme Figura 1.

Figura 1 – Escala fenológica da soja.



Fonte: Fehr e Caviness (1977).

Existem algumas doenças e pragas que são mais comuns a um determinado período do ciclo fenológico. No período entre VE e V3 há mais pragas do solo; nos estádios entre VC a R7 é mais comum o início de lagartas nas lavouras; e entre o período do estágio R1 a R7 é possível encontrar percevejos. Os nematoides podem ser encontrados nas plantações de soja durante todo o ciclo fenológico, tornando uma ameaça para as culturas agrícolas.

Quando observada a vegetação natural ou agrícola por meio de sensores remotos, é interessante levar em consideração o comportamento espectral da cobertura vegetal. Na faixa do espectro eletromagnético, na região do visível (400 nm a 700 nm), verifica-se que

há uma forte influência dos pigmentos foliares que absorvem a radiação eletromagnética. Na faixa do infravermelho próximo (700 nm a 1.300 nm) é predominante a reflexão, em virtude da estrutura interna das folhas. E na faixa do infravermelho de ondas curtas (1.300 nm a 2.500 nm), ocorre a predominância de absorção devido ao conteúdo de umidade interna das folhas (FORMAGGIO; SANCHES, 2017).

2.2 Nematoides na soja

Das formas nocivas às plantas mais populares, encontram-se: dois tipos de animais metazoários, geralmente chamados de pragas, que são os insetos e os ácaros; os variados microrganismo considerados como agentes causadores de doenças, como, os fungos, bactérias ou vírus; e diferentes espécies de animais metazoários pertencentes ao filo Nematoda (ou Nemata), nomeado nematoides ou nematódeos, este último age como parasitas de tecidos vegetais, principalmente das raízes (FERRAZ; BROWN, 2016).

Os nematoides podem ser divididos em três grupos a partir de seus hábitos alimentares, os de vida livre, os zooparasitas e os fitoparasitas ou parasitas de plantas (ou ainda, fitonematoides), que é o foco do presente estudo. Mesmo causando grandes prejuízos na agricultura, os fitonematoides ainda são pouco explorados. Os nematoides parasitas de plantas afetam o fluxo de absorção e translocação de água e nutrientes na planta, podendo causar uma perda significativa na produtividade das plantas.

Os fitonematoides são seres vivos pequenos, medindo entre 0,2 a 3,0 mm de comprimento, que pode dificultar a percepção de sua presença nas plantações. Uma outra característica da maioria dos fitonematoides, que faz com que sua presença passe despercebida, é a hospedagem no sistema radicular das plantas, as raízes. Além das características citadas, que dificultam identificar este parasita, os nematoides não apresentam sintomas específicos nas plantas infectadas que possam ser detectados com precisão, ou seja, sua presença pode ser confundida facilmente com sintomas provocados por outros tipos de organismos ou por fatores abióticos, como, por exemplo, o solo e o clima. Com a falta de sintomas específicos e que possam ser caracterizados com precisão, comumente são adotadas práticas ineficientes e com pouco benefício pelos agricultores (FERRAZ; BROWN, 2016).

Existem mais de cem espécies de nematoides associadas aos cultivos de soja em todo mundo. No Brasil, os nematoides que mais afetam este tipo de cultura são o de cisto (*Heterodera glycines*), o reniforme (*Rotylenchulus reniformis*), o das lesões radiculares (*Pratylenchus brachyurus*) e os formadores de galhas (*Meloidogyne spp.*) (DIAS et al., 2010), sendo que as duas últimas espécies já foram encontradas na área de estudo desta pesquisa.

O nematoide *Pratylenchus brachyurus*, ou nematoide das lesões radiculares, é co-

mum no Brasil. As lavouras de soja que são atacadas por esses nematoides geralmente apresentam sintomas como reboleiras, e as raízes das plantas parasitadas ficam, parcialmente ou totalmente, escuras, conforme Figura 2. Contudo, a intensidade dos sintomas são dependentes de outros fatores, como, por exemplo, a textura do solo.

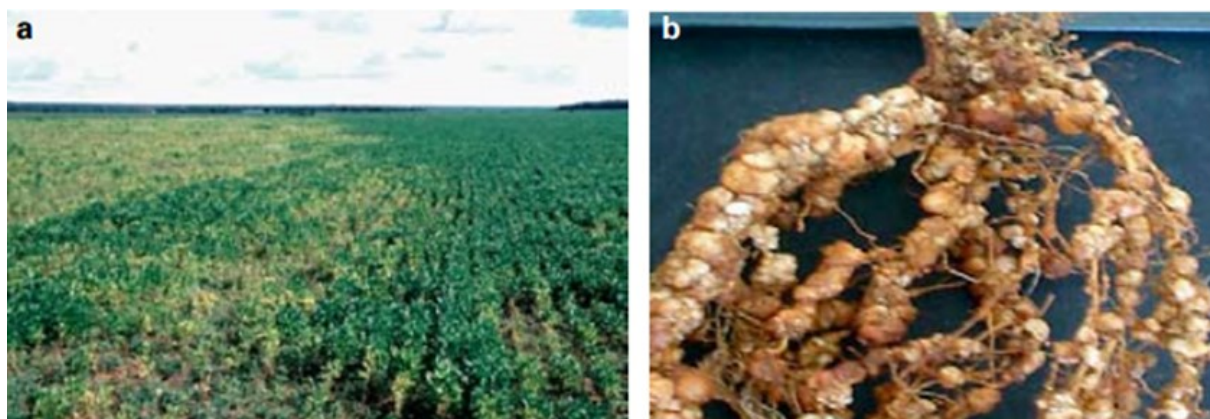
Figura 2 – Raízes da soja: (a) plantas saudáveis e (b) com nematoides das lesões radiculares.



Fonte: Dias et al. (2010).

O nematoide *Meloidogyne.spp* ou nematoide das galhas (Figura 3) é a outra espécie encontrada nas plantações de soja. Este é considerado o mais importante tipo de nematoide fitopatogênico causador de grandes danos às culturas, principalmente porque apresenta ampla distribuição geográfica e grande gama de hospedeiros (FREITAS, 2006).

Figura 3 – Soja atacada pelo nematoide de galhas: (a) lavoura e (b) raízes.



Fonte: Dias et al. (2010).

As lavouras que são infectadas por essa espécie de nematoide podem apresentar

manchas em reboleiras, em que as plantas ficam pequenas e amareladas. As folhas das plantas atacadas podem sofrer o efeito de folhas "carijó", em que formam manchas cloróticas ou necroses entre as nervuras. Nem sempre ocorre a redução do tamanho das plantas, mas é possível observar aumento no abortamento das vagens e amadurecimento prematuro das plantas infectadas. As raízes das plantas hospedeiras apresentam galhas em número e tamanho variados, podendo variar de acordo com a intensidade da população de nematoides no solo. As fêmeas dos nematoides ficam localizadas no interior das galhas, sua coloração é branco-pérola e seu formato é de pera (DIAS et al., 2010).

O gênero *Criconemella* e *Helicotylenchus* são compostos por nematoides que atacam uma ampla variedade de plantas, incluindo a soja. Esses são ectoparasitas que se alimentam da raiz da soja, causando danos diretos na planta e predispondo-a a infecções por patógenos e secundários. Os sintomas da infestação incluem raízes ralas, amarelamento das folhas, deficiência de nutrientes e redução do rendimento da cultura (DIAS et al., 2010).

Vários estudos relataram a redução do teor de clorofila, crescimento das plantas, taxa fotossintética e concentrações de nutrientes das plantas devido à infecção por nematoides (CARNEIRO et al., 2002; HASEEB; SRIVASTAVA; PANDEY, 1990; LU et al., 2014). Sintomas acima do solo decorrentes da infecção, como clorose e retardamento grave do crescimento, podem ser observados (NUTTER et al., 2002). Essas alterações fisiológicas nas plantas causadas por nematoides podem ser avaliadas por meio de sensoriamento remoto.

2.3 Missão Sentinel-2

A missão de imageamento multiespectral Copernicus Sentinel-2 compreende uma constelação de dois satélites, Sentinel-2A e Sentinel-2B, em órbita polar sincronizada com o Sol, divididos em 180° entre si. O satélite Sentinel-2A foi lançado em 23 de junho de 2015 e o satélite Sentinel-2B foi lançado em 07 de março de 2017, ambos equipados com o sensor MSI (*MultiSpectral Instrument*) que é capaz de obter imagens ópticas de alta resolução. Esta missão é destinada a monitorar a variabilidade nas condições da superfície da terra contando com uma faixa orbital de 290 km e uma revisita de dez dias no Equador com um satélite e cinco dias com dois satélites (ESA, 2021).

O sensor MSI possui resolução radiométrica de 12 bits e 13 bandas espectrais com três diferentes resoluções espaciais: quatro bandas com 10 m, seis bandas com 20 m e três bandas com 60 m. Na Tabela 1, são apresentadas as bandas com suas respectivas resoluções espaciais e comprimentos de onda central.

As imagens adquiridas pelo Sentinel-2 são disponibilizadas aos usuários nos níveis de correção 1C e 2A. As imagens disponíveis no nível 1C contém pixels expressos em reflectância de topo de atmosfera e o nível 2A pixels expressos em reflectância de superfície.

Os produtos disponíveis do Sentinel-2 são ortoimagens de 100 km² na projeção/Datum UTM/WGS84. O produto de nível 2A pode ser adquirido através do *download* pelo site da Agência Espacial Europeia (ESA), ou pelo processamento utilizando a ferramenta criada pela agência chamada *sen2cor*, que utiliza as imagens de nível 1C como base do processamento.

Tabela 1 – Bandas espectrais do Sentinel-2.

Resolução espacial(m)	Número da banda	Nome da banda	Comprimento de onda central (nm)	
			S2A	S2B
10	B2	Azul	492,4	492,1
	B3	Verde	559,8	559
	B4	Vermelho	664,6	664,9
	B8	Infravermelho Próximo	832,8	832,9
20	B5	<i>Red Edge</i> 1	704,1	703,8
	B6	<i>Red Edge</i> 2	740,5	739,1
	B7	<i>Red Edge</i> 3	782,8	779,7
	B8A	<i>Red Edge</i> 4	864,7	864
	B11	Infravermelho de Ondas Curtas 1	1613,7	1610,4
	B12	Infravermelho de Ondas Curtas 2	2202,4	2185,7
60	B1	Aerossol	442,7	442,2
	B9	<i>Water Vapor</i>	945,1	943,2
	B10	Cirrus	1373,5	1376,9

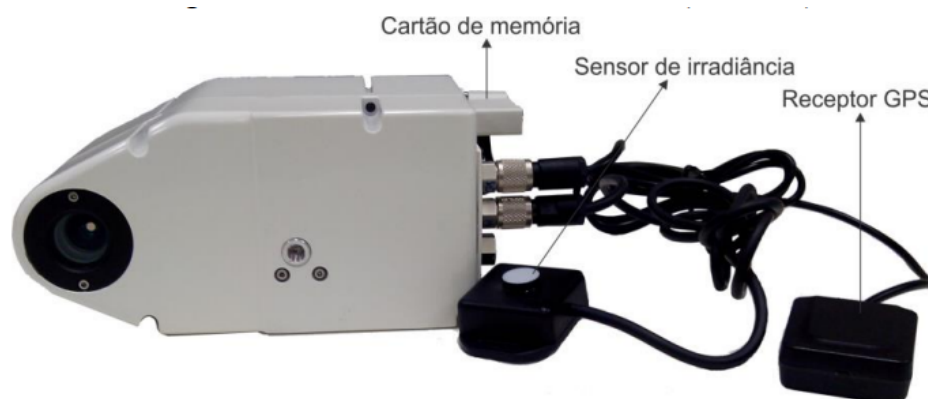
Fonte:(ESA, 2021).

Os produtos da missão Sentinel-2 disponíveis gratuitamente são ideais para a realização de um estudo prévio da área da cultura de soja, já que seu intervalo de revisita pode chegar a 5 dias. Além disso, o sensor possui bandas com 10 m e 20 m de resolução espacial, que possibilita visualizar algumas características da cultura.

2.4 Câmera hiperespectral Rikola

A câmera Rikola modelo FPI2015 possui sensor hiperespectral que adquire imagens no formato de quadro (Figura 4), desenvolvida pelo VTT - Centro de Pesquisa Técnico finlandês em parceria com a Rikola Ltd., ela é equipada com interferômetro *Fabry Perot* (FPI - *Fabry-Perot Interferometer*). Caracterizado como um sensor leve, a sua operação é definida pelo princípio de aquisição sequencial por tempo. Os hipercubos cobrem o espectro eletromagnético de 500 nm a 900 nm, em que os intervalos de largura podem ser variados. No modo manual, as bandas podem ser configuradas em um comprimento de onda específico. Entretanto, a largura da banda FWHM (*full width at half maximum*) não pode ser fixada pelo operador, pois o FWHM é predefinido pela câmera.

Figura 4 – Câmera hiperespectral Rikola.



Fonte: Miyoshi (2016).

A câmara está equipada com sensores auxiliares, incluindo um sensor de irradiância que mede a quantidade de energia solar que chega à superfície terrestre. Além disso, a câmara possui um receptor GPS que registra a posição geográfica em latitude e longitude no momento de aquisição de cada imagem. Os dados desses sensores são disponibilizados para o usuário por meio do arquivo de metadados que acompanha cada imagem adquirida. Na Tabela 2, são apresentadas as especificações técnicas da câmara.

Tabela 2 – Especificações técnicas da câmara hiperespectral Rikola.

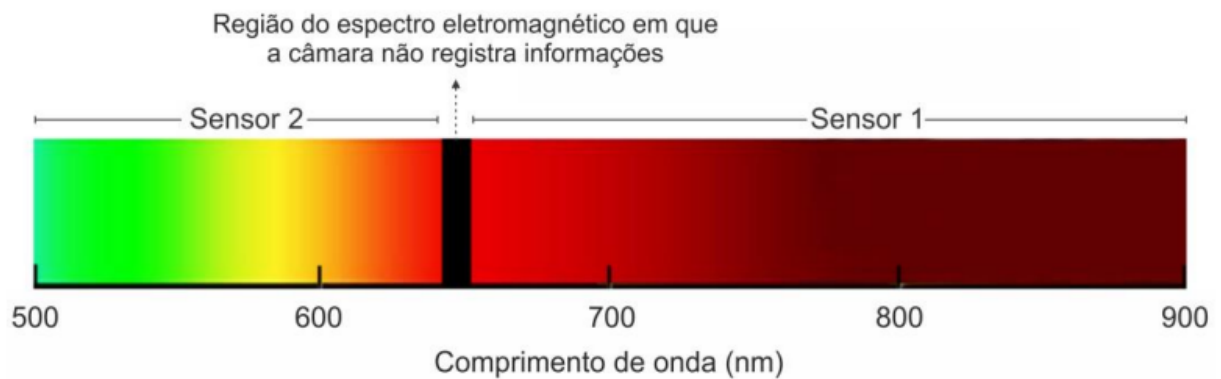
Modelo	Rikola FPI2015
Focal nominal	9 mm
Tamanho do pixel	5,5 x 5,5 μm
Dimensões da imagem	1017 x 648 pixel
Tipo do sensor	CMOS
Resolução espectral FWHM	10 nm
Faixa espectral	500-900 nm
Dimensões da câmara	75 x 89,5 x 151,5 mm
Peso	700g

A câmara FPI2015 da Rikola é operada por computador e pode exibir imagens em tempo real, além de salvá-las no próprio computador ou no cartão de memória Compact Flash no modo autônomo. É importante lembrar que o modo autônomo limita o número de bandas espectrais adquiridas, pois o tempo de transferência das imagens formadas do sensor para o cartão de memória é de aproximadamente 0,8 s para 24 bandas espectrais. No entanto, para o modelo FPI2015, é possível adquirir até 50 bandas espectrais em imagens de 1023 x 648 pixels.

No modo autônomo, é necessário configurar o tempo de atraso para iniciar a aquisição de imagens após a ativação da câmara, o intervalo de aquisição entre cubos de imagens sequenciais e o tempo de exposição de cada imagem.

Para cobrir o intervalo espectral de 500 a 900 nm, a câmara utiliza dois sensores CMOS (*Complementary Metal Oxide Semiconductor* -Semicondutor óxido metálico complementar) monocromáticos baseados no FPI que trabalham juntos. O primeiro sensor registra a energia correspondente aos comprimentos de onda a partir de 650 nm, enquanto o segundo sensor registra a energia dos comprimentos de onda até 636 nm. É importante observar que há uma lacuna radiométrica entre 637 nm e 649 nm, na qual a câmara não registra dados (MIYOSHI, 2016), como observado na Figura 5.

Figura 5 – Intervalo espectral dos sensores da câmara Rikola.



Fonte: Miyoshi (2016).

A câmara Rikola permite a aquisição de bandas espectrais em diferentes comprimentos de onda devido à existência do FPI. O usuário pode selecionar o espaçamento entre os espelhos do interferômetro, sendo que para cada espaçamento já são pré-definidos o comprimento de onda central da banda e sua resolução espectral em termos de FWHM.

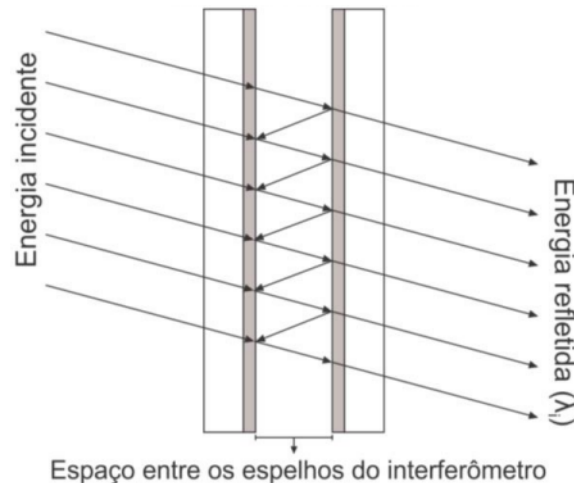
Entretanto, devido ao princípio de funcionamento do FPI e à existência de dois sensores CMOS, há um deslocamento relativo entre as bandas de uma mesma imagem (HONKAVAARA et al., 2013). Esse deslocamento é dependente da velocidade da plataforma onde a câmara é acoplada e da escala da imagem, tornando necessária a aplicação de uma transformação geométrica para o registro das bandas de uma mesma imagem. Estudos e resultados sobre o registro de imagens para a câmara baseada em FPI foram apresentados por Oliveira et al. (2014) e Honkavaara et al. (2013).

2.4.1 FPI

O sistema interno da câmara hiperspectral é baseada na tecnologia FPI *Fabry-Perot Interferometer* e atuadores piezoelétricos (SAARI et al., 2009). O FPI consiste em duas superfícies parcialmente refletoras, que são posicionadas paralelas uma à outra (Figura 6). Deste modo, os raios de luz que passam por essas superfícies sofrem múltiplas reflexões entre elas (OLIVEIRA et al., 2014). O comprimento do caminho óptico da cavidade formada entre as superfícies refletoras (a lacuna) determina os comprimentos de onda que

podem ser transmitidos através do interferômetro (MAKYNEN et al., 2011). Diferentes comprimentos de onda podem ser adquiridos modificando esta lacuna, tornando possível reconstruir o espectro para cada pixel na imagem. A lacuna de ar entre as placas do interferômetro é controlada por capacitivos e atuadores piezoelétricos (MAKYNEN, 2009) que são acoplados às placas do interferômetro e tem como função converter sinais elétricos em deslocamento.

Figura 6 – Princípio do Interferômetro de Fabry-Perot (FPI).



Fonte: Miyoshi (2016).

A luz emitida pelo objeto é filtrada por filtros *low-pass* e *high-pass* antes de chegar à lente. A lente tem como função focalizar a luz no módulo FPI, que por sua vez alcança o sensor RGB (vermelho/verde/azul) (SAARI et al., 2009; MAKYNNEN et al., 2011). O sensor é composto por um filtro mosaico Bayer com os três canais RGB, o que permite a aquisição simultânea de até três imagens correspondentes a diferentes comprimentos de onda. Através da variação da lacuna de ar, outras bandas podem ser geradas (OLIVEIRA et al., 2014).

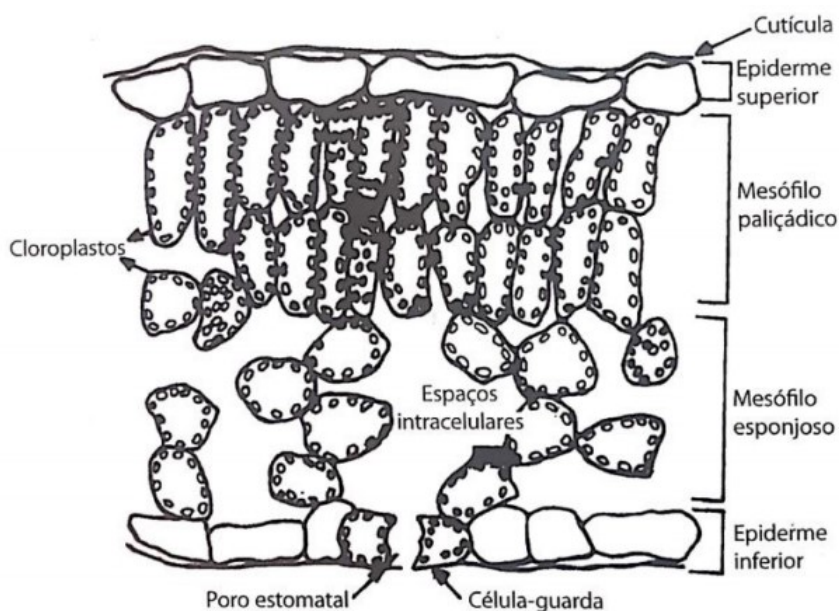
2.5 Interação da radiação eletromagnética com a vegetação

Segundo Jensen (2007), o fato da superfície terrestre ser coberta em sua maioria por variados tipos de vegetação contribuiu para o desenvolvimento de tecnologias capazes de entender informações biofísicas da vegetação a partir de dados oriundo de sensoriamento remoto. As imagens obtidas pelo sensoriamento remoto auxiliam no mapeamento de classes específicas da cobertura vegetal brasileira e estão em pleno desenvolvimento, podendo ser útil na previsão de safras e no suporte para preservar o meio ambiente (PONZONI; SHIMABUKURO; KUPLICH, 2012).

As folhas presentes na vegetação são o elemento mais importante para a interação com a radiação eletromagnética (REM), elas contribuem de forma considerável para o sinal que os sensores que operam na região espectro óptico detecta. De acordo com Colwell (1974), pode-se dizer que o comportamento espectral das folhas está relacionado com sua composição bioquímica (concentração de pigmentos fotossintetizantes) e por sua morfologia interna.

Na Figura 7, é apresentada a estrutura interna da planta, representada por um corte transversal em uma determinada folha. Ressalta-se que existem variações nas estruturas internas das plantas, dependendo de sua espécie. A radiação eletromagnética trafega por sua estrutura e dependendo do comprimento de onda dessa radiação, a interação se comporta de maneira diferente.

Figura 7 – Estrutura interna da planta.



Fonte: Ponzoni, Shimabukuro e Kuplich (2012).

A interação da radiação eletromagnética com os alvos presentes na superfície terrestre é baseada nos princípios da reflexão, transmissão e absorção, estes são dependentes de fatores como o comprimento de onda, ângulo em que a energia incidente atinge a superfície e a composição dos alvos que estão no terreno. Essas diferenças permitem que os diferentes alvos possam ser discriminados nas imagens obtidas pelos sensores.

A magnitude das porções da radiação eletromagnética absorvidas, transmitidas e refletidas pela planta faz com que seja possível distinguir diferentes espécies ou até mesmo eventos entre a mesma espécie. Abordando esta questão do ponto de vista fisiológico e bioquímico, a parte da energia que é absorvida é mais relevante. Cerca de 50% da REM que chega até a planta é absorvida pelos pigmentos contidos nas folhas. Essa energia absorvida é responsável por alguns fenômenos, como, por exemplo, a fotossíntese e

fotoconversão. Das principais funções da planta, a fotossíntese é a única ligada à radiação solar diretamente. Neste processo, é absorvida a radiação solar de comprimentos de onda entre 400 e 700 nm, em outras palavras, a energia utilizada pela planta corresponde à luz azul (400-530 nm) e à vermelha (600-730 nm), as quais são absorvidas pelos pigmentos de cloroplastos (MOREIRA, 2012).

Entre os elementos da planta, os principais responsáveis pela reflectância são: pigmentos, espaços ocupados pela água e pelo ar e estruturas celulares que possuem mesma dimensão do comprimento de onda da radiação incidente. As células da camada superficial são responsáveis pela reflectância de uma parte pequena da radiação incidente. A maior parte dessa energia é transmitida para o mesófilo esponjoso, em que os raios interagem com o meio e são refletidos.

2.6 Calibração Radiométrica

A calibração radiométrica é um processo fundamental no sensoriamento remoto para garantir que as imagens adquiridas sejam corretamente interpretadas e utilizadas em aplicações científicas e práticas. Segundo Barsi, Schott e Raqueno (2014), a calibração radiométrica é o processo de converter os valores de ND registrados pelos sensores em unidades de radiância espectral ou refletância. No processamento de dados de sensoriamento remoto, a calibração radiométrica é responsável por estabelecer uma relação entre a energia radiante registrada pelos sensores e as quantidades de radiação refletida ou emitida pelos objetos presentes na superfície analisada (PONZONI; KUPLICH; RUDORFF, 2007). Tal relação é estabelecida por meio da determinação dos modelos de resposta radiométrica dos sensores.

Existem dois tipos principais de calibração radiométrica: a absoluta e a relativa. A calibração radiométrica absoluta é realizada para determinar os coeficientes de calibração dos sensores, que relacionam as quantidades de energia radiante com as unidades de medida de interesse, como a reflectância ou a radiância. Esses coeficientes são obtidos por meio da comparação das medidas dos sensores com as medidas de fontes radiométricas conhecidas, como as lâmpadas de tungstênio ou os painéis difusores (HONKAVAARA et al., 2009).

Já a calibração radiométrica relativa é utilizada para corrigir as variações na resposta radiométrica dos sensores ao longo do tempo, decorrentes de mudanças na iluminação, no clima e nas condições atmosféricas. Esse tipo de calibração envolve a utilização de imagens de referência, que são coletadas em condições ideais e servem como padrões para a correção das imagens subsequentes (HONKAVAARA et al., 2009).

O processo de calibração de sensores é fundamental para garantir a precisão e a qualidade dos dados obtidos por sensores remotos. Essa calibração pode ser realizada em laboratório, antes da colocação do sensor na plataforma, ou durante o voo.

A calibração em laboratório pode ser realizada utilizando esferas ou hemisferas de integração, que são capazes de determinar os coeficientes de calibração do sensor. Além disso, instrumentos como colimadores podem ser utilizados para determinar a radiância através de um espectrorradiômetro calibrado (DINGUIRARD; SLATER, 1999). É importante destacar que as calibrações realizadas durante o voo são essenciais para a correção de variações temporais nas condições de iluminação e atmosféricas.

A calibração em voo pode utilizar lâmpadas de calibração ou luz natural proveniente do Sol. Já na calibração radiométrica "vicarious", são utilizados alvos de referência posicionados na superfície terrestre ou objetos da própria cena, caracterizados espectralmente para aferir a sensibilidade dos detectores.

Existem métodos baseados na radiância e reflectância, sendo que este último pode ser corrigido atmosféricamente com base em um modelo de transferência radiativa. A caracterização espectral de alvos deve ser realizada simultaneamente à passagem do sensor sobre ele, para comparação com os dados obtidos pelo sensor. Algumas campanhas de calibração de sensores envolvem alvos em tons de cinza, placas coloridas e estrelas de Siemens para avaliar a resolução espacial (SLATER; HOUGH, 1987; TEILLET; GUINDON; GOODENOUGH, 1990; THOME; BIGGAR; SLATER, 2001; PONZONI; KUPLICH; RUDORFF, 2007; DINGUIRARD; SLATER, 1999; HONKAVAARA et al., 2009).

Smith e Milton (1999) apresentaram um método chamado *Empirical Line Method* (ELM) para calibrar dados obtidos por sensores remotos e converter os valores digitais desses dados em valores de reflectância. É uma técnica que utiliza um conjunto de dados adquiridos por um sensor remoto em um ambiente que é considerado ter uma reflectância conhecida, como uma área de referência terrestre, para calibrar dados obtidos em outras áreas. O método envolve a criação de uma linha empírica que relaciona os valores digitais do sensor com a reflectância do terreno, permitindo assim que os dados obtidos em outras áreas sejam convertidos em valores de reflectância. O trabalho descreve o processo de aplicação do ELM em dados obtidos pelo sensor *Thematic Mapper* (TM), que é o sensor multiespectral embarcado no satélite Landsat. Os resultados mostraram que o ELM é um método eficaz para a calibração de dados de sensoriamento remoto, produzindo valores de reflectância comparáveis aos obtidos por outras técnicas de calibração.

Em Sahoo et al. (2013), é abordada a calibração de dados hiperespectrais utilizando o método de calibração pelo ELM. Este método é um dos mais comuns de calibração radiométrica para dados hiperespectrais e é baseado na relação empírica entre a reflectância de um objeto em uma cena e a radiância medida pelo sensor. O ELM consiste em selecionar alvos de referência na cena que possuam reflectância conhecida e medir sua radiância correspondente com o sensor. Em seguida, o método utiliza esses valores medidos para gerar uma curva de calibração linear, que é utilizada para converter as radiâncias medidas pelo sensor em reflectância espectral. O ELM é uma técnica bastante utilizada para a

correção atmosférica de dados de sensores multiespectrais, como o do Landsat TM, mas pode ser aplicado também a dados hiperespectrais, como demonstrado nos estudos de Miyoshi (2016) e Moriya (2015).

2.7 Aprendizado de máquina na agricultura de precisão

O aprendizado de máquina é um campo científico que se beneficia da capacidade dos computadores de analisar grandes volumes de dados sem a necessidade de instruções de programação diretas sobre as entradas de dados para gerar dados de saída (LIAKOS et al., 2018). O aprendizado de máquina teve início com a tecnologia *big data*, que possibilitou a compreensão de grandes quantidades de dados provenientes de atividades agrícolas (LIAKOS et al., 2018). As técnicas de aprendizado de máquina aplicadas aos dados fotogramétricos e de sensoriamento remoto facilitam a compreensão de grandes áreas agrícolas de uma forma não destrutiva em comparação às características *in situ*, com conhecimentos científicos sobre os atributos das plantas, solo e ambiente (KAMILARIS; PRENAFETA-BOLDÚ, 2018).

As imagens adquiridas por sensoriamento remoto e fotogrametria podem ser utilizadas para a extração de informações específicas sobre a cobertura da terra. Entre as técnicas populares para essa finalidade, encontra-se a classificação. Nesse procedimento, para cada pixel da imagem é atribuído um rótulo de acordo com suas características espectrais e/ou espaciais (NOVO, 2008).

A aplicação de estratégias de aprendizado de máquina pode gerar uma estimativa mais precisa, ponto a ponto, das variações dos rendimentos, aliando elementos de tecnologia às características do ambiente (CAMPOS et al., 2016). Um dos fatores mais importantes para a determinação do rendimento das culturas é a sua demanda e disponibilidade hídrica, já que influencia em diversos processos metabólicos do desenvolvimento da planta (MLADENOVA et al., 2017). Assim, métodos para estimar a relação entre a aplicação de água e a produtividade das plantas são oportunas, principalmente devido à grande quantidade de dados obtidos por fotogrametria e sensoriamento remoto, e à utilização do aprendizado de máquina para analisá-los.

2.7.1 Random Forest

Uma árvore de decisão é um método de aprendizado de máquina supervisionado não-paramétrico, muito utilizado em atividades de classificação e regressão. Algoritmos de aprendizagem supervisionada baseados em árvores de decisão são comumente utilizados em problemas de classificação.

O interesse por algoritmos de aprendizagem de máquina, como o *Random Forest* (RF) ou Floresta Aleatória, vem crescendo por se tratar de métodos precisos e robustos (BREIMAN, 1996; DIETTERICH, 2000). O *Random Forest* é um algoritmo que usa um conjunto de *Classification and Regression Trees* (CART) que possibilita fazer uma predição por regressão ou classificação (BREIMAN, 2001).

O algoritmo de aprendizado de máquina RF é constituído por um conjunto de árvores de decisão. Árvore de decisão é um tipo de estrutura de dados que, a partir de uma amostra de treinamento, efetua a classificação e a predição de dados (SILVA, 2005). O conjunto de treinamento utilizado para a aprendizagem do algoritmo é formado por entradas e saídas. Cada entrada dispõe de um conjunto de características que definem a saída (classe), de modo que a partição do espaço de atributos seja dada da melhor forma possível.

De acordo com Silva (2005), a capacidade de discriminação de uma árvore de decisão depende da eficiência em dividir o espaço definido pelos atributos em sub-espacos, sendo estes associados a uma classe no final do processo. Uma das vantagens em se utilizar as árvores de decisão consiste na possibilidade de avaliar atributos de diferentes fontes, podendo essas serem: espectral, espacial, geométrica e derivadas.

A estrutura hierárquica da árvore de decisão é composta por nós, ramos e folhas. Nos nós, são tomadas as decisões, ou seja, onde uma condição é avaliada com base em um determinado atributo. A escolha do melhor atributo, ou seja, aquele que maximiza uma função ganho, é feita a partir do valor de entropia ou índice de Gini.

Os ramos são os possíveis caminhos tomados a partir da decisão, podendo estes se degenerarem em um novo nó, caso a profundidade da árvore continue, ou em uma folha, sendo essa a classe na qual o elemento foi classificado. Em aplicações de Sensoriamento Remoto, o elemento referido anteriormente é o pixel, e as folhas normalmente são tipos de cobertura do solo (água, mata, área urbana, solo exposto, entre outros).

O RF é um algoritmo de aprendizado de máquina que requer a definição de dois parâmetros principais: o número de árvores que compõem a floresta (*Ntree*) e o número de atributos a serem avaliados em cada nó para construir cada uma das árvores da floresta (*Mtry*) (BELGIU; DRĂGUȚ, 2016). Em outras palavras, o *Ntree* define o número de árvores que serão criadas e combinadas para formar a floresta, enquanto o *Mtry* determina quantos atributos serão selecionados aleatoriamente em cada nó para decidir qual será a melhor divisão do nó em dois ramos. É importante ajustar esses parâmetros adequadamente para alcançar uma boa precisão na classificação dos dados. Segundo Breiman (2001), o algoritmo cria árvores bem variadas e não tendenciosas.

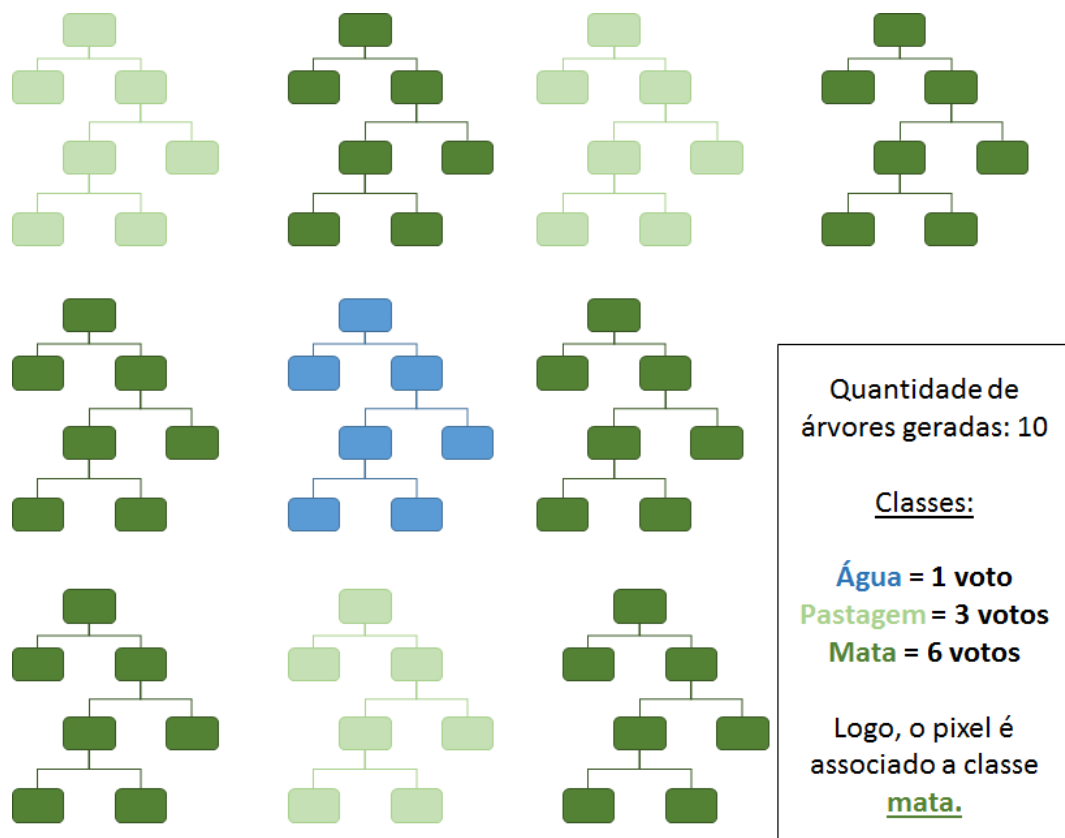
Depois que o algoritmo é treinado, de modo similar ao treinamento de uma árvore de decisão conforme apresentado anteriormente, cada um dos pixels da imagem passa pelo

processo de classificação. Cada pixel a ser classificado passa por uma árvore de decisão da floresta formada, o resultado gerado a partir dessa árvore é armazenado, como se fosse um voto, sendo esse processo repetido para as demais. Ao final do processo, é feita a contagem dos votos (*bagging*), e a classe mais votada é associada ao pixel.

O parâmetro *Mtry* é responsável por definir o número de atributos que serão testados para separar o espaço de atributos da melhor forma possível. Em relação aos parâmetros do algoritmo, estudos mostraram que a acurácia da classificação é menos sensível ao parâmetro *Ntree* do que ao parâmetro *Mtry* (BELGIU; DRĂGUT, 2016).

Segundo Lawrence, Wood e Sheley (2006 apud BELGIU; DRĂGUT, 2016), alguns estudos consideram o parâmetro *Ntree* igual a 500, visto que a partir dessa quantidade de árvores a acurácia da classificação é estabilizada. O parâmetro *Mtry* geralmente é adotado como a raiz quadrada da dimensão do vetor de atributos, ou seja, da quantidade de características que definem as entradas (GISLASON; BENEDIKTSSON; SVEINSSON, 2006 apud BELGIU; DRĂGUT, 2016).

Figura 8 – Exemplo do funcionamento do algoritmo *Random Forest*.



Fonte: A autora (2021).

Digno de nota que o parâmetro *Ntree* não influencia na performance computacional, enquanto que o tempo computacional exigido para o processamento é diretamente proporcional ao parâmetro *Mtry*, visto que a informação de ganho deve ser calculada para cada

um dos atributos, para que em seguida seja selecionado o que apresenta melhor resultado (BELGIU; DRĂGUȚ, 2016).

O algoritmo RF é um dos métodos de aprendizado de máquina mais utilizados em aplicações na agricultura de precisão. Os principais motivos para sua ampla utilização é sua habilidade em lidar com dados de alta dimensionalidade e com múltiplas variáveis que podem influenciar o resultado final de uma análise. Além disso, o RF é capaz de fornecer uma estimativa da importância relativa das variáveis na análise, o que pode auxiliar no entendimento dos fatores que mais influenciam o resultado final.

O critério para a decisão de fazer as divisões dos nós afeta a precisão de uma árvore. Existem diversos algoritmos para dividir um nó em dois ou em mais sub-nós, o Índice de Gini e a Entropia são os algoritmos de partição mais usados na árvore de decisão utilizadas para classificação. O critério para realizar as partições é o da utilidade do atributo para a classificação.

O índice de Gini mede o grau de heterogeneidade dos dados utilizados, sendo assim este índice pode ser usado para medir a impureza de um nó (ONODA, 2001) e pode ser determinado pela Equação 2.1 .

$$IG = 1 - \sum_{i=1}^c p_i^2 \quad (2.1)$$

Em que: p_i é a frequência de cada classe em cada nó e c é o número de classes.

O índice de Gini é analisado de forma que se o resultado obtido for zero, o nó é considerado puro. E quando o resultado se aproxima de um, o nó é impuro.

A entropia é normalmente usada na teoria da informação para medir a pureza ou a impureza de um determinado conjunto, ou seja, em um conjunto de dados, é uma medida da falta de homogeneidade dos dados de entrada em relação a sua classificação. Diz-se que o conjunto de dados é heterogêneo quando a entropia é máxima (igual a 1).

De acordo com Belgiu e Drăguț (2016), o classificador RF pode ser usado com sucesso para selecionar e classificar as variáveis com maior capacidade de discriminação entre as classes. Esse é um recurso importante, considerando que a alta dimensionalidade dos dados obtidos por sensoriamento remoto torna a seleção das variáveis mais relevantes uma tarefa demorada, propensa a erros e subjetiva (BELGIU; DRĂGUȚ; STROBL, 2014). O algoritmo RF tem sido utilizado com sucesso para mapear classes de cobertura terrestre (COLDITZ, 2015; HAAS; BAN, 2014), tem sido aplicado na classificação de níveis de desfolhamento de insetos usando a banda do *red edge* de imagens do RapidEye (ADELABU; MUTANGA; ADAM, 2014). Além disso, também tem sido usado para classificação de biomassa usando dados temporais do Landsat (FRAZIER et al., 2014) e para identificar a saúde das árvores usando dados do sensor IKONOS (WANG et al., 2015).

O algoritmo RF tem se mostrado eficiente também ao ser utilizado em dados hiperespectrais, como no mapeamento de espécies de árvores (CAVALLARO et al., 2015; FASSNACHT et al., 2014), plantas invasoras (LAWRENCE; WOOD; SHELEY, 2006; PEERBHAY; MUTANGA; ISMAIL, 2015), espécies de gramíneas (MANSOUR et al., 2012) e níveis de desfolhamento de insetos em áreas de mopane (ADELABU; MUTANGA; ADAM, 2014).

2.7.2 Support Vector Machine

O classificador supervisionado *Support Vector Machine* (SVM) define um hiperplano na discriminação entre duas classes (VAPNIK, 1998). Este classificador foi introduzido por Cortes e Vapnik (1995), em que os vetores de entrada são transformados para um espaço com dimensão superior através de uma transformação linear e, então um hiperplano ótimo que classifica (separa) os dados pode ser encontrado. Existem inúmeros hiperplanos capazes de dividir os dados em duas classes, porém o objetivo é estabelecer um que apresente a maior separação entre as margens das classes, sendo esse delimitador chamado de hiperplano otimizado. Assim, os erros na aplicação de rótulos aos dados são minimizados.

Sendo um classificador não paramétrico, na regra de decisão para a atribuição de rótulos aos dados, o SVM, não utiliza atributos estatísticos sobre o conjunto de treinamento, como a média e o desvio padrão dos valores de NDs dos pixels que pertencem a alguma classe de uso e cobertura da terra. Deste modo, a quantidade de elementos que são utilizados no treinamento não são de extrema relevância, mas toda a variabilidade de um tipo específico de classe que esteja inserida no conjunto (FOODY; MATHUR, 2004). Uma das principais vantagens do SVM é sua capacidade de lidar com conjuntos de dados complexos e com alta dimensionalidade, o que o torna adequado para análises de imagens e dados de sensoriamento remoto.

No SVM, as funções utilizadas para construir o modelo são chamadas de *kernels*. Existem vários tipos de *kernels* disponíveis, incluindo *Radial Basis Function* (RBF), linear, polinomial e sigmoide. O RBF é comumente utilizado para dados que não possuem uma separação linear clara, pois é capaz de transformar os dados de entrada em um espaço de alta dimensão, onde é mais fácil separá-los em classes. Já o *kernel* linear é mais adequado para dados que podem ser separados linearmente e é mais rápido de calcular do que outros tipos de *kernels*. O *kernel* polinomial é adequado para dados que não são linearmente separáveis, esses possuem uma relação não linear entre as características, permitindo mapeá-los para um espaço de alta dimensão e separar as classes com um hiperplano polinomial. Por fim, o *kernel* sigmoide é recomendado para dados que não são linearmente separáveis e possuem uma relação não linear entre as características, semelhante ao *kernel* polinomial, mas usando uma função de ativação sigmoide em vez de uma função polinomial.

As imagens adquiridas por sensoriamento remoto e fotogrametria podem ser utilizadas para a extração de informações específicas sobre a cobertura da terra. Entre as técnicas populares para essa finalidade, encontra-se a classificação. Nesse procedimento, para cada pixel da imagem, é atribuído um rótulo de acordo com suas características espectrais e/ou espaciais (NOVO, 2008).

Em Wang et al. (2017), o SVM foi utilizado para prever a produtividade da soja a partir de dados hiperespectrais, com resultados promissores. Outro estudo realizado por Barros et al. (2018) também utilizou o SVM para prever produtividade da soja a partir de dados de sensoriamento remoto, demonstrando sua eficácia em relação a outros algoritmos.

Além disso, o SVM tem sido utilizado para outras aplicações na área de agricultura de precisão, como a classificação de plantas daninhas e doenças (WANG et al., 2018), o mapeamento de solos (CAI et al., 2021) e a previsão de propriedades do solo (YI et al., 2018).

De maneira geral, este algoritmo é eficaz e versátil e tem sido amplamente utilizado na agricultura de precisão, sobretudo em aplicações com dados provenientes do sensoriamento remoto. É uma ferramenta importante quando se trata de dados complexos e de alta dimensionalidade.

2.7.3 K-Nearest Neighbors

O *K-Nearest Neighbors* (KNN) é um algoritmo de aprendizado de máquina supervisionado que tem como principal objetivo classificar um objeto baseado nas classes dos k objetos mais próximos a ele, seguindo uma medida de distância definida (ALTMAN, 1992). O valor de k é selecionado de maneira a maximizar a acurácia do modelo e pode variar de acordo com o conjunto de dados em questão (BERRAR, 2006). Este algoritmo é um dos mais simples e populares de aprendizado de máquina para classificação e regressão. Seu funcionamento baseia-se em encontrar os k exemplos mais próximos (vizinhos) de uma amostra de teste e usar a classe ou valor médio desses vizinhos para prever a classe ou valor da amostra de teste.

Uma das principais vantagens do KNN é sua maneira simples de funcionamento e fácil interpretação, tornando-o um algoritmo popular em estudos de classificação (BERRAR, 2006). Porém, o desempenho do KNN pode ser afetado pela escolha da medida de distância utilizada e pela escala dos atributos (ALTMAN, 1992). Além disso, este algoritmo pode ser considerado computacionalmente caro para conjuntos de dados grandes, já que é necessário calcular a distância entre o objeto em questão e todos os outros objetos do conjunto de dados (BERRAR, 2006). Apesar das limitações, o KNN continua sendo uma técnica de classificação popular e eficiente, especialmente em conjuntos de dados pequenos e com poucas classes (ALTMAN, 1992).

Segundo Altman (1992), o KNN tem sido aplicado com sucesso em diversas áreas do conhecimento, como na medicina, na análise de imagens e na previsão de dados econômicos. Em aplicações agrícolas, este algoritmo tem sido aplicado para classificar diferentes culturas de acordo com suas características físicas, como em Arribas et al. (2014), em que foi utilizado o KNN para classificar diferentes variedades de uva de acordo com suas características de coloração. Além disso, muitas pesquisas na área de agricultura de precisão têm utilizado o uso do KNN para prever produtividade da soja e outras variáveis de interesse (GARG et al., 2021; SHARMA et al., 2020; KAMIR; WALDNER; HOCHMAN, 2020; ZHANG et al., 2020).

Em Han et al. (2013), os autores propuseram um modelo de previsão de produtividade da soja baseado em dados de sensoriamento remoto e utilizando o KNN como algoritmo de classificação. Os resultados mostraram que o modelo proposto apresentou desempenho satisfatório na previsão da produtividade da soja. Já em Wang et al. (2017), os autores avaliaram o desempenho de diferentes algoritmos de aprendizado de máquina, incluindo o KNN, na previsão da produtividade da soja e utilizando imagens hiperespectrais. Os resultados obtidos mostraram que o KNN apresentou desempenho inferior em relação aos demais algoritmos avaliados.

De maneira geral, as pesquisas mostram que o KNN tem sido amplamente utilizado na agricultura de precisão, tanto para prever a produtividade da soja quanto para estudar outras variáveis de interesse. Embora esse algoritmo apresente resultados satisfatórios em alguns casos, em outros o KNN foi superado por outros algoritmos de aprendizado de máquina, o que indica a importância de avaliar diferentes algoritmos e técnicas. Também é importante destacar que o desempenho do KNN pode ser influenciado por diversos fatores, como escolha do número de vizinhos, a métrica de distância utilizada, entre outros. Deste modo, é fundamental realizar uma análise dos dados e ajustar os parâmetros do algoritmo de acordo com as características do problema em questão.

3 Material e métodos

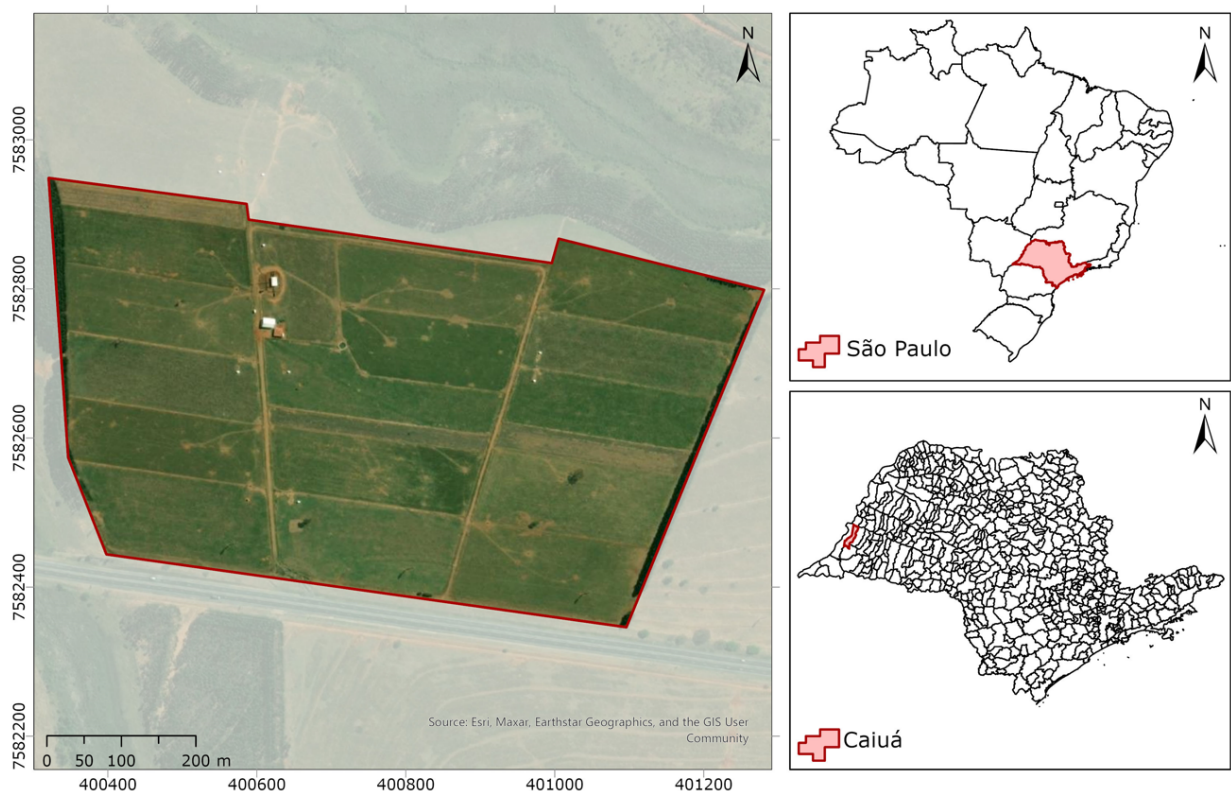
Nesta Seção, serão apresentados os principais materiais necessários ao desenvolvimento da pesquisa. Na subseção 3.1.1, será apresentada a área de estudo e os dados. Os recursos de software e equipamentos utilizados para a realização do trabalho serão apresentados nas subseção 3.1.3. Na Subseção 3.2, será apresentado o método proposto para a execução da pesquisa.

3.1 Materiais

3.1.1 Área de estudo e dados

A área selecionada para o desenvolvimento desta pesquisa, localiza-se na fazenda experimental Facholi (Grupo Facholi), nessa área há dois tipos de culturas agrícolas: soja e capim forrageiro. Essa fazenda está localizada no município de Caiuá, no estado de São Paulo, e possui uma área aproximada de 417.000 m², conforme Figura 9. O sistema de cultivo nessa área foi implantado em 2015, com rotação de culturas e pastoreio anual.

Figura 9 – Área de estudo localizada no Município de Caiuá.



Fonte: A autora (2023).

A área experimental da fazenda é dividida em três blocos (A, B e C), como apresentado na Figura 10. Os blocos ainda são subdivididos em parcelas com dimensões aproximadas de 200x100 m. No momento da tomada das imagens aéreas da área de estudo, haviam cultivos de capim e soja. No entanto, para o o propósito deste estudo, foram consideradas apenas as parcelas com cultivo de soja.

Figura 10 – Divisão dos blocos da fazenda experimental.



Fonte: A autora (2023).

Para este trabalho, foram considerados dois grupos de parcelas, um grupo para a análise das respostas espectrais, a partir de imagens adquiridas pela missão Sentinel-2 e outro grupo para extrair informações espectrais para o treinamento, validação e classificação, utilizando algoritmos de aprendizado de máquina.

Na Figura 11, é apresentado o primeiro grupo com seis parcelas (A1, A6, B5, B6, C2 e C4), utilizados para analisar as respostas espectrais da soja parasitada com nematoides, utilizando as imagens adquiridas pelo sensor MSI. Para a seleção desse grupo, optou-se apenas pelas parcelas em que as plantas de soja estivessem com cobertura vegetal densa, de modo que o solo entre as linhas de plantio não fosse capturado pelas imagens e houvesse menos influência deste alvo nas respostas espectrais. Ressalta-se que a escolha das parcelas para a análise das respostas espectrais deve ser cuidadosa, pois deve ser considerada a capacidade da resolução espacial do sensor, que é de 20 metros.

Figura 11 – Grupo de parcelas utilizadas para a análise das respostas espectrais a partir de imagens da missão Sentinel-2.



Fonte: A autora (2023).

Na Figura 12, é apresentado o segundo grupo de parcelas, contendo seis áreas (A1, A3, A6, B3, B5 e B6). Estas áreas foram utilizadas na etapa da metodologia em que são extraídas informações do conjunto de imagens hiperespectrais de alta resolução espacial disponíveis para o treinamento, validação e classificação por meio de algoritmos de aprendizado de máquina. Para esta etapa, foram utilizadas imagens capturadas pela câmera Rikola que possui alta resolução espacial.

Figura 12 – Grupo de parcelas utilizadas para extrair informações para treinamento dos algoritmos.



Fonte: A autora (2023).

Para o desenvolvimento desta pesquisa, foram utilizadas três conjuntos de imagens, com dez bandas, coletadas em três datas distintas pela missão Sentinel-2, com resolução espacial de 20 metros. Essas imagens foram utilizadas para analisar as respostas espectrais das plantas de soja parasitadas por nematoides em períodos fenológicos diferentes, com o objetivo de investigar qual o estágio fenológico apresenta maior diferença espectral entre plantas parasitadas e não parasitadas por fitonematoides. Assim, foram utilizados conjuntos de imagens capturadas nos dias 26 de dezembro de 2021, 22 de janeiro de 2022 e 16 de fevereiro de 2022. Além disso, foram utilizadas vinte bandas da câmera Rikola variando entre 500 nm e 900 nm coletadas no dia 01 de fevereiro de 2022.

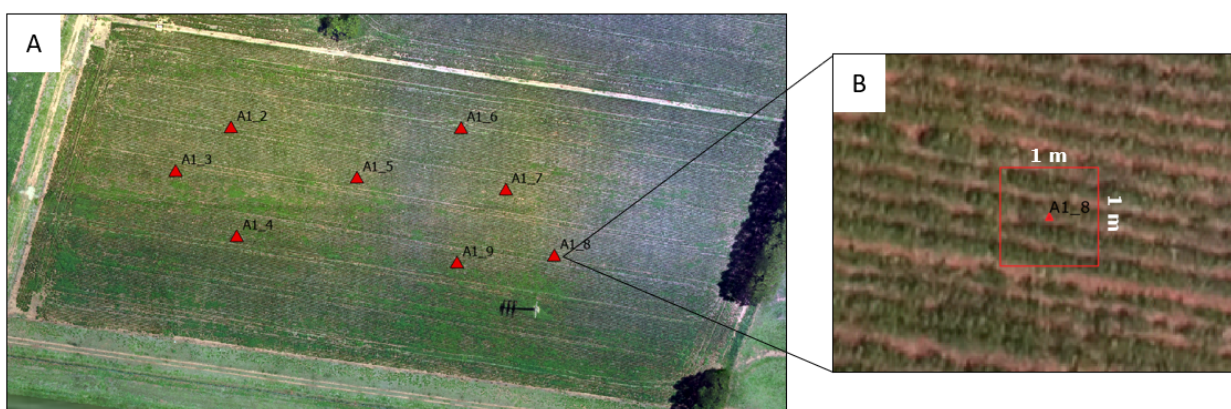
3.1.2 Dados de Referência

A coleta e análise das amostras nematológicas foram realizadas no laboratório de biologia do solo da Universidade do Oeste Paulista (UNOESTE). A coleta ocorreu em 19 de janeiro de 2022, quando observou em campo que a soja estava na fase reprodutiva final e foram utilizadas $n = 39$ amostras no total.

Os pontos georreferenciados selecionados para análise nematológica foram distribuí-

dos pelas parcelas buscando seguir uma distribuição em zigue e zague, conforme mostra a Figura 15. Para a coleta do solo, foi estabelecida uma área de 1 m^2 no ponto selecionado, conforme apresentado na Figura 13 (B). Dessa forma, foi coletado o solo presente na raiz de 1×1 metro. O solo de cada amostra foi homogeneizado antes da extração dos nematoides, utilizando uma alíquota de 100 cm^3 ou 100 mL de solo. A extração foi realizada por meio da metodologia de flotação centrífuga em solução de sacarose, conforme descrito por Jenkins et al. (1964). Os nematoides encontrados nessa área foram quantificados em laboratório e o número de indivíduos, por 100 cm^3 de solo ou por 100 mL , foi registrado.

Figura 13 – (A) Distribuições dos pontos de amostras nematológicas de uma parcela e (B) área considerada por ponto de amostra.

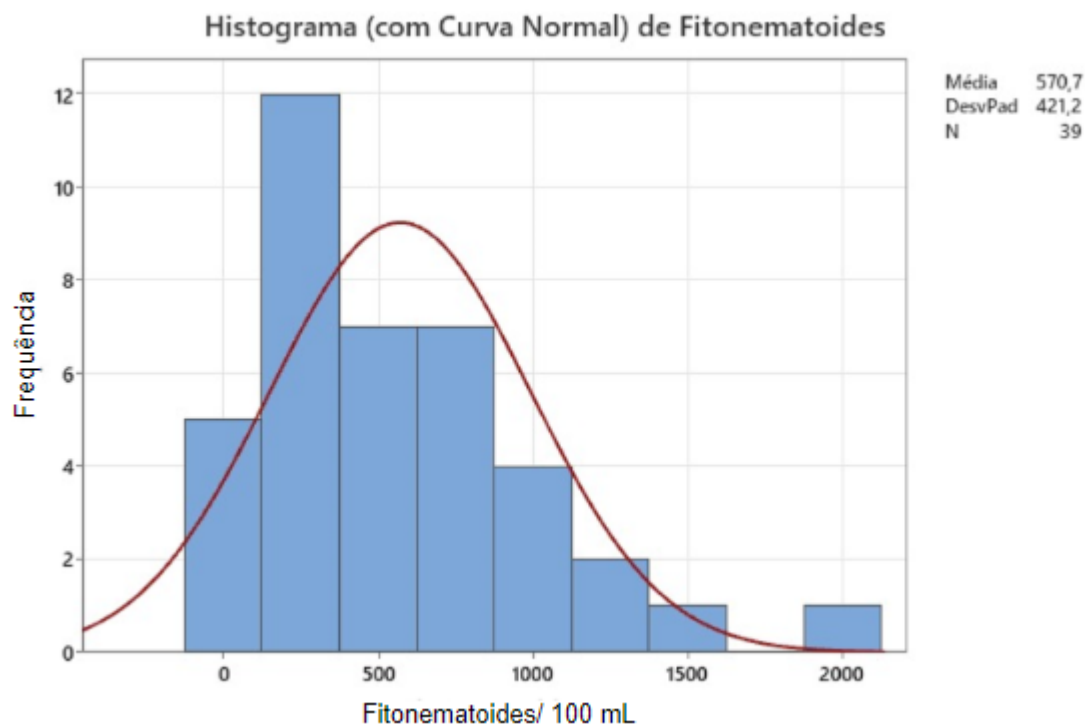


Fonte: A autora (2023).

Os nematoides foram separados e contabilizados por grupos e gêneros, sendo que, na área de estudo, foram encontrados dois grupos: fitonematoides e vida livre. Dentre os fitonematoides, foram contabilizados os seguintes gêneros: *Pratylenchus*, *Meloidogyne*, *Helicotylenchus* e *Criconemella*. O gênero de fitonematoide mais comum foi o *Helicotylenchus*, presente em todas as amostras da área. Também foram contabilizados os nematoides da espécie vida livre, encontrados em todas as áreas analisadas.

Para a soma dos fitonematoides contabilizados nas amostras coletadas, foram calculadas as estatísticas descritivas utilizando o software Minitab. A média e o desvio padrão foram calculados e um histograma de frequência com a curva normal foi gerado, conforme apresentado na Figura 14. A média calculada é de $570,70/100\text{ mL}$, o desvio padrão é de $421,20/100\text{ mL}$ e o valor mediano é $468/100\text{ mL}$.

Figura 14 – Estatísticas descritivas das amostras utilizadas como referência.



Fonte: A autora (2023).

De acordo com o histograma de frequência apresentado na Figura 14, é possível observar que a maioria das áreas analisadas apresenta uma concentração média de aproximadamente 250 fitonematoides/100 mL de solo, representando 30,80% das amostras totais. A frequência de amostras sem presença de fitonematoides foi de 12,80%, enquanto que amostras com mais infestação representaram apenas 2,60% do total.

Observa-se que há diferentes quantidades de fitonematoides distribuídas em torno da média (570,70/100 cm^3), com um desvio-padrão de 421,20/100 mL, indicando a existência de variabilidade nas amostras. Portanto, o nível de infestação por fitonematoides nas áreas de estudo não é homogêneo.

3.1.3 Equipamentos e aplicativos

No laboratório de fotogrametria e sensorimento remoto da FCT/UNESP, estão disponíveis os recursos de equipamentos, *hardware* e software, como:

- Microcomputadores e componentes (Processo FAPESP nº 2013/13138-0): para realização das etapas propostas e de experimento;
- Câmara hiperespectral Rikola com interferômetro Fabry-Perot (FPI);
- Receptores GNSS Topcon Hiper (L1/L2);

- Espectrorradiômetro FieldSpec HandHeld (325-1075 nm);
- VANT X800 GEO;
- DJI Phantom 4;
- Software Agisoft Metashape para processamento das imagens;
- Software ENVI;
- Software HyperSpectral Imager da Rikola;
- Software CoRegister da Rikola;
- Google Colaboratory (Colab);
- Biblioteca *scikit-learn*;
- Software estatístico Minitab;
- Software Quantum GIS (QGIS): para manipulação de arquivos e visualização;
- Software ArcGIS Pro para representação final dos resultados obtidos;

3.2 Metodologia

3.2.1 Avaliação das respostas espectrais na cultura da soja com diferentes níveis de infestação por fitonematoides nos estádios fenológicos

Foi realizado um estudo prévio com a intenção de avaliar as respostas espectrais da planta de soja parasitada com fitonematoides em diferentes estádios fenológicos. Para isso, foram utilizados três conjuntos de imagens coletadas pela missão Sentinel-2 em períodos fenológicos diferentes. A finalidade foi verificar a resposta espectral em diferentes estádios da soja visando relacionar com a intensidade do parasitismo por nematoides.

As imagens multiespectrais capturadas pelo sensor MSI foram obtidas gratuitamente pela plataforma *Copernicus Open Access Hub* (<<https://scihub.copernicus.eu/dhus/#/home>>), administrada pela *European Space Agency* (ESA). As imagens foram adquiridas no nível de correção 2A, em que os pixels são expressos em reflectância de superfície e com 20 metros de resolução espacial. Foi optado utilizar as imagens com 20 metros de resolução espacial, devido a disponibilidade de bandas na região do *Red Edge*.

Foram analisados três conjuntos de imagens em diferentes estádios fenológicos da soja. O primeiro conjunto foi coletado em 26 de dezembro de 2021, setenta e um dias

após o plantio, quando a soja estava estimada entre os estádios V4 e V6. O segundo conjunto foi coletado em 22 de janeiro de 2022, noventa e oito dias após o plantio, quando se estimou que a soja estava nos estádios fenológicos R5 e R7. Já o terceiro conjunto foi coletado em 16 de fevereiro de 2022, cento e vinte e três dias após o plantio, quando se estimou que a soja já estava no estágio fenológico R8.

Na Figura 15, é apresentada a distribuição dos 38 pontos das amostras nematológicas que foram utilizados. As respostas espectrais dos pontos das amostras nematológicas georreferenciadas foram extraídas para as três datas dos conjuntos de imagem do sensor MSI. A partir dessas informações, foi possível gerar gráficos para análises das respostas espectrais nas respectivas datas.

Figura 15 – Distribuição dos pontos das amostras nematológicas utilizadas para análise da resposta espectral.



Fonte: A autora (2023).

Foram coletadas as respostas espectrais dos pixels de cada banda para cada uma das 38 amostras georreferenciadas. Devido à resolução espacial do sensor MSI, que é de 20 metros, apenas uma resposta espectral foi coletada para um único pixel correspondente a cada amostra. Para simplificar a análise das respostas espectrais das plantas de soja parasitadas com fitonematoides, os pontos das amostras nematológicas foram divididos em quatro classes (C1, C2, C3 e C4) de densidade de população de nematoides por amostra, onde C1 representa a classe de menor infestação por fitonematoides e C4 a classe com

maior infestação. Foi calculada a média dos valores de cada classe e para cada banda, gerando um gráfico para representar as classes em cada período fenológico. A análise visual desses gráficos permite identificar o estágio fenológico em que os espectros das classes apresentam maior diferença.

3.2.2 Levantamento Geodésico

Foi realizado um levantamento geodésico para obter as coordenadas dos pontos de controle e de verificação utilizando receptores GNSS Topcon Hiper (L1/L2). Esses pontos foram empregados durante o processamento das imagens obtidas tanto pela Rikola quanto pelo VANT DJI Phantom 4. A Figura 16 mostra a distribuição desses pontos na área de estudo.

Esses pontos, cujas coordenadas são conhecidas, são usados como referência durante o processamento das imagens para corrigir possíveis distorções e imprecisões, garantindo a acurácia dos dados gerados. Além disso, foi realizado um segundo levantamento geodésico com o receptor RTK (*Real Time Kinematic*) Topcon para coletar as coordenadas de pontos onde foram coletadas as amostras nematológicas das 3 parcelas pertencentes ao bloco A (Figura 12).

Figura 16 – Distribuição dos pontos de controle e de verificação para fototriangulação das imagens.



Fonte: A autora (2023).

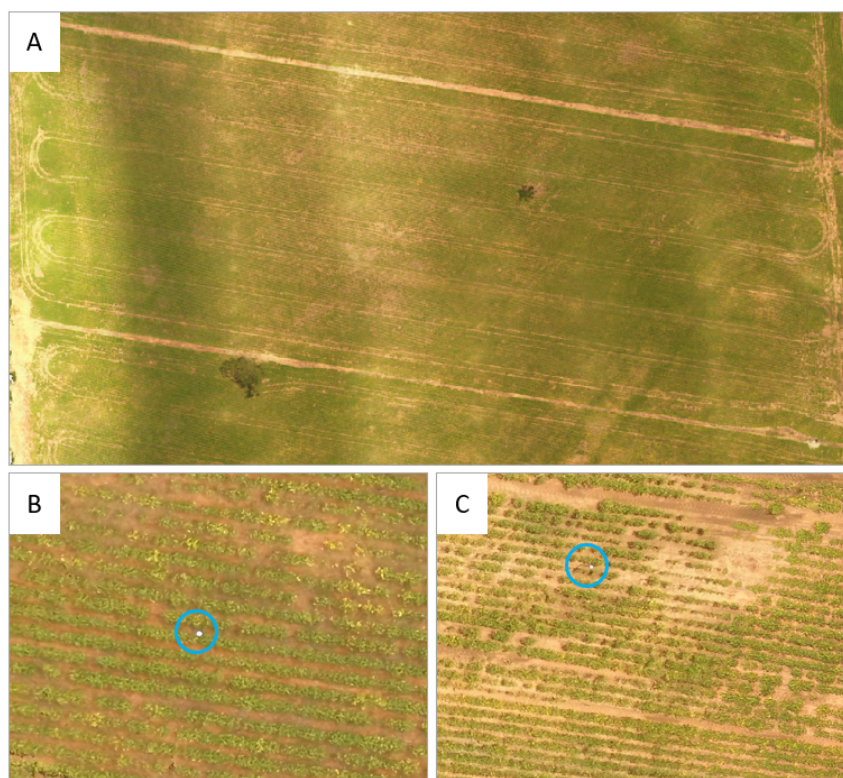
3.2.3 Levantamento aéreo e processamento de imagens digitais

Foram realizados dois levantamentos aéreos com diferentes datas e sensores. O primeiro levantamento foi realizado em 15 de janeiro de 2022, utilizando o VANT DJI Phantom 4, que coletou 464 imagens dos três blocos da fazenda experimental. A câmera RGB de 20 MP do DJI Phantom 4 possui três bandas no espectro do visível (vermelha, verde e azul) e tamanho de 5472x3648 pixels. A altura do voo foi de cerca de 50 metros, resultando em um *Ground Sampling Distance* (GSD) de 1,48 cm. O objetivo deste voo foi visualizar os alvos de isopor colocados no local onde foram coletadas as amostras nematológicas, para georreferenciar esses pontos.

As imagens coletadas neste levantamento foram processadas utilizando o software Agisoft Metashape 1.7.5, no momento das tomadas das fotos o VANT contava com integração PPK (*Post Processed Kinematic*) e assim foram utilizados apenas dois pontos de verificação. O processamento das imagens no software Agisoft Metashape incluem várias etapas, entre elas as principais são: (1) criação de um projeto com as imagens, (2) alinhamento das fotos para extração dos pontos de correspondência, que foram utilizados como *tie points* para gerar a nuvem de pontos, (3) geração da nuvem densa de pontos e (4) geração do ortomosaico.

O ortomosaico resultante deste processamento foi utilizado para georreferenciar as amostras nematológicas distribuídas pelos blocos B e C (o georreferenciamento das amostras nematológicas do bloco A foi realizado a partir do levantamento geodésico com receptor RTK). Para isso, os locais onde as informações nematológicas foram coletadas foram marcados com um disco de isopor branco, a fim de serem localizados no ortomosaico e suas coordenadas medidas. Um exemplo de como foi visualizado os discos na imagem é apresentado na Figura 17.

Figura 17 – Exemplos dos alvos de isopor no ortomosaico. (A) Parcela, (B) e (C) alvos de isopor.



Fonte: A autora (2023).

O segundo levantamento aéreo foi realizado com um VANT do modelo X800 (GEO) da fabricante XFLY (Figura 18) equipado com o sensor hiperespectral Rikola (Figura 4) em 01 de fevereiro de 2022. Neste levantamento, a altura do voo foi de 140 m, o que resultou em um GSD de 9 cm. A área coberta por este levantamento incluiu os blocos A e B da fazenda experimental. As especificações da câmera Rikola estão disponíveis na Tabela 2.

Figura 18 – VANT modelo X800 (GEO).



Na estrada entre as parcelas de soja, foram colocados 3 alvos de Etileno Vinil Acetato (EVA) nas cores branca, preto e cinza (Figura 19). Foram coletadas 499 imagens em cada uma das 25 bandas, com faixa espectral entre 500 e 900 nm. As imagens obtidas neste levantamento foram submetidas a um processo de correção de *dark current*, que consiste na remoção de sinais elétricos gerados aleatoriamente pelo detector da câmera durante o tempo de exposição da imagem. Para essa correção, utilizou-se uma imagem de um alvo preto de EVA como referência. A interferência térmica acumulada em cada pixel da imagem durante o tempo de integração com a REM pode causar alterações na resposta espectral dos objetos capturados, sendo assim a correção é essencial para assegurar a qualidade dos dados obtidos (MORIYA, 2015). As imagens no formato RAW foram calibradas, corrigidas do efeito *dark current* e convertidas para tiff a partir do software HyperSpectral Imager da Rikola.

Figura 19 – Alvos de referência no processo de calibração por linha empírica.



Além disso, como as bandas de tempo-sequencial não são sincronizadas, foi utilizado o software CoRegister da Rikola para realizar o co-registro das bandas, garantindo que houvesse a sobreposição correta entre elas. As imagens co-registradas de cada banda foram fototrianguladas no software Agisoft Metashape utilizando os pontos do levantamento geodésico, o resultado final foi um ortomosaico por banda. Nesta etapa, alguns ortomosaicos foram excluídos do âmbito deste trabalho devido à sua qualidade insatisfatória. Na Tabela 3, são apresentados os comprimentos de onda central das 20 bandas que foram utilizadas.

Tabela 3 – Comprimentos de onda central das bandas utilizadas da câmera Rikola.

Banda	λ (nm)	FWHM (nm)	λ inicial (nm)
B1	509,35	15,80	501,45
B2	522,47	17,55	513,68
B3	538,12	16,80	529,71
B8	606,36	14,83	598,94
B9	620,22	16,26	612,08
B10	633,41	18,95	623,93
B11	654,21	16,28	646,07
B12	663,02	18,10	653,97
B13	672,61	19,87	662,67
B14	684,10	20,29	673,95
B15	691,90	18,33	682,73
B16	701,27	19,03	691,75
B18	723,16	19,15	713,58
B19	731,22	19,03	721,70
B20	741,37	17,85	732,44
B21	751,06	18,12	741,99
B22	771,87	18,61	762,56
B23	790,30	17,39	781,60
B24	810,46	18,00	801,46
B25	829,39	16,79	820,99

Realizou-se a calibração radiométrica a partir das medidas radiométricas das placas de EVA com o espectrorradiômetro e dos ortomosaicos das imagens hiperespectrais. O espectrorradiômetro utilizado foi o ASD FieldSpec® HandHeld (Figura 20), modelo UV/NIR, capaz de coletar dados espectrais no intervalo de 325 nm a 1075 nm com um intervalo de amostragem de 1,6 nm (ASD Inc., 2003). Este equipamento é portátil, com um campo de visão de 25° (FOV - *Field of View*) e filtros que permitem FOVs de 1° e 10°.

Figura 20 – Espectrorradiômetro FieldSpec® Handheld.



Utilizou-se o software ENVI, que empregou o método de correção por linha empírica. Esse método calcula uma relação empírica entre a radiância e a reflectância, utilizando alvos escuros e claros na área de estudo, sendo definida pela medição espectral na imagem e

em campo (MORIYA, 2015). A partir de uma regressão linear, são calculados os fatores de transformação dos números digitais (ND) em valores físicos, como a reflectância (SAHOO et al., 2013). O software ENVI oferece essa opção para calibração radiométrica usando o ELM e permite ao usuário selecionar um região de interesse (ROI - *Region of Interest*) na imagem para determinar os coeficientes de calibração usando uma ou mais áreas de ROI com características espectrais conhecidas.

3.2.4 Índice fitonematoides e nematoides de vida livre

Como comentado na Seção 2.2, os fitonematoides causam grandes prejuízos nas culturas, pois afetam o fluxo de absorção e translocação de água e nutrientes da planta. Segundo Ferraz e Brown (2016) os nematoides de vida livre não afetam diretamente as atividades humanas, mas eles podem ter uma influência significativa em processos importantes, como, a reciclagem de compostos contendo carbono, compostos nitrogenados e nutrientes minerais, com efeitos que beneficiam a fertilidade do solo e a produtividade das culturas agrícolas.

A presença de nematoides do grupo vida livre pode equilibrar os danos causados pelos fitonematoides, agindo como uma proteção natural. Dessa maneira, foi calculado um índice sobre a razão entre a soma das ocorrências de todos os gêneros de fitonematoides encontrados e a ocorrência de nematoides de vida livre (IFVL), conforme a Equação 3.1.

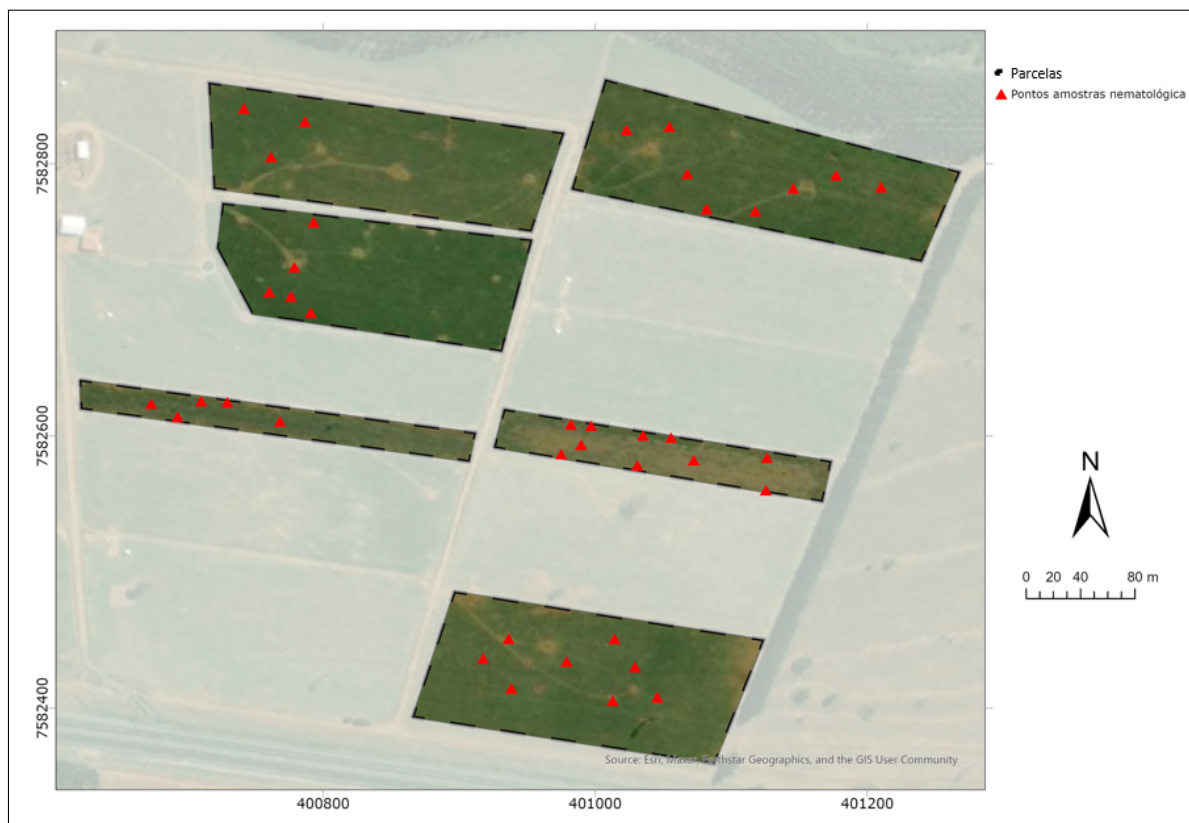
$$IFVL_i = \frac{\sum Fitonematoides_i}{\sum VidaLivre_i} \quad (3.1)$$

Onde: *Fitonematoides* é a somatória de todos os gêneros de fitonematoides contabilizados em número de indivíduos/mL de solo; *VidaLivre* é a ocorrência de vida livre em número de indivíduos/100 mL de solo; $i = 1, 2, \dots, n$ e n é o número de amostra.

3.2.5 Extração de informações para treinamento de algoritmo de aprendizado de máquina

Para treinar os algoritmos de aprendizado de máquina, foram utilizadas informações extraídas dos 20 ortomosaicos gerados a partir das bandas do sensor hiperespectral e 39 amostras nematológicas. Essas informações consistem nas respostas espectrais em fator de reflectância, que foram coletadas em locais correspondentes aos pontos onde foram amostrados nematoides nos blocos A e B. Isso foi possível graças ao georreferenciamento das amostras nematológicas. A distribuição dos 39 pontos das amostras nematológicas é apresentada na Figura 21.

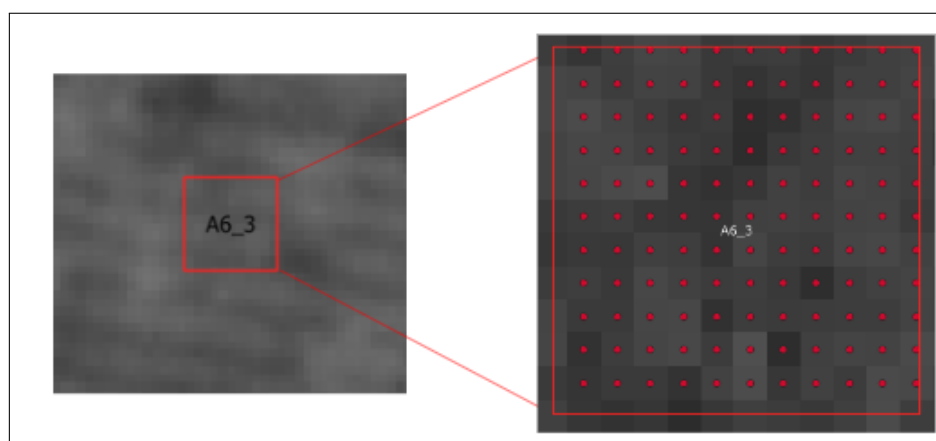
Figura 21 – Distribuição dos pontos das amostras nematológicas.



Fonte: A autora (2023).

Considerando que as amostras nematológicas foram obtidas a partir de uma área de 1x1 metro centrada em cada ponto amostrado, para extrair informações das respostas espectrais, foi criado um *buffer* de 1x1 metro ao redor do ponto. A média dos valores espectrais dos pixels dentro dessa área delimitada foi então calculada para todas amostras e para todas as bandas do sensor hiperespectral, o que corresponde a aproximadamente 121 pixels (pontos vermelho), como mostrado na Figura 22.

Figura 22 – Área considerada para cálculo da média das respostas espectrais.

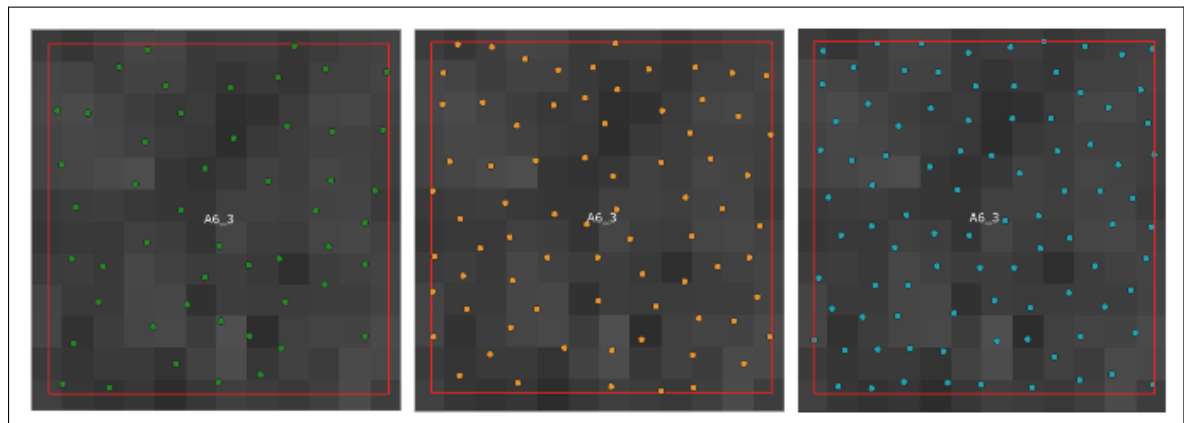


Fonte: A autora (2023).

Além da média das respostas espectrais de todos os pixels dentro da área delimitada, foram calculadas outras três médias, considerando 50, 75 e 95 pixels selecionados aleatoriamente dentro da área delimitada. Essa estratégia foi adotada para aumentar o conjunto de dados de treinamento do algoritmo, que passou de 39 para 156 elementos. A seleção dos pixels para cada conjunto foi feita de forma aleatória. Na Figura 23 é apresentado um exemplo da seleção de 50, 75 e 95 pixels dentro da área delimitada em volta do ponto da amostra nematológica.

Antes da extração das características, os dados a serem classificados foram ajustados para garantir a compatibilidade com os dados de treinamento. Foi gerada uma grade com células de 1x1 metro e calculada a média da resposta espectral para cada banda. Essa manipulação permitiu que os dados a serem classificados ficassem com o mesmo padrão dos dados de treinamento, que também foram considerados em uma área de 1x1 metro para cálculo da média da resposta espectral. Dessa forma, foi possível garantir que os dados estivessem em uma mesma escala e prontos para serem utilizados na classificação.

Figura 23 – Exemplo de seleção de conjunto de pixels aleatórios para cálculo da média das respostas espectrais: (A) considerando 50 pixels dentro da área, (B) considerando 75 e (C) considerando 95.



Fonte: A autora (2023).

3.2.6 Algoritmos de aprendizado de máquina

Foram utilizados três algoritmos para o desenvolvimento desta pesquisa: *Random Forest* (RF), *Support Vector Machine* (SVM) e *K-Nearest Neighbors* (KNN). São algoritmos não paramétricos que possuem boa performance (LU; WENG, 2007). O algoritmo *Random Forest* é um dos algoritmos não paramétricos mais populares, e é conhecido por apresentar alta acurácia em seus resultados (PAL, 2005; DU et al., 2015; BREIMAN, 2001). Além disso, ele também é usado para classificar variáveis por ordem de importância (SARTONO; SYAFITRI, 2010). O algoritmo SVM utiliza vetores separados por hiperplanos ou fronteiras de decisão. São baseadas na teoria de aprendizado estatístico e são eficazes em problemas complexos, tanto de classificação como de regressão (CORTES; VAPNIK, 1995). Ainda,

alguns trabalhos estudam o desempenho que diferentes algoritmos, como em Wang, Li e Li (2020) que os autores avaliam os algoritmos SVM, *Random Forest*, KNN e *Gradient Boosting* para estimar a produtividade da soja utilizando imagens hiperspectrais obtidas por VANT.

Para treinar e validar os algoritmos, foram utilizadas inicialmente 39 amostras. No entanto, por meio do processo de aumento de dados, a quantidade de amostras aumentou para 156. Os estudos recentes de técnicas baseadas em aprendizado de máquina utilizam diferentes proporções para dividir os dados de treinamento e de validação, como 80/20, 60/40, 70/30 e ainda 50/50. Em Nguyen et al. (2021) foi apresentado um estudo em que são avaliadas diferentes técnicas de aprendizado de máquina com várias proporções e relataram que 70/30 apresentou o melhor resultado. Portanto, seguindo a proporção recomendada a divisão dos dados para treinamento e validação foi feita em 70% e 30%, respectivamente.

Os algoritmos foram implementados em linguagem Python utilizando o ambiente de desenvolvimento integrado Google Colaboratory (Colab)¹, um serviço de nuvem gratuito fornecido pelo Google que permite aos usuários executarem códigos Python em notebooks interativos e compartilháveis. O Colab oferece recursos de processamento e armazenamento na nuvem, bem como acesso a GPUs e TPUs, o que torna mais fácil e eficiente a execução de modelos de aprendizado de máquina mais complexos. A principal biblioteca utilizada é a `scikit-learn`², que possui os três classificadores testados neste trabalho implementados e a função *GridSearchCV*.

Antes de utilizar os classificadores, foram investigados alguns parâmetros para encontrar a melhor combinação e alcançar os melhores resultados possíveis em cada algoritmo (*Random Forest*, *SVM* e *K-Nearest Neighbors*). Para isso, foi utilizada a função *GridSearchCV*³, que realiza uma pesquisa exaustiva nos valores dos parâmetros de um estimador.

No estudo sobre o classificador *Random Forest*, foram testados diferentes parâmetros, incluindo a profundidade das árvores com os valores de 3, 5 e 10, e o número de árvores na floresta com 100 e 500. Para o algoritmo SVM, o parâmetro chave foi o tipo de *kernel* utilizado, com testes realizados utilizando os *kernels* RBF, linear, polinomial e sigmoide. No caso do algoritmo *K-Nearest Neighbors*, foi avaliado o número de vizinhos com 10, 100 e 1000 como valores testados. Para cada parâmetro, foram realizados testes empíricos para determinar o conjunto de parâmetros que apresentaram os melhores resultados na classificação dos dados. É importante ressaltar que a seleção adequada de parâmetros é crucial para a performance do modelo de aprendizado de máquina, e deve ser realizada de

¹ <<https://colab.research.google.com/>>

² <<https://scikit-learn.org/>>

³ <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>

maneira criteriosa para garantir resultados confiáveis e válidos. Dessa forma, por meio da otimização dos parâmetros, foi possível encontrar a melhor combinação para cada algoritmo e, assim, alcançar resultados mais precisos e confiáveis.

Foram realizados testes em três números diferentes de divisões de classes para os três classificadores utilizados: 2, 5 e 10 classes, respectivamente. As classes foram divididas em intervalos iguais e foram calculadas com base nos valores obtidos pelo IFVL. O objetivo dessa etapa foi verificar o nível de detalhamento possível na classificação do nível de infestação de fitonematoides na soja, sem prejudicar o desempenho do algoritmo.

A escolha da quantidade de classes é uma questão crucial, pois classificações com poucas classes podem não refletir adequadamente a variação da infestação de nematoides na soja, enquanto classificações com muitas classes podem ser mais complexas e demandar mais trabalho computacional para serem processadas. Além disso, quando há poucas amostras e classes desbalanceadas, a classificação com muitas classes pode não ser muito interessante.

Dessa forma, optou-se por testar três opções de números de classes. A classificação binária, com apenas duas classes, é uma opção mais simples e direta, enquanto a classificação em 10 classes é mais complexa e detalhada. O resultado desse teste indicará qual é a melhor opção de divisão de classes para a classificação da intensidade de parasitas na soja. É importante destacar que a escolha da melhor opção para a classificação deve levar em consideração não apenas a precisão do algoritmo, mas também a facilidade de interpretação dos resultados obtidos. Além disso, é importante avaliar a disponibilidade e qualidade dos dados, a fim de garantir que o resultado da classificação seja confiável e preciso.

3.2.7 Variáveis de importância

No total, foram utilizadas 20 variáveis correspondentes as 20 bandas do sensor hiperespectral. É possível verificar a influência de cada variável no resultado do classificador *Random forest*. Para isso, foi utilizado o atributo *feature_importances_* da biblioteca `scikit-learn`⁴, que é utilizado para avaliar a importância relativa de cada variável na predição da variável de destino.

O cálculo da importância é feito a partir do índice de ganho de informação obtido em cada nó durante a construção da árvore de decisão. A partir daí, o algoritmo calcula uma pontuação de importância para cada variável, que é proporcional à soma total dos ganhos de informação gerados por aquele recurso em todas as árvores da floresta.

A importância de cada variável é normalizada de forma que a soma total de todas as importâncias seja igual a 1. Portanto, quanto maior a pontuação de importância de

⁴ <<https://scikit-learn.org/>>

uma variável, maior é a contribuição dele na classificação das amostras.

O atributo `feature_importances_` retorna um array com a importância de cada variável na mesma ordem em que foram fornecidos os dados de treinamento. Esse array pode ser usado para identificar as variáveis mais importantes para o modelo e, se necessário, selecionar apenas os mais relevantes para melhorar a eficiência computacional do modelo ou reduzir a dimensionalidade dos dados.

3.2.8 Métricas de avaliação dos classificadores

A fim de facilitar a comparação do desempenho do modelo proposto neste projeto com os encontrados na literatura, serão utilizadas as métricas comumente apresentadas nos trabalhos da área. Além da acurácia e RMSE (*Root Mean Squared Error*), essas medidas incluem sensibilidade, precisão, F_1 -score e acurácia balanceada. Os resultados das métricas foram obtidos a partir da média de 100 iterações para cada classificador.

Embora os trabalhos relacionados apresentem a acurácia como principal medida de desempenho do modelo, neste estudo foi utilizada a acurácia balanceada, tendo em vista que os dados contêm classes desbalanceadas e que a acurácia apresenta uma estimativa otimista, quando um classificador é testado em um conjunto de dados desequilibrado (BRODERSEN et al., 2010). Como pode ser observado na Equação 3.3, a acurácia balanceada é a média das sensibilidades (*recall*) de todas as c classes. Na Equação 3.2, VP_n é o número de verdadeiros positivos da classe n e FN_n é o número de falsos negativos da classe n .

$$\text{Sensibilidade} = \frac{VP_n}{VP_n + FN_n} \quad (3.2)$$

$$\text{Acurácia balanceada} = \frac{\sum_{n=1}^{n=c} \text{sensibilidade}_n}{c} \quad (3.3)$$

A precisão de c classes é definida conforme a Equação 3.4, sendo FP_n os falsos positivos da classe n .

$$\text{Precisão} = \frac{\sum_{n=1}^{n=c} VP_n}{\sum_{n=1}^{n=c} (VP_n + FP_n)} \quad (3.4)$$

A Equação 3.5 descreve a métrica F_1 de uma classe n .

$$F_{1n} = \frac{2 \times \text{precisão}_n \times \text{sensibilidade}_n}{\text{precisão}_n + \text{sensibilidade}_n} \quad (3.5)$$

E a Equação 3.6 descreve a métrica RMSE.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3.6)$$

Onde $\hat{y}_i$ é o valor previsto pelo modelo para a i -ésima observação, y_i é o valor real observado e n é o número total de observações.

4 Resultados e Discussões

Nesta seção, são apresentados e discutidos os resultados obtidos a partir dos experimentos descritos anteriormente.

4.1 Análise das respostas espectrais na cultura da soja com diferentes níveis de infestação por fitonematoídeos nos estádios fenológicos utilizando imagens do sensor MSI/Sentinel-2

As respostas espectrais dos pontos das amostras nematológicas (Figura 15) foram extraídas para os três conjuntos de imagens do sensor MSI/ Sentinel-2 referentes aos dias 26 de dezembro de 2021, 22 de janeiro de 2022 e 16 de fevereiro de 2022, em que cada data corresponde a um período fenológico.

A fim de simplificar a análise das respostas espectrais das plantas de soja parasitadas com nematoídeos, os pontos das amostras nematológicas foram divididos em quatro classes (C1, C2, C3 e C4) de densidade de comunidade de nematoídeos por amostras, onde C1 é a classe de menor densidade e C4 a classe com maior densidade. Nas Figuras 24, 25 e 26 são apresentadas as médias das respostas espectrais de cada classe para cada banda e para cada data respectivamente.

Figura 24 – Média das respostas espectrais das classes da data 1 (Dezembro).

26 DE DEZEMBRO 2021										
Bandas	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12
Média C1	0,04982	0,08052	0,09102	0,14014	0,23336	0,26826	0,27679	0,29781	0,31596	0,22165
Média C2	0,05075	0,08411	0,09229	0,14723	0,24442	0,27934	0,28490	0,30939	0,32140	0,22306
Média C3	0,05244	0,08742	0,09006	0,14830	0,25360	0,29256	0,29596	0,32104	0,31522	0,21354
Média C4	0,05370	0,08843	0,09570	0,14960	0,25210	0,29127	0,29960	0,32100	0,32477	0,22333

Fonte: A autora (2023).

Figura 25 – Média das respostas espectrais das classes da data 2 (Janeiro).

22 DE JANEIRO DE 2022										
Bandas	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12
Média C1	0,04343	0,07517	0,07477	0,1332	0,25871	0,29690	0,29822	0,32958	0,30330	0,19949
Média C2	0,04319	0,07411	0,07257	0,13069	0,26751	0,30961	0,311	0,34122	0,30059	0,19556
Média C3	0,04176	0,07456	0,06384	0,12558	0,28198	0,3269	0,32856	0,3596	0,27982	0,1745
Média C4	0,04366	0,07613	0,07376	0,13073	0,2719	0,31546	0,31226	0,35126	0,3045	0,1957

Fonte: A autora (2023).

Figura 26 – Média das respostas espectrais das classes da data 3 (Fevereiro).

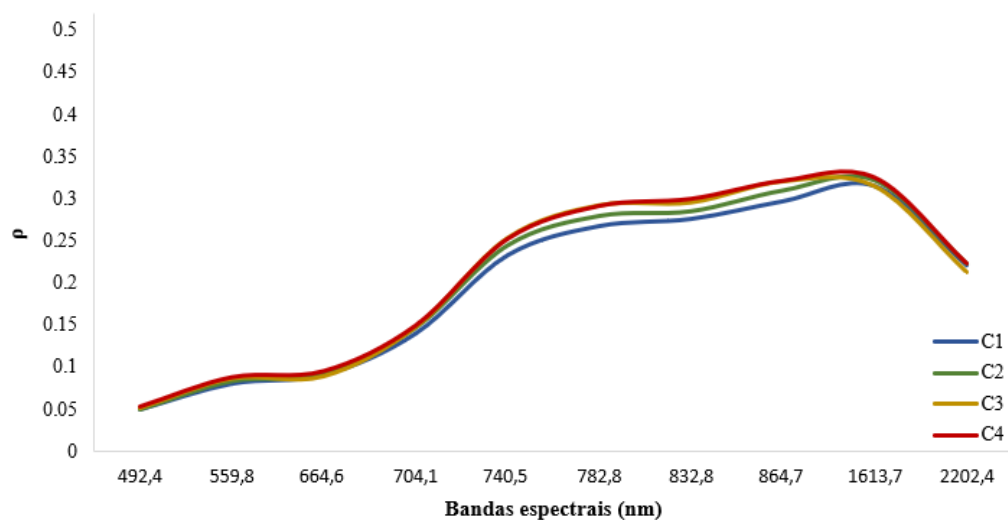
16 DE FEVEREIRO DE 2022										
Bandas	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12
Média C1	0,1721	0,21502	0,22831	0,30842	0,37271	0,41016	0,41627	0,44684	0,45297	0,34681
Média C2	0,16644	0,20851	0,21074	0,2934	0,38858	0,43286	0,4405	0,46668	0,42144	0,3169
Média C3	0,16354	0,20608	0,194	0,28052	0,40234	0,45182	0,45864	0,48986	0,38724	0,27968
Média C4	0,17006	0,2147	0,2146	0,2998	0,39143	0,43596	0,44253	0,4761	0,43153	0,32026

Fonte: A autora (2023).

A partir das médias das respostas espectrais de cada classe, foram gerados gráficos das três datas para visualizar as mudanças nas respostas ao longo dos períodos. As bandas compõem o eixo das abscissas e as respostas espectrais em reflectância compõem o eixo das ordenadas. Nas Figuras 27, 28 e 29 são apresentados os gráficos das médias das respostas espectrais.

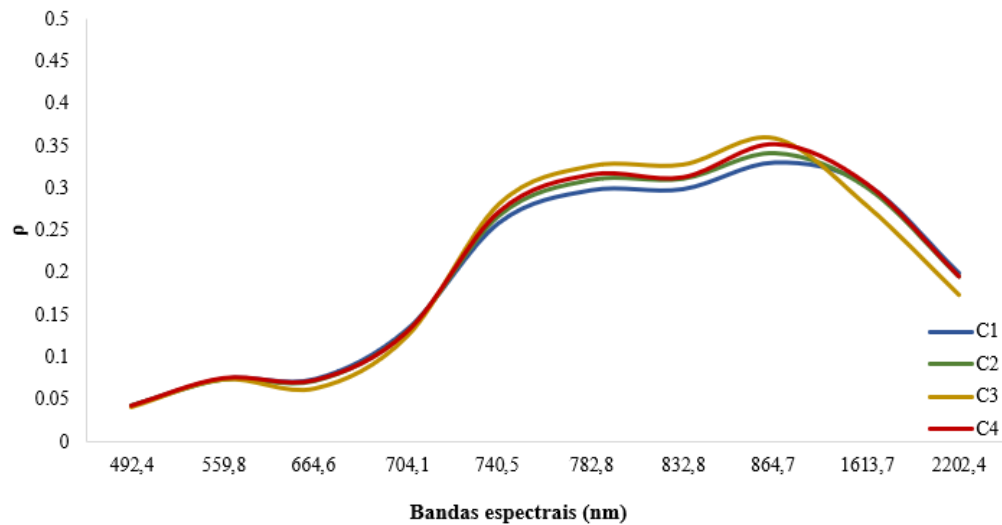
Além disso, foram gerados dois gráficos da diferença simples. O primeiro (Figura 30) mostra a diferença entre as respostas espectrais das imagens das datas de janeiro e dezembro, enquanto o segundo (Figura 31) mostra a diferença simples entre as respostas espectrais das imagens das datas de fevereiro e janeiro.

Figura 27 – Gráfico das médias das respostas espectrais em dezembro de 2021.



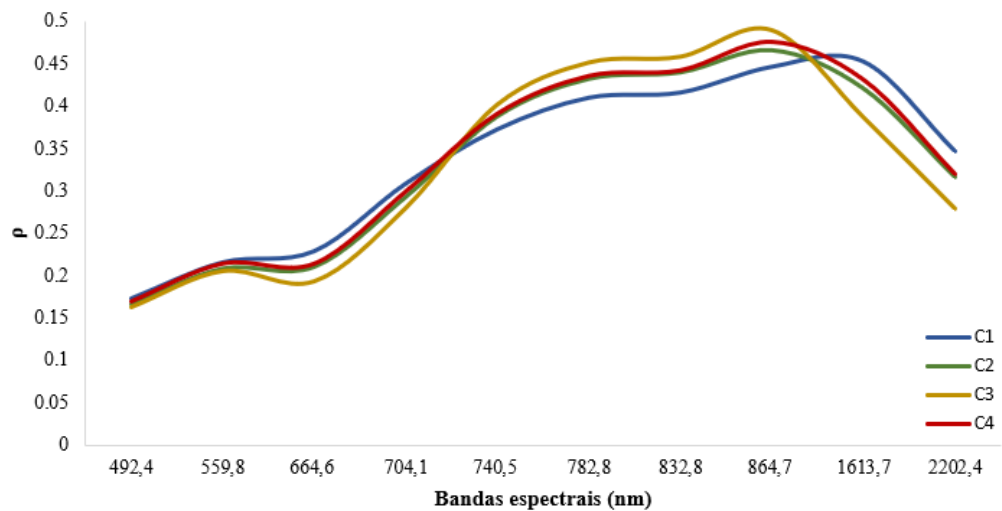
Fonte: A autora (2023).

Figura 28 – Gráfico das médias das respostas espectrais em janeiro de 2022.



Fonte: A autora (2023).

Figura 29 – Gráfico das médias das respostas espectrais em fevereiro de 2022.



Fonte: A autora (2023).

É possível observar que o gráfico correspondente às respostas espectrais da data de dezembro (Figura 27), as 4 classes não apresentam muita separabilidade entre elas até a banda 704,1 nm (*red edge 1*). No entanto, entre 740,5 nm e 864,7 nm (*red edge* e infravermelho próximo), ocorre a região com maior separabilidade entre as classes, sendo a classe 4 com as maiores respostas espectrais. Nas bandas 1613,7 nm e 2202,4 nm (infravermelho de ondas curtas), a separabilidade entre as classes não é perceptível.

Já nas respostas espectrais referentes a janeiro (Figura 28), nas bandas da região azul e verde, as classes não possuem muita separabilidade, enquanto na região do vermelho, a classe 3 apresenta uma leve separabilidade em relação às outras classes. A partir da banda 740,5 nm (*red edge*), a separabilidade entre as classes aumenta. Entre 740,5 nm e

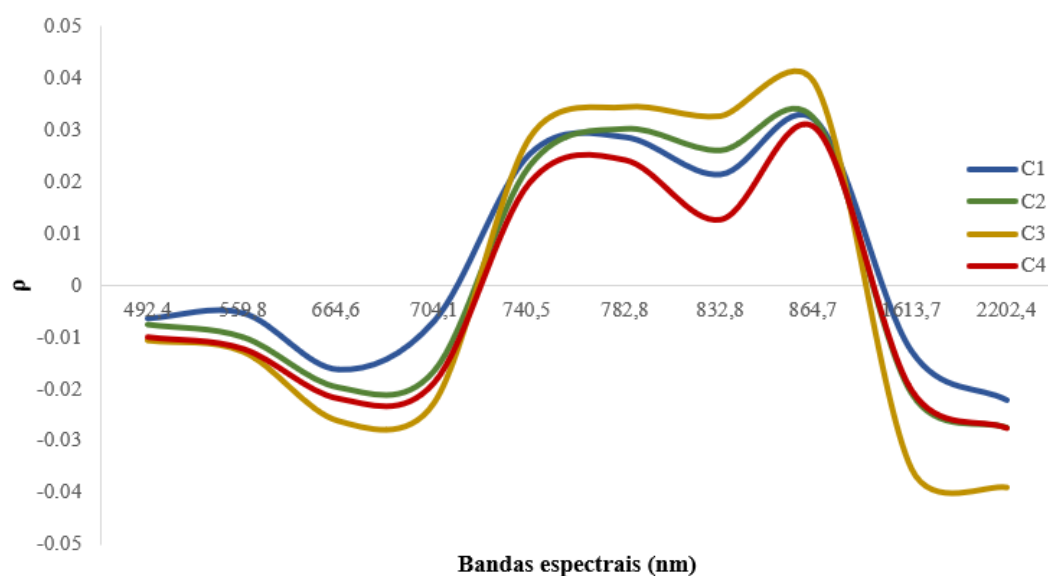
864,7 nm (*red edge* e infravermelho próximo), as repostas espectrais da classe 1 (menor nível de infestação de nematoides) são menores em relação às outras classes, e a classe 3 apresenta as maiores respostas espectrais. A partir de 1613,7 nm (infravermelho de ondas curtas), as respostas espectrais da classe 3 são menores em relação às outras classes.

No gráfico das repostas espectrais da data de fevereiro (Figura 29), a separabilidade entre as classes começa a se tornar perceptível já no fim da região do visível, na banda do vermelho, e permanece assim em todas as bandas, com exceção da banda 704,1 nm. A partir do comprimento de onda 740,5 nm, a separabilidade entre as classes se torna mais clara e definida e continua assim por toda a região do *red edge*, infravermelho próximo e infravermelho de ondas curtas.

As informações referentes às classes em relação a cada data são importantes para compreender as diferenças nas repostas espectrais das diferentes classes em cada estágio fenológico analisado. É válido ressaltar que as repostas espectrais do gráfico referente à data de dezembro refletem o estágio fenológico da soja entre os estádios V4 e V6. Na imagem de janeiro, a soja estava entre os estádios R5 e R7, e na data de fevereiro, encontrava-se no estágio fenológico final, estágio R8.

A Figura 30 apresenta o gráfico da diferença entre as repostas espectrais das imagens da data de janeiro e dezembro. Ao analisar visualmente os gráficos e observar o gráfico da diferença (Figura 30), pode-se verificar que não há mudanças significativas entre as repostas das imagens obtidas em dezembro e janeiro.

Figura 30 – Gráfico da diferença entre as repostas espectrais da data de janeiro e dezembro.



Fonte: A autora (2023).

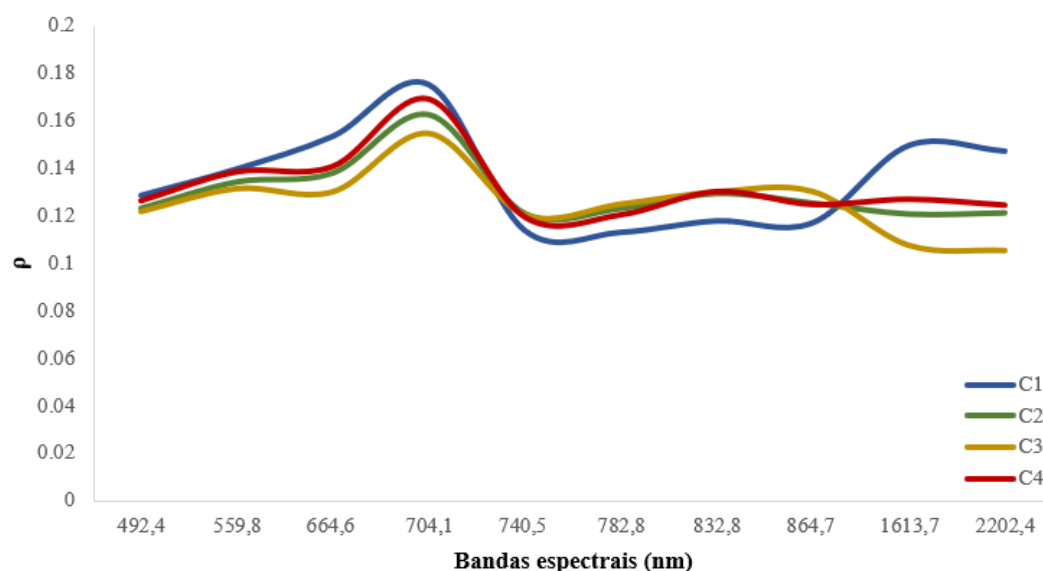
Na análise do gráfico da diferença entre as repostas espectrais das imagens das

datas de janeiro e dezembro, é possível observar que não foram encontradas diferenças significativamente altas. Na região do visível, a maior diferença entre as duas datas foi na banda do vermelho (representada pela banda 664,6 nm no Sentinel-2) e na classe 3 (em que o nível de infestação de nematoides é médio). Entre as bandas de 740,5 nm e 2202,4 nm (região do infravermelho próximo, *red edge* e infravermelho de ondas curtas, a diferença entre as datas aumenta, mas não ultrapassa 0,04 (em termos de reflectância), sendo a classe 3 a que apresenta o maior valor na diferença.

Também é possível observar que na região entre 740,5 nm e 864,7 nm (região correspondente ao infravermelho próximo e *red edge*), a classe 4 (alto nível de infestação por nematoides) apresenta a menor diferença nas respostas espectrais entre as imagens das datas de janeiro e dezembro. Isso sugere que, para a classe 4, as respostas espectrais se mantiveram relativamente estáveis ao longo dessas duas datas.

Entretanto, é notável a diferença de resposta espectral entre as datas de janeiro e fevereiro, na qual a curva se desloca para cima nas 4 classes. Isso indica que as diferenças entre as respostas na data de fevereiro foram significativamente maiores, conforme o gráfico da diferença entre as respostas espectrais das imagens correspondentes à data de fevereiro e janeiro, apresentado na Figura 31. É digno de nota que as plantas de soja na terceira data (fevereiro) estavam em seu estágio final, já que as imagens são do dia 16 de fevereiro e a colheita da cultura foi realizada no dia 22 de fevereiro de 2022.

Figura 31 – Gráfico da diferença entre as respostas espectrais da data de fevereiro e janeiro.



Fonte: A autora (2023).

Houve aumento nas respostas espectrais de todas as bandas na data de fevereiro, porém algumas bandas se destacaram. Na classe 1, as bandas que tiveram maior aumento de resposta entre a data de janeiro e fevereiro foram as bandas do 704,1 nm (*Red Edge 1*) e do 664,6 nm (vermelho). Já na classe 2, as maiores diferenças das respostas entre a

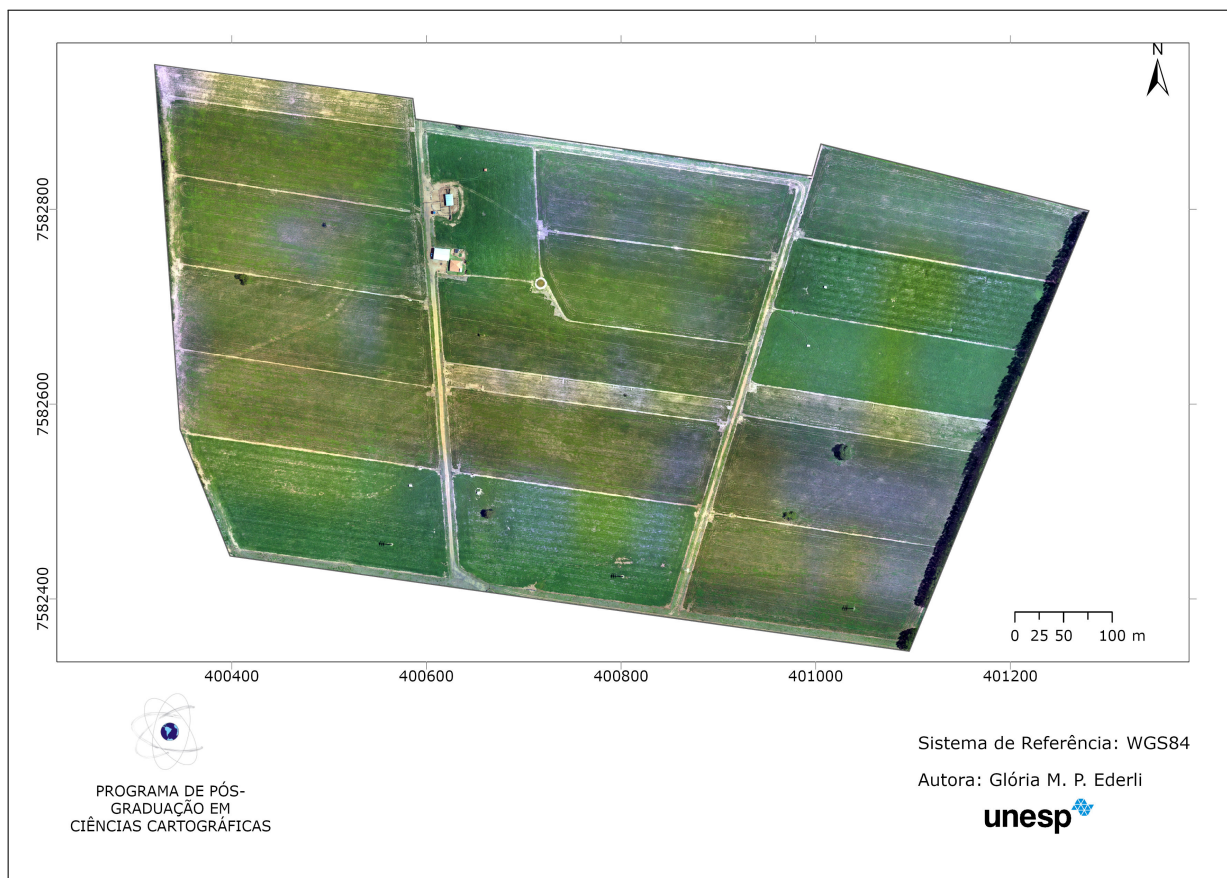
data 2 e a data 3 foram nas bandas do 704,1 nm (*Red Edge 1*), 832,8 nm (infravermelho próximo) e 864,7 nm (*Red Edge 4*). Nas classes 3 e 4, analogamente à classe 3, houve maior aumento de resposta espectral nas bandas 704,1 nm (*Red Edge 1*), 832,8 nm (infravermelho próximo) e 864,7 nm (*Red Edge 4*).

Com base na análise deste estudo sobre as respostas espectrais em diferentes estádios fenológicos, observa-se que o estágio final do reprodutivo é o período mais apropriada para realizar levantamentos aéreos, uma vez que todas as bandas espectrais analisadas apresentaram um aumento nas respostas espectrais. Além disso, neste período fenológico, as classes de infestação de fitonematoides apresentaram uma maior separabilidade entre elas, o que sugere que algoritmos de aprendizado de máquina podem ser capazes de distinguir essas classes com maior precisão.

4.2 Análise das respostas espectrais com diferentes níveis de infestação por fitonematoides utilizando imagens do sensor hiperespectral

Na Figura 32, é apresentado o ortomosaico gerado a partir do levantamento aéreo realizado com o DJI Phantom 4, utilizando o próprio sensor do equipamento, conforme descrito na Seção 3.2.3. Foram coletadas um total de 464 imagens, e o processamento foi realizado no software Agisoft Metashape. O ortomosaico RGB foi utilizado para identificar e obter as coordenadas dos pontos de amostras nematológicas marcados com disco de isopor branco (Figura 17).

Figura 32 – Ortomosaico gerado a partir das imagens adquiridas pelo DJI Phantom 4.

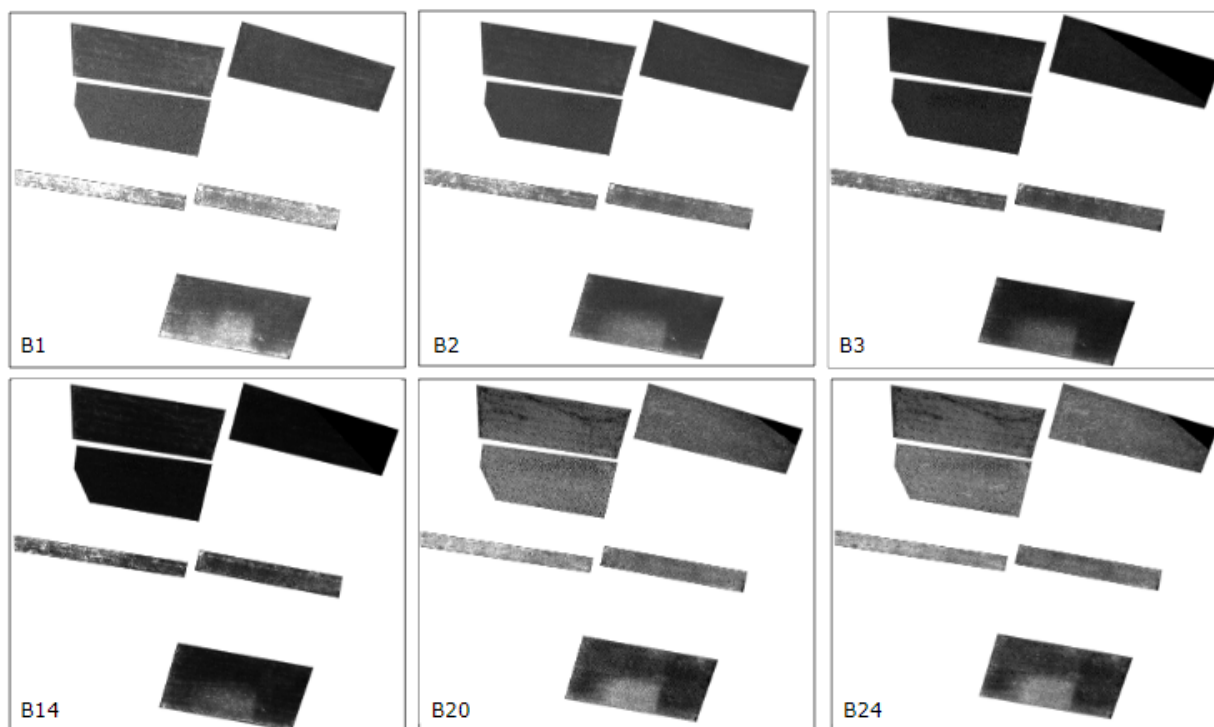


Fonte: A autora (2023).

Embora as respostas espectrais dessas imagens não tenham sido utilizadas no desenvolvimento da pesquisa, essa etapa foi necessária para georreferenciar os pontos de amostras nematológicas a partir dessa imagem. Essa abordagem permitiu uma localização precisa das amostras, estabelecendo uma relação entre a localização das plantas afetadas por nematoides e as características espectrais observadas nas imagens.

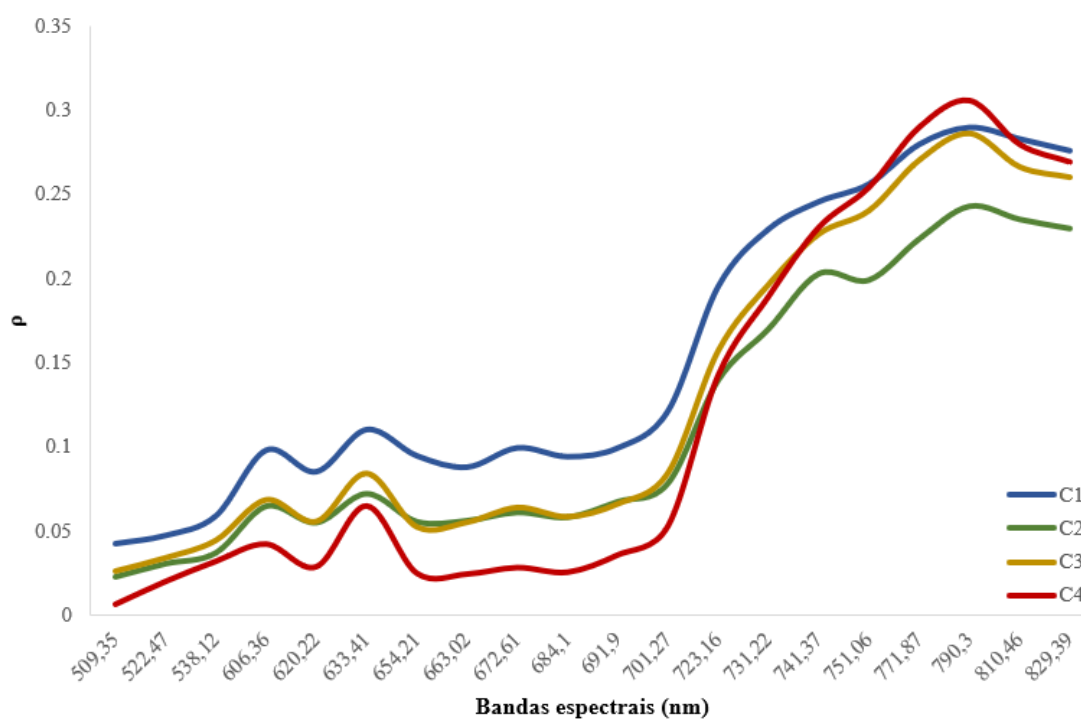
Na Figura 33, são apresentados exemplos de ortomosaicos gerados a partir das imagens coletadas pelo sensor hiperespectral. As bandas B1, B2, B3, B14, B20 e B24 são apresentadas em tons de cinza, o que dificulta a extração visual de informações. Para melhorar a visualização e a análise dos dados, foram plotadas as médias dos espectros de reflectância para 4 classes de infestação de fitonematoides calculadas a partir do IFVL, como apresentada na Figura 34. As classes foram separadas a partir de intervalos iguais dos valores de IFVL e a média de cada classe para cada uma das 20 bandas foi calculada. A escolha de 4 classes foi feita para que fosse compatível com os espectros analisados do sensor MSI na Seção 4.1.

Figura 33 – Exemplos de ortomosaicos gerados a partir das imagens adquiridas pelo sensor hiperespectral.



Fonte: A autora (2023).

Figura 34 – Gráfico das médias das respostas espectrais para o sensor hiperespectral.



Fonte: A autora (2023).

Analisando o gráfico apresentado na Figura 34, é possível notar que as classes

possuem uma boa separabilidade entre elas, principalmente entre as classes C1 e C4, que representam, respectivamente, o menor e o maior nível de infestação de fitonematoides com base no IFVL. Observa-se que as plantas mais saudáveis (associadas à classe C1) apresentam uma resposta espectral maior do que as plantas com maior nível de infestação de fitonematoides (associadas à classe C4) na região do visível e no início do *red edge*. Por outro lado, na região entre 771,87 nm e 790,30 nm, a classe C4 apresenta as maiores respostas espectrais.

É válido destacar que, embora os espectros de reflectância das classes C2 e C3 pareçam sobrepor-se em algumas regiões, ainda é possível notar diferenças sutis que podem ser exploradas pelos algoritmos de classificação. Essas classes, no geral, apresentam respostas semelhantes na região do visível, mas nos comprimentos de onda de 538,12 nm e 633,41 nm, há uma separabilidade entre elas. A partir da região espectral *red edge*, a diferença entre as classes aumentam, sendo que a maior diferença no comprimento de onda 751,06 nm.

A boa separabilidade entre as classes é importante, pois indica que os classificadores baseados em aprendizado de máquina provavelmente serão capazes de distingui-las e classificá-las com precisão. Além disso, é importante mencionar que a separabilidade das classes é um indicativo de que a abordagem de análise espectral pode ser uma ferramenta eficiente para detecção e mapeamento de infestação de fitonematoides em lavouras de soja.

4.3 Métricas de avaliação dos classificadores

4.3.1 Matriz de confusão

Nesta Seção, serão apresentadas as métricas de avaliação dos classificadores utilizados no estudo. Os parâmetros dos algoritmos RF, SVM e KNN foram selecionados por meio da função GridSearchCV, que realiza uma pesquisa exaustiva nos valores dos parâmetros de um estimador. Como descrito na Seção 3.2, diferentes valores foram testados para os principais parâmetros dos classificadores.

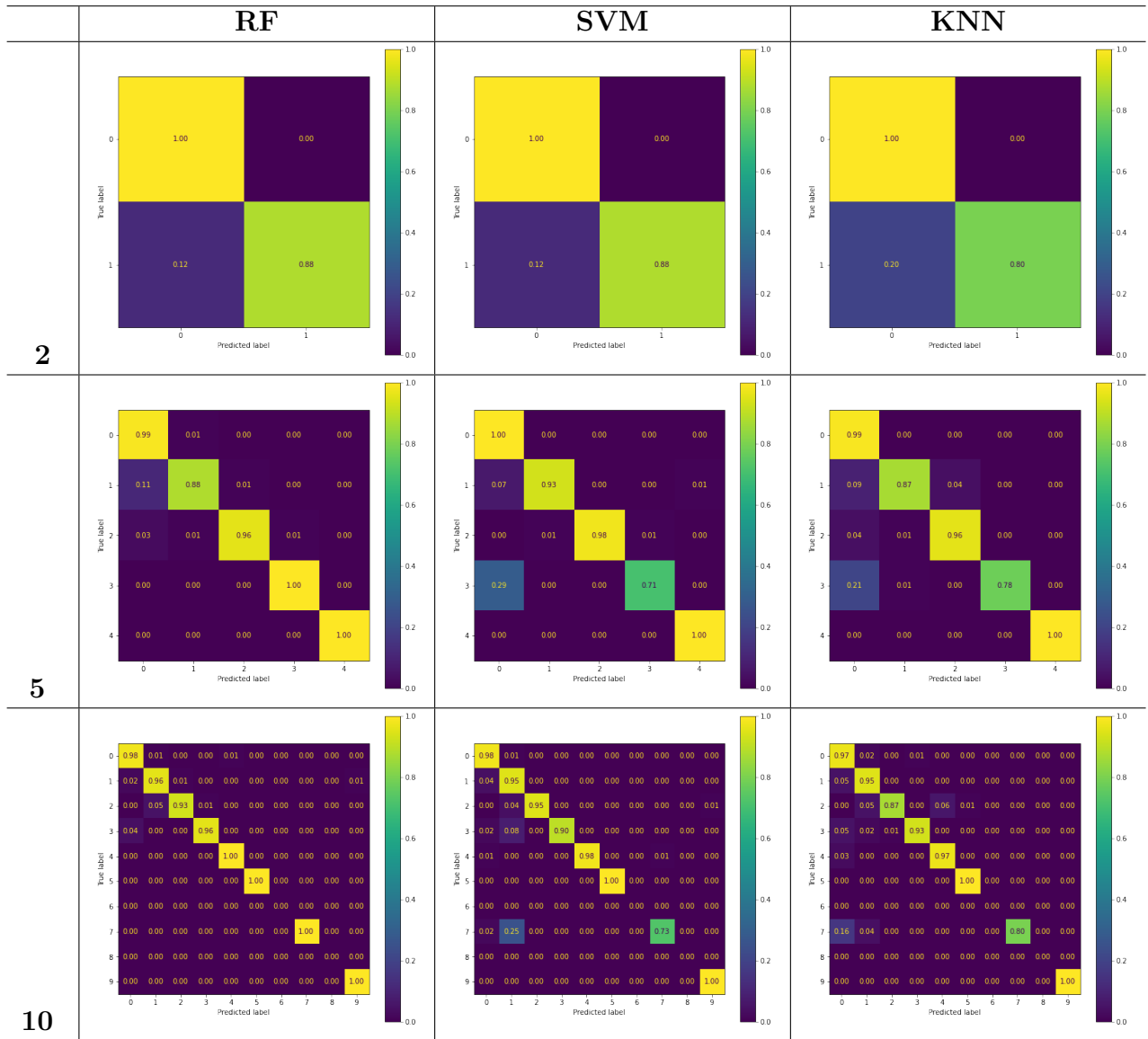
No caso do RF, foram testados diferentes valores para a profundidade das árvores (3, 5 e 10) e para o número de árvores na floresta (100 e 500). Os parâmetros que apresentaram o melhor desempenho foram uma profundidade de árvore de 10 e um número de árvores na floresta de 500.

Já para o SVM, o parâmetro chave testado foi o tipo de kernel utilizado, com testes realizados utilizando os kernels RBF, linear, polinomial e sigmoid. O kernel RBF obteve o melhor desempenho, o que é comum em dados que não possuem uma separação linear clara. Por fim, no caso do KNN, foram testados diferentes números de vizinhos (10, 100 e 1000), sendo que o melhor resultado foi obtido com o número de vizinhos igual a 10.

As matrizes de confusão são amplamente utilizadas para avaliar o desempenho de classificadores em aprendizado supervisionado. Elas permitem avaliar quantitativamente os acertos e erros do classificador para cada classe, destacando os valores de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). A partir das matrizes de confusão, é possível calcular diversas métricas de avaliação, como acurácia, precisão, sensibilidade, F1-score, entre outras.

Na Tabela 4, são apresentadas as matrizes de confusão normalizadas geradas para cada algoritmo, com 2, 5 e 10 classes divididas em intervalos iguais. Os resultados apresentados são as médias de 100 iterações, como descrito na Seção 3.2.6.

Tabela 4 – Matrizes de confusão normalizadas.



Na análise das matrizes de confusão, é possível observar que, ao considerar a divisão do problema em 10 classes, os classificadores não obtêm um bom desempenho. Para as classes 7 e 9, nenhum dado de validação foi classificado corretamente, ou seja, não houve

verdadeiros positivos nem verdadeiros negativos. Essa análise por classes é importante, pois as métricas que resultam em um único valor para todas as classes podem mascarar as classes que não foram aprendidas pelos classificadores. Dessa forma, a análise individual das matrizes de confusão possibilita uma avaliação mais precisa e específica do desempenho do classificador para cada classe, permitindo que sejam feitas melhorias e ajustes para o melhor aprendizado dessas classes.

4.3.2 Sensibilidade, precisão, F_1 score, RMSE e Acurácia balanceada

Foram utilizadas 156 amostras, resultantes do processo de aumento de dados, para o treinamento e validação dos algoritmos. A divisão dos dados foi feita de forma que 70% das amostras foram destinadas ao treinamento e 30% para a validação. Para uma análise mais completa dos modelos gerados, foram calculados diversas métricas, como o RMSE, a sensibilidade, a precisão, o F_1 -score e a acurácia balanceada. Na Tabela 5, são apresentados os resultados obtidos para o cálculo do RMSE para cada classificador, considerando 2, 5 e 10 classes de infestações de nematoides.

O RMSE é uma métrica importante para avaliar a acurácia do modelo, já que representa a diferença média entre os valores preditos e os valores observados. Quando o RMSE é grande, isso pode indicar que o modelo possui alta variância e, portanto, está fazendo previsões menos acuradas. Por outro lado, se o RMSE apresentar um valor pequeno, isso indica que o modelo possui baixa variância e faz previsões mais acuradas. Assim, o RMSE é uma medida importante para avaliar a qualidade das previsões geradas pelos modelos.

Tabela 5 – RMSE.

	RF	SVM	KNN
2 Classes	0,0663	0,0758	0,0983
5 Classes	0,1717	0,2124	0,2524
10 Classes	0,2423	0,4735	0,5191

A análise dos resultados de RMSE revela que os modelos com apenas 2 classes apresentaram valores menores, o que era esperado, já que um problema de classificação binária é geralmente mais simples. Além disso, observa-se que os modelos criados com o classificador RF apresentaram valores de RMSE mais baixos do que os modelos criados com os classificadores SVM e KNN, em todas as variações do número de classes testadas. O classificador SVM obteve resultados intermediários, enquanto o classificador KNN apresentou os maiores valores de RMSE.

As métricas sensibilidade, precisão e F_1 -score foram calculadas para cada classificador e para cada classe. Assim, a Tabela 6 apresenta os resultados obtidos para o cálculo

da sensibilidade, precisão e do F_1 -score para cada classificador e para 2 classes (C1, C2).

Tabela 6 – Sensibilidade, precisão e F_1 para 2 classes.

Classificador	Classe	Sensibilidade	Precisão	F_1 -score
RF	C1	0,99	1,00	0,99
	C2	0,97	0,88	0,90
SVM	C1	0,99	1,00	0,99
	C2	0,97	0,88	0,91
KNN	C1	0,98	1,00	0,99
	C2	1,00	0,80	0,87

Do mesmo modo, foram calculadas as métricas considerando o problema com 5 classes. A Tabela 7 apresenta os resultados obtidos para o cálculo da sensibilidade, precisão e do F_1 -score para cada classificador e para 5 classes (C1, C2, C3, C4 e C5).

Tabela 7 – Sensibilidade, precisão e F_1 para 5 classes.

Classificador	Classe	Sensibilidade	Precisão	F_1 -score
RF	C1	0,96	0,99	0,97
	C2	0,98	0,88	0,92
	C3	0,99	0,96	0,97
	C4	0,99	1,00	0,99
	C5	0,99	1,00	1,00
SVM	C1	0,97	1,00	0,98
	C2	0,99	0,93	0,95
	C3	1,00	0,98	0,98
	C4	0,70	0,71	0,70
	C5	0,98	1,00	0,99
KNN	C1	0,95	0,99	0,97
	C2	0,99	0,87	0,92
	C3	0,94	0,96	0,94
	C4	0,78	0,78	0,78
	C5	0,99	1,00	1,00

As métricas foram calculadas mais uma vez, dessa vez considerando 10 classes. A Tabela 8 apresenta os resultados obtidos para o cálculo da sensibilidade, precisão e do F_1 -score para cada classificador e para 10 classes (C1, C2, C3, C4, C5, C6, C7, C8, C9 e C10).

Tabela 8 – Sensibilidade, precisão e F_1 para 10 classes.

Classificador	Classe	Sensibilidade	Precisão	F_1 -score
RF	C1	0,98	0,98	0,98
	C2	0,96	0,96	0,96
	C3	0,99	0,93	0,95
	C4	0,99	0,96	0,96
	C5	0,98	1,00	0,99
	C6	1,00	1,00	1,00
	C7	0,00	0,00	0,00
	C8	1,00	1,00	1,00
	C9	0,00	0,00	0,00
	C10	0,98	1,00	0,99
SVM	C1	0,97	0,98	0,97
	C2	0,92	0,95	0,93
	C3	0,99	0,95	0,97
	C4	0,98	0,90	0,94
	C5	0,99	0,98	0,98
	C6	1,00	1,00	1,00
	C7	0,00	0,00	0,00
	C8	0,72	0,73	0,72
	C9	0,00	0,00	0,00
	C10	0,99	1,00	0,99
KNN	C1	0,94	0,97	0,95
	C2	0,94	0,95	0,94
	C3	0,99	0,87	0,92
	C4	0,97	0,93	0,94
	C5	0,94	0,97	0,95
	C6	0,98	1,00	0,99
	C7	0,00	0,00	0,00
	C8	0,80	0,80	0,80
	C9	0,00	0,00	0,00
	C10	1,00	1,00	1,00

Para as classes C7 e C9 utilizando os 3 classificadores não há informação calculada, ou seja, o valor é zero para as 3 métricas estatísticas analisadas. Isso ocorreu, como comentado anteriormente, porque essas classes não tiveram resultados classificados como verdadeiros positivos ou verdadeiros negativos.

Considerando os resultados das matrizes de confusão (Tabela 4) e as métricas de sensibilidade, precisão e F_1 -score, é possível concluir que a escolha de um modelo que utilize 10 classes não é recomendada para essa aplicação. Embora um modelo com um alto número de classes possa fornecer um nível de detalhamento maior na localização das áreas infestadas com nematoides, os resultados indicam que os classificadores não obtiveram um bom desempenho nessa configuração. Um grande número de elementos de treinamento poderia potencializar um desempenho mais efetivo dos algoritmos para classificação em

um grande número de classes. No entanto, devido a limitações de disponibilidade de dados, não foi possível utilizar um número tão elevado de elementos de treinamento neste trabalho.

Por outro lado, uma classificação binária também não seria interessante, pois apenas dois níveis de infestação de nematoides seriam detectados, o que limitaria a análise do nível de infestação das áreas. Além disso, uma classificação binária poderia limitar a análise, focalizando apenas na presença ou ausência de nematoides em uma determinada área, sem considerar o nível de infestação. Nesse sentido, para um estudo mais completo e detalhado sobre a infestação de nematoides, é necessário considerar um número maior de classes.

A última métrica foi a acurácia balanceada, que foi calculada para os três classificadores e considerando as três variações de números de classes. Entre as métricas analisadas, a acurácia balanceada teve mais peso na escolha do classificador para a geração do resultado final, devido aos elementos de treinamento não possuírem número de classes balanceadas. Isso significa que há mais elementos pertencentes a uma classe do que a outras.

Ao considerar a divisão em cinco classes, a classe que mais possui elementos é a C1 (baixa infestação de nematoides) e a que menos possui é a C5 (maior nível de infestação). Portanto, a Tabela 9 apresenta os resultados do cálculo da acurácia balanceada para cada classificador e considerando 2, 5 e 10 classes de infestações de nematoides.

Vale ressaltar que um valor alto de acurácia balanceada indica que o modelo está com um bom desempenho na classificação das diferentes classes, mesmo considerando a desigualdade na quantidade de elementos em cada classe.

Tabela 9 – Acurácia balanceada.

	RF (%)	SVM (%)	KNN (%)
2 Classes	93,8139	93,5174	90,0784
5 Classes	96,5571	92,1500	91,9400
10 Classes	97,7756	93,7780	93,5491

Analisando a Tabela 9, pode-se observar que o maior valor foi para o modelo considerando o classificador RF e 10 classes, entretanto, como apontado anteriormente, os modelos considerando 10 classes não foram capazes de aprender as classes C7 e C9 o que auxiliou na decisão de não considerar os modelos utilizando 10 classes, uma vez que o número de elementos para treinamento não iria garantir um bom desempenho para um modelo com 10 classes.

Além disso, como também já foi citado anteriormente, os modelos com apenas duas classes não seria interessante para alcançar o objetivo do estudo, que é mapear

a intensidade de infestação de nematoides e não a presença e ausência de nematoides. Assim, o número de classes ideal para esta pesquisa é 5, que pode alcançar um nível de detalhamento satisfatório e não afeta o bom desempenho de nenhum dos classificadores para a detecção dos nematoides a partir de dados de sensoriamento remoto.

Deste modo, a análise dos resultados obtidos pela acurácia balanceada concentra-se na avaliação dos classificadores utilizados. Observa-se que entre os classificadores RF, SVM e KNN, o que obteve a maior acurácia balanceada foi o RF com um valor de 96,56%, seguido pelo algoritmo SVM com 92,15%, enquanto o KNN teve a acurácia mais baixa, com 91,94%. É importante destacar que as diferenças entre as acurácias balanceadas são consideráveis, sugerindo que a variação não é atribuída a ruídos, mas sim ao desempenho dos classificadores.

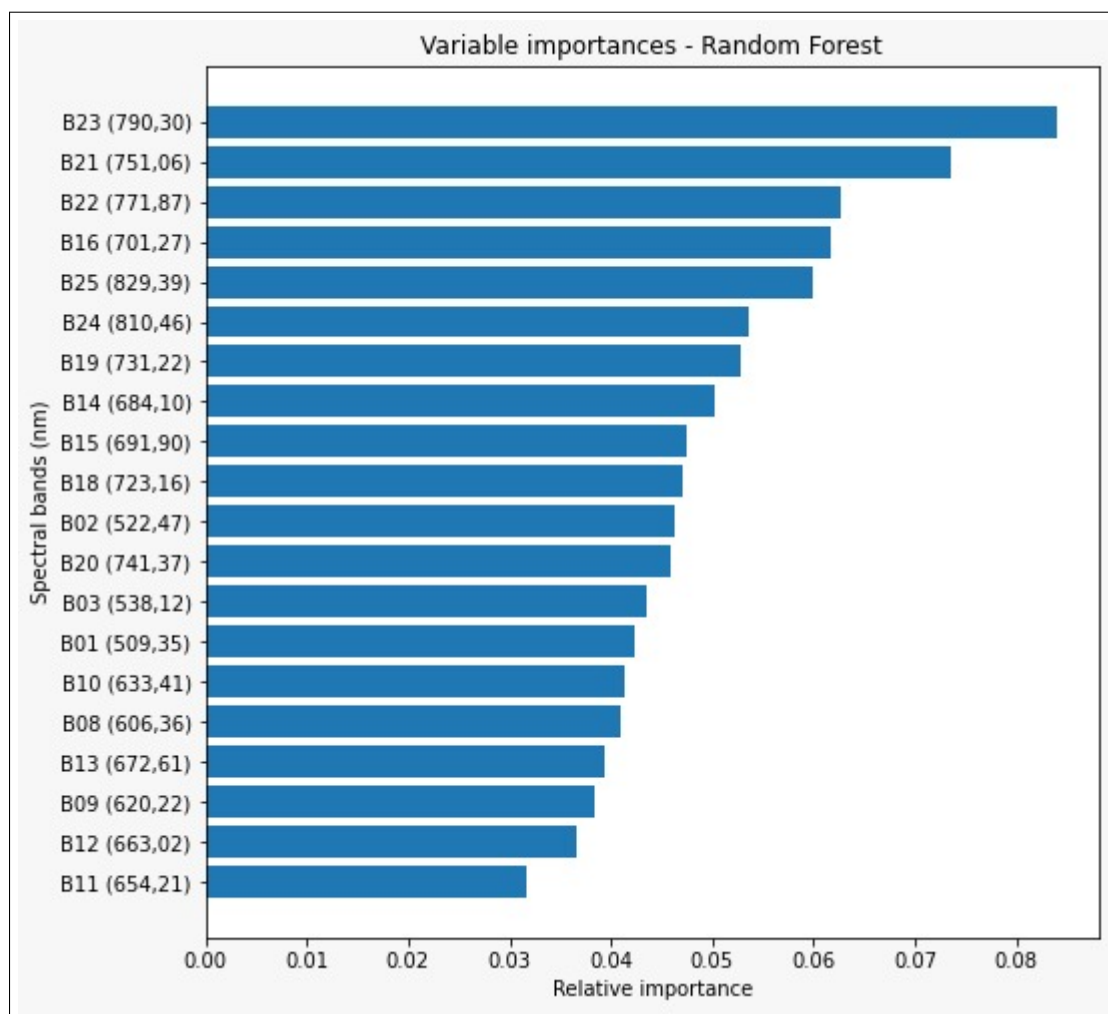
Vários trabalhos recentes indicam o bom desempenho do algoritmo RF em aplicações na área de agricultura de precisão, como o estudo de Wang, Li e Li (2020), que avaliou o desempenho de diferentes algoritmos de aprendizado de máquina, incluindo SVM, RF e *Gradient Boosting*, e relatou que o RF obteve a melhor precisão na estimativa da produtividade da soja usando imagens hiperespectrais obtidas por VANT. Além disso, em Almeida et al. (2018), o classificador RF também obteve o melhor desempenho para estimar a produtividade da soja com base em imagens capturadas por VANT, quando comparado com os algoritmos KNN, SVM, Rede Neural Artificial (RNA) e Regressão Linear Múltipla (RLM).

Com base nos resultados apresentados pelas métricas estatísticas, sobretudo pela acurácia balanceada, pode-se concluir que o modelo utilizando o classificador RF com 5 classes foi o mais adequado para a aplicação de mapear a intensidade de infestações de nematoides na soja. Uma possível causa para o algoritmo RF ter um bom desempenho é que ele consegue trabalhar bem com dados de alta dimensionalidade e com interações complexas presentes nas variáveis de entrada, o que pode acontecer comumente em dados provenientes de Sensoriamento Remoto e Fotogrametria que são utilizadas em agricultura de precisão. O algoritmo RF também é interessante quando há um número limitado de elementos de treinamento e para casos não lineares, já que ele não requer muita sintonia de hiperparâmetros.

4.3.3 Variáveis por ordem de importância

A partir dos resultados obtidos pelo classificador RF, foi possível identificar as variáveis mais influentes para este classificador em uma análise de dados. A classificação das variáveis por ordem de importância gerada a partir dos resultados do algoritmo *Random Forest* é apresentada na Figura 35. A partir dessa classificação, é possível analisar as relações mais relevantes e menos relevantes consideradas pelo RF, o que pode auxiliar na compreensão do problema.

Figura 35 – Classificação das variáveis por ordem de importância.



Fonte: A autora (2023).

As quatro variáveis mais influentes na classificação a partir do modelo do RF foram as bandas com comprimento de onda central de 790,30 nm, 751,06 nm, 771,87 nm e 701,27 nm, respectivamente. Esses comprimentos de onda pertencem à região do *Red Edge* do espectro eletromagnético. No espectro visível, a banda do vermelho também desempenha um papel importante, especialmente próxima à banda de absorção da clorofila (670 nm), que é representada aqui pela Rikola em torno de 684 nm. Isso significa que a banda do vermelho também pode ser utilizada em estudos de estresse vegetal. Portanto, é possível concluir que essas regiões contribuíram significativamente para o objetivo da pesquisa de identificar plantas de soja parasitadas por nematoides e seu nível de infestação.

Jensen e Epiphanyo (2009) abordam o *Red Edge* como uma região importante do espectro eletromagnético para o monitoramento das plantas, pois essa região é altamente sensível à clorofila e pode indicar mudanças na saúde da planta, como estresse hídrico, doenças e deficiências nutricionais. Os sensores remotos que capturam dados nessa região são ferramentas úteis para avaliar a saúde das plantas e destacar os problemas, permitindo a implementação de medidas para evitar uma perda significativa na produção.

4.4 Mapa temático

Como discutido na Seção 4.3, o modelo que apresentou o melhor desempenho foi aquele que utilizou o algoritmo RF e considerou 5 classes com intervalos iguais. Assim, foi possível gerar um mapa temático da classificação da soja parasitada por fitonematoides em níveis de infestação, baseada no IFVL. O modelo selecionado, como demonstrado anteriormente, foi o algoritmo *Random Forest*, com 5 classes, em que a classe C1 representa o menor nível de infestação e a classe C5 representa o maior nível de infestação. Na Tabela 10, são apresentados os valores numéricos das classes, calculados a partir dos valores do IFVL.

Tabela 10 – Valores numérico de cada classe.

Classe	Intervalo numérico (IN)
C1	$0 \leq \text{IN} < 0,5668$
C2	$0,5668 \leq \text{IN} < 1,1336$
C3	$1,1336 \leq \text{IN} < 1,7004$
C4	$1,7004 \leq \text{IN} < 2,2672$
C5	$2,2672 \leq \text{IN} \leq 2,8340$

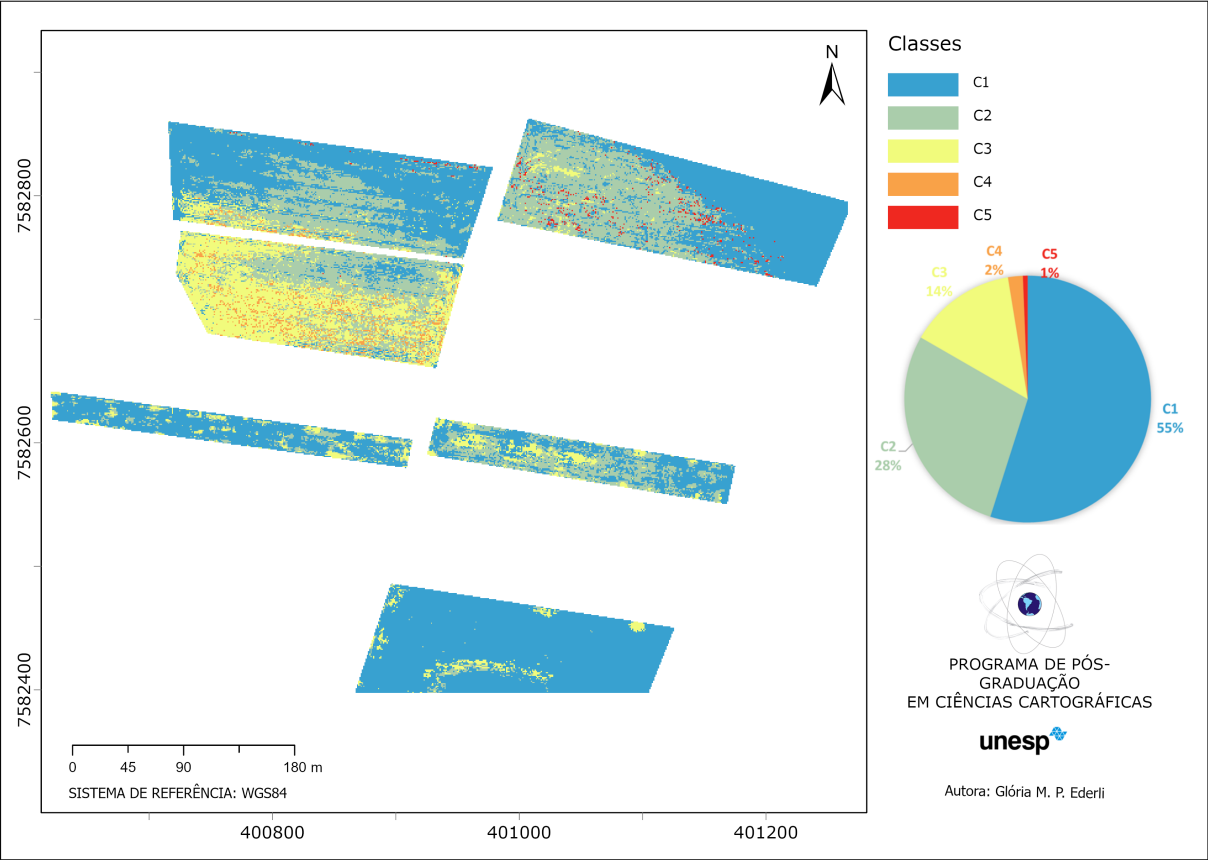
Na Figura 36, é apresentado o mapa temático gerado para a classificação da soja parasitada por nematoides, com base no nível de infestação do IFVL. O mapa foi elaborado utilizando o software ArcGIS Pro. Como o resultado da classificação foi obtido a partir de pontos (em que cada ponto representa a média de uma área de 1x1m), esses pontos foram convertidos em um raster.

Com base no mapa de infestação por fitonematoides a partir do índice IFVL (Figura 36), é possível identificar as parcelas que apresentam os maiores níveis de infestação. Nota-se que as parcelas B5, A6 e B6 apresentam os maiores índices de infestação. É interessante destacar que as parcelas A6 e B6 apresentam altos níveis de infestação, representados pela cor vermelha (C5) no mapa, mas com baixa frequência. Por outro lado, a parcela B5 apresenta alta frequência de ocorrência nas classes C4 e C3, representadas pelas cores laranja e amarela, respectivamente.

Já a parcela A1 apresenta o menor nível de infestação, com predominância da classe C1, que representa baixa infestação por fitonematoides. A parcela B3 também apresenta predominância na classe C1, indicando baixa infestação, porém com algumas regiões nas classes C2 e C3, representando um nível moderado de infestação.

Cabe ressaltar que o uso do IFVL como indicador da comunidade de nematoides no solo é uma ferramenta útil na tomada de decisão para o manejo desses patógenos, permitindo identificar as parcelas que requerem maior atenção e intervenção.

Figura 36 – Resultado utilizando o classificador RF e 5 classes.



Fonte: A autora (2023).

5 Conclusões e Recomendações

O objetivo geral deste estudo é avaliar o potencial das imagens hiperespectrais na detecção e classificação da intensidade de infestação de fitonematoides na cultura de soja em uma fazenda experimental localizada em Caiuá, São Paulo. Para alcançar este objetivo, foi adotada uma abordagem de aprendizado de máquina, utilizando um conjunto de amostras de campo e um índice proposto que relaciona a quantidade de nematoides de vida livre e fitonematoides (IFVL).

É importante destacar que o número reduzido de amostras em campo é uma escolha estratégica deste estudo, considerando que a coleta de amostras nematológicas em campo é uma tarefa onerosa e exaustiva. Dessa forma, a limitação do número de amostras contribui para a redução dos custos e das horas de trabalho necessárias para o levantamento dos dados.

Neste estudo, foi apresentado uma abordagem para utilizar imagens multiespectrais como uma ferramenta para auxiliar no planejamento de um levantamento aéreo por VANT em uma fazenda experimental, utilizando uma câmera hiperespectral de alta resolução espacial. As imagens da missão Sentinel-2 foram utilizadas principalmente para validar se o período fenológico em que as imagens do sensor hiperespectral foram capturadas foi o mais adequado para extrair as informações mais relevantes para o estudo, conforme apresentado na Seção 4.1. Os resultados demonstraram que as respostas espectrais das plantas de soja parasitadas com nematoides ficam em maior evidência no período fenológico no final da fase reprodutiva. Com isso, foi possível determinar que a data de captura das imagens com o sensor hiperespectral é satisfatória para a extração das informações espectrais das plantas de soja parasitadas com nematoides.

Além disso, foram analisadas as curvas espectrais das médias das respostas espectrais para cada classe (calculadas a partir do IFVL) do sensor hiperespectral utilizado (Seção 4.2). Verificamos que as classes possuem separabilidade entre elas, o que indica que os classificadores de aprendizado de máquina conseguem distingui-las com maior precisão e que o IFVL é um índice válido para ser utilizado para dividir as classes.

Depois de extrair as informações das imagens obtidas por técnicas de sensoriamento remoto, foram testados três algoritmos baseados em aprendizado de máquina comumente utilizados em trabalhos realizados dentro do tópico agricultura de precisão, o RF, o SVM e o KNN. Esses algoritmos foram testados com diferentes valores de parâmetros. Também foram testados diferentes números de classes para representar o resultados de nível de infestação de fitonematoides a partir do IFVL, que foram 2, 5 e 10 classes.

Após a análise dos resultados obtidos neste estudo, pode-se concluir que o algoritmo RF com 5 classes, levando em consideração a quantidade de amostras de campo disponíveis,

foi o mais adequado para mapear a intensidade de infestação de fitonematoides na cultura de soja, como apresentado na Seção 4.3.2. A escolha desse modelo baseou-se principalmente na acurácia balanceada, que foi de 96,5%. Esses resultados indicam que o RF é um algoritmo eficaz para trabalhar com dados de alta dimensionalidade e complexidade, como os provenientes de Sensoriamento Remoto e Fotogrametria. Seus resultados têm sido consistentes na literatura, e sua habilidade em lidar com dados de alta dimensionalidade e em fornecer estimativas da importância relativa das variáveis na análise o tornam uma ferramenta importante para pesquisas na área.

A partir do algoritmo RF, foi gerado um gráfico que apresenta as variáveis por ordem de importância que contribuíram para o desempenho deste modelo. Os resultados mostraram que as quatro variáveis mais influentes na classificação foram as bandas com comprimentos de onda centrais de 790,30 nm, 751,06 nm, 771,87 nm e 701,27 nm, respectivamente. Todas essas bandas pertencem à região do *red edge* do espectro eletromagnético, que é conhecida por ser sensível a mudanças no conteúdo de clorofila e na estrutura da vegetação. Esses resultados são consistentes com estudos anteriores que mostram a importância das bandas do *red edge* para a detecção de estresse e doenças em plantas.

Além disso, o IFVL apresentou resultados satisfatórios para indicar o nível de infestação de fitonematoides, considerando-se também os nematoides de vida livre presentes na região. Isso pode ser um fator importante a ser considerado em estudos futuros, já que esses tipos de nematoides podem gerar um certo equilíbrio.

É importante ressaltar que a limitação do número de amostras de campo pode ter afetado os resultados obtidos. Porém, isso também indica que a utilização de dados de Sensoriamento Remoto e Fotogrametria pode ser uma alternativa interessante para reduzir custos e horas de trabalho.

Em conclusão, este estudo contribui para a aplicação de técnicas de agricultura de precisão na detecção e classificação da intensidade de infestação de fitonematoides na cultura de soja, utilizando dados de Sensoriamento Remoto e Fotogrametria e algoritmos de aprendizado de máquina.

Como recomendações futuras para o desenvolvimento de trabalhos com interesses similares, é importante explorar e investigar índices de vegetação que utilizem as bandas que se mostraram mais eficientes na detecção de áreas parasitadas por fitonematoides, bem como diferentes sensores, como o sensor multiespectral. Além disso, é interessante considerar um número maior de espécies de nematoides e analisar se algumas delas têm maior influência nas respostas do que outras.

No que diz respeito aos estudos de algoritmos, é válido testar com diferentes valores de parâmetros e com mais algoritmos, incluindo abordagens de aprendizado profundo.

Outra sugestão é aumentar o tamanho do conjunto de dados de treinamento para melhorar a capacidade de generalização do modelo e avaliar o desempenho em diferentes áreas geográficas e em diferentes safras agrícolas.

Referências

- ADELABU, S.; MUTANGA, O.; ADAM, E. Evaluating the impact of red-edge band from rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 95, p. 34–41, 2014. Citado 2 vezes nas páginas 40 e 41.
- ALMEIDA, R. d. O. et al. Predictive models for soybean yields using unmanned aerial vehicles and machine learning techniques. *Precision Agriculture*, Springer, v. 19, n. 4, p. 593–608, 2018. Citado na página 78.
- ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Taylor Francis, v. 46, n. 3, p. 175–185, 1992. Citado 2 vezes nas páginas 42 e 43.
- ARAGAO, A.; CONTINI, E. *O AGRO NO BRASIL E NO MUNDO: UMA SÍNTESE DO PERÍODO DE 2000 A 2020*. 2020. <<https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/pimenta/producao/doencas-e-pragas/doencas/nematoides>>. Citado na página 17.
- ARANTES, B. H. T. et al. Spectral detection of nematodes in soybean at flowering growth stage using unmanned aerial vehicles. *Ciência Rural*, SciELO Brasil, v. 51, 2021. Citado 3 vezes nas páginas 19, 22 e 24.
- ARRIBAS, J. I. et al. Comparison of different algorithms for varietal classification of grapevine by leaf spectral reflectance. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 14, n. 3, p. 4903–4918, 2014. Citado na página 43.
- ASD Inc. *ASD. FieldSpec®UV/VNIR: HandHeld Spectroradiometer. User's Guide*. Version 1.19. Boulder, CO, USA, 2003. 72 p. Citado na página 56.
- BARROS, J. F. D. et al. Estimac o da produtividade da soja utilizando t cnicas de sensoriamento remoto. In: *Congresso Brasileiro de Engenharia Agr cola*. [S.l.: s.n.], 2018. p. 1–8. Citado na p gina 42.
- BARSI, J. A.; SCHOTT, J. R.; RAQUENO, N. G. Radiometric and geometric analysis of worldview-2 imagery acquired over dunhuang megadunes. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 6, n. 12, p. 11945–11964, 2014. Citado na p gina 35.
- BELGIU, M.; DR GUT , 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, 2016. Citado 3 vezes nas p ginas 38, 39 e 40.
- BELGIU, M.; DR GUT , L.; STROBL, J. Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using worldview-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 87, p. 205–215, 2014. Citado na p gina 40.
- BERRAR, D. A brief review of the basics of classification. *Environmetrics*, Wiley Online Library, v. 17, n. 5, p. 557–568, 2006. Citado na p gina 42.

- BONGERS, T.; FERRIS, H. Nematode community structure as a bioindicator in environmental monitoring. *Trends in Ecology and Evolution*, Elsevier, London, v. 4, p. 224–228, 1999. Citado na página 17.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 38.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado 2 vezes nas páginas 38 e 59.
- BRODERSEN, K. H. et al. The balanced accuracy and its posterior distribution. In: IEEE. *International Conference on Pattern Recognition*. [S.l.], 2010. p. 3121–3124. Citado na página 62.
- CAI, L. et al. Mapping soil properties using ensemble learning methods: A case study of the soybean fields in northeast china. *Geoderma*, v. 404, p. 115291, 2021. Citado na página 42.
- CAMPOS, I. et al. Combining a water balance model with evapotranspiration measurements to estimate total available soil water in irrigated and rainfed vineyards. *Agricultural Water Management*, Elsevier, v. 165, p. 141–152, 2016. Citado na página 37.
- CARNEIRO, R. G. et al. Uptake and translocation of nitrogen, phosphorus and calcium in soybean infected with *meloidogyne incognita* and *m. javanica*. *Fitopatologia Brasileira*, Sociedade Brasileira de Fitopatologia, v. 27, n. 2, p. 141–150, 2002. Citado na página 29.
- CAVALLARO, G. et al. Extended self-dual attribute profiles for the classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 12, n. 8, p. 1690–1694, 2015. Citado na página 41.
- COLDITZ, R. R. An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing*, MDPI, v. 7, n. 8, p. 9655–9681, 2015. Citado na página 40.
- COLOMINA, I.; MOLINA, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of photogrammetry and remote sensing*, Elsevier, v. 92, p. 79–97, 2014. Citado na página 19.
- COLWELL, J. E. Vegetation canopy reflectance. *Remote sensing of environment*, Elsevier, v. 3, n. 3, p. 175–183, 1974. Citado na página 34.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado 2 vezes nas páginas 41 e 59.
- DIAS, W. et al. Nematóides em soja: identificação e controle. *Embrapa Soja-Circular Técnica (INFOTECA-E)*, Londrina: CNPSo., 2010. Citado 4 vezes nas páginas 17, 27, 28 e 29.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *International workshop on multiple classifier systems*. [S.l.], 2000. p. 1–15. Citado na página 38.
- DINGUIRARD, M.; SLATER, P. Radiometric calibration. *Manual of Remote Sensing*, John Wiley & Sons, v. 3, p. 183–225, 1999. Citado na página 36.

- DU, P. et al. Random forest and rotation forest for fully polarized sar image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 105, p. 38–53, 2015. Citado na página 59.
- ESA, S.-. *ESA*. 2021. Disponível em: <<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>. Citado 2 vezes nas páginas 29 e 30.
- FASSNACHT, F. E. et al. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central european test sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 7, n. 6, p. 2547–2561, 2014. Citado na página 41.
- FEHR, W. R.; CAVINESS, C. E. *Stages of Soybean Development*. Ames, 1977. 11 p. Citado na página 26.
- FERRAZ, L.; BROWN, D. Nematologia de plantas: fundamentos e importância. *Manaus: Norma Editora*, v. 1, p. 251, 2016. Citado 3 vezes nas páginas 17, 27 e 57.
- FOODY, G. M.; MATHUR, A. Toward intelligent training of supervised image classifications: directing training data acquisition for svm classification. *Remote Sensing of Environment*, v. 93, n. 1, p. 107–117, 2004. ISSN 0034-4257. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S003442570400207X>>. Citado na página 41.
- FORMAGGIO, A. R.; SANCHES, I. D. *Sensoriamento remoto em agricultura*. [S.l.]: Oficina de Textos, 2017. Citado 2 vezes nas páginas 25 e 27.
- FRAZIER, R. J. et al. Characterization of aboveground biomass in an unmanaged boreal forest using landsat temporal segmentation metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 92, p. 137–146, 2014. Citado na página 40.
- FREITAS, L. *Introdução à nematologia*. [S.l.]: Universidade Federal de Viçosa, 2006. Citado na página 28.
- GARG, S. et al. Towards a multimodal system for precision agriculture using iot and machine learning. In: IEEE. *2021 12th international conference on computing communication and networking technologies (ICCCNT)*. [S.l.], 2021. p. 1–7. Citado na página 43.
- GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests for land cover classification. *Pattern recognition letters*, Elsevier, v. 27, n. 4, p. 294–300, 2006. Citado na página 39.
- HAAS, J.; BAN, Y. Urban growth and environmental impacts in jing-jin-ji, the yangtze, river delta and the pearl river delta. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 30, p. 42–55, 2014. Citado na página 40.
- HAN, L. et al. A machine learning-based method for yield prediction of soybean based on remote sensing data. *Field Crops Research*, Elsevier, v. 149, p. 11–18, 2013. Citado na página 43.

- HASEEB, A.; SRIVASTAVA, N.; PANDEY, R. The influence of *meloidogyne incognita* on growth, physiology, nutrient concentration and alkaloid yield of *hyoscyamus niger*. *Nematologia Mediterranea*, Firenze University Press, v. 18, n. 2, p. 127–129, 1990. Citado na página 29.
- HEATH, W. et al. The potential use of spectral reflectance from the potato crop for remote sensing of infection by potato cyst nematodes. *Aspects of Applied Biology*, v. 60, p. 185–188, 2000. Citado 3 vezes nas páginas 18, 22 e 24.
- HILLNHÜTTER, C. et al. Use of imaging spectroscopy to discriminate symptoms caused by *heterodera schachtii* and *rhizoctonia solani* on sugar beet. *Precision Agriculture*, Springer, v. 13, n. 1, p. 17–32, 2012. Citado 2 vezes nas páginas 22 e 24.
- HILLNHÜTTER, C. et al. Remote sensing to detect plant stress induced by *heterodera schachtii* and *rhizoctonia solani* in sugar beet fields. *Field Crops Research*, Elsevier, p. 70–77, jan 2011. Citado 2 vezes nas páginas 18 e 19.
- HONKAVAARA, E. et al. Radiometric calibration of a hyperspectral camera—specifications and results. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 64, n. 2, p. 223–230, 2009. Citado 2 vezes nas páginas 35 e 36.
- HONKAVAARA, E. et al. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight uav spectral camera for precision agriculture. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 5, p. 5006–5039, 2013. Citado na página 32.
- INAMASU, R. Y. et al. Agricultura de precisão: um novo olhar. *Embrapa Instrumentação-Livro científico (ALICE)*, São Carlos, SP: Embrapa Instrumentação, 2011. 334 p., 2011. Citado 2 vezes nas páginas 22 e 23.
- IPPOLITI, G. A.; EPIPHANIO, J. C.; SHIMABUKURO, Y. E. Utilização de sensoriamento remoto na previsão de área a ser plantada com culturas de verão em três municípios do estado de são paulo. *CD-ROM do IX Simpósio Brasileiro de Sensoriamento Remoto*, 1998. Citado na página 21.
- JENKINS, W. et al. A rapid centrifugal-flotation technique for separating nematodes from soil. *Plant disease reporter*, v. 48, n. 9, 1964. Citado na página 48.
- JENSEN, J. R. *Remote Sensing of the Environment: An Earth Resource Perspective*. [S.l.]: Prentice Hall, 2007. Citado na página 33.
- JENSEN, J. R.; EPIPHANIO, J. C. N. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. [S.l.]: Parêntese Editora São José dos Campos, 2009. Citado na página 79.
- KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, Elsevier, v. 147, p. 70–90, 2018. Citado na página 37.
- KAMIR, E.; WALDNER, F.; HOCHMAN, Z. Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 160, p. 124–135, 2020. Citado na página 43.

- LARY, D. J. et al. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, Elsevier, v. 7, n. 1, p. 3–10, 2016. Citado na página 18.
- LAUDIEN, R.; BARETH, G.; DOLUSCHITZ, R. Analysis of hyperspectral field data for detection of sugar beet diseases. In: *Proceedings of the EFITA Conference, 2003*. Debrecen, Hungary: [s.n.], 2003. Citado na página 18.
- LAWRENCE, G. W. et al. Remote sensing and precision nematocide applications for *Rotylenchulus reniformis* management in cotton. In: COOK, R.; HUNT, D. J. (Ed.). *Nematology Monographs and Perspectives*. Leiden, Netherlands: Brill, 2004. p. 13–21. Citado na página 18.
- LAWRENCE, R. L.; WOOD, S. D.; SHELEY, R. L. Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (randomforest). *Remote Sensing of Environment*, Elsevier, v. 100, n. 3, p. 356–362, 2006. Citado 2 vezes nas páginas 39 e 41.
- LIAKOS, K. G. et al. Machine learning in agriculture: A review. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 8, p. 2674, 2018. Citado na página 37.
- LU, D.; WENG, Q. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, Taylor & Francis, v. 28, n. 5, p. 823–870, 2007. Citado na página 59.
- LU, P. et al. Physiological effects of meloidogyne incognita infection on cotton genotypes with differing levels of resistance in the greenhouse. *Journal of Nematology*, Society of Nematologists, v. 46, n. 4, p. 352–359, 2014. Citado na página 29.
- MAKYNEN, J. et al. Unmanned aerial vehicle (uav) operated megapixel spectral camera. *Electro-Optical Remote Sensing, Photonic Technologies, and Applications V. SPIE*, v. 8186, p. 81860Y–1–81860Y–9, 2011. Citado na página 33.
- MAKYNEN, J. A lightweight hyperspectral imager. 64 p. Dissertação (Mestrado) — Helsinki University of Technology, Helsinki, Finland, 2009. Disponível em: <<http://lib.tkk.fi/Dipl/2009/urn100105.pdf>>. Citado na página 33.
- MANSOUR, K. et al. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to aisa eagle resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 70, p. 56–65, 2012. Citado na página 41.
- MARIN, D. B. et al. Multispectral radiometric monitoring of bacterial blight of coffee. *Precision Agriculture*, Springer, v. 20, n. 5, p. 959–982, 2019. Citado na página 18.
- MARRA, A. B. Detecção e mapeamento de nematoides na cultura cafeeira por meio de imagens multiespectrais do msi/sentinel-2 e classificação baseada em aprendizado de máquina. Universidade Estadual Paulista (UNESP), 2020. Citado na página 19.
- MARTINS, G. D. *CARACTERIZAÇÃO ESPECTRAL E ESPACIAL DE ÁREAS INFESTADAS POR NEMATOIDES E MIGDOLUS FRYANUS EM CULTURA CANAVIEIRA*. Dissertação (Mestrado) — Unesp, Presidente Prudente, 2013. Dissertação de Mestrado - Curso de Ciências Cartográfica. Citado na página 18.

- MARTINS, G. D.; GALO, M. d. L. B. T.; VIEIRA, B. S. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 10, n. 12, p. 5395–5403, 2017. Citado na página 18.
- MARTINS, G. D. et al. Detecting and mapping root-knot nematode infection in the coffee crop using remote sensing measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 12, n. 5, p. 1325–1333, 2019. Citado 2 vezes nas páginas 22 e 24.
- MELAKEBERHAN, H. Embracing the emerging precision agriculture technologies for site-specific management of yield-limiting factors. *Journal of Nematology*, v. 34, n. 2, p. 185–188, 2002. Citado na página 18.
- MIYOSHI, G. T. Caracterização espectral de espécies de mata atlântica de interior em nível foliar e de copa. Universidade Estadual Paulista (Unesp), 2016. Citado 4 vezes nas páginas 31, 32, 33 e 37.
- MLADENOVA, I. E. et al. Intercomparison of soil moisture, evaporative stress, and vegetation indices for estimating corn and soybean yields over the us. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 10, n. 4, p. 1328–1343, 2017. Citado na página 37.
- MOREIRA, M. A. Fundamentos do sensoriamento remoto e metodologia de aplicação. 4ª edição. Viçosa. Editora UFV, 2012. Citado na página 35.
- MORIYA, É. A. S. Identificação de bandas espectrais para detecção de cultura de cana-de-açúcar sadia e doente utilizando câmara hiperespectral embarcada em vant. Universidade Estadual Paulista (UNESP), 2015. Citado 4 vezes nas páginas 18, 37, 55 e 57.
- NEUMAIER, N. et al. Estádios de desenvolvimento da cultura de soja. *Embrapa Soja-Capítulo em livro científico (ALICE)*, In: BONATO, ER (Ed.) Estresses em soja. Passo Fundo: Embrapa Trigo, 2000., 2000. Citado na página 26.
- NGUYEN, Q. H. et al. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, v. 2021, 2021. Citado na página 60.
- NOVO, E. M. L. d. M. *Sensoriamento Remoto: princípios e aplicações*. São Paulo: Edgard Blucher, 2008. v. 01. 363 p. ISBN 9788521204411. Disponível em: <www.blucher.com.br>. Acesso em: 13 mar. 2021. Citado 2 vezes nas páginas 37 e 42.
- NUTTER, F. et al. Use of remote sensing to detect soybean cyst nematode-induced plant stress. *Journal of Nematology*, v. 34, n. 3, p. 222–231, 2002. Citado 3 vezes nas páginas 22, 24 e 29.
- OLIVEIRA, L. de et al. Conhecendo a fenologia do feijoeiro e seus aspectos fitotécnicos. *Embrapa Arroz e Feijão-Livro técnico (INFOTECA-E)*, Brasília, DF: Embrapa, 2018., 2018. Citado na página 25.

- OLIVEIRA, R. A. et al. Avaliação do corregristo de imagens coletadas por uma câmara hiperespectral de quadro. In: SOCIEDADE BRASILEIRA DE CARTOGRAFIA. *XXVI Congresso Brasileiro de Cartografia (CBC)*. Gramado, 2014. Disponível em: <http://www.cartografia.org.br/cbc/trabalhos/4/585/CT04-98_1404422611.pdf>. Citado 2 vezes nas páginas 32 e 33.
- ONODA, M. *Estudo sobre um algoritmo de árvores de decisão acoplado a um sistema de banco de dados relacional*. 2001. 110p. Tese (Doutorado) — Dissertação (Mestrado)-Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001. Citado na página 40.
- Organização das Nações Unidas (ONU). *Objetivos de Desenvolvimento Sustentável*. acesso em 26 de fevereiro de 2023. <<https://nacoesunidas.org/pos2015/ods>>. Citado 2 vezes nas páginas 22 e 24.
- OTOBONI, C. E. de M. *Precisão em Proteção de plantas Fatec Shunji Nishimura Pompeia/SP*. 2018. Disponível em: <<http://avantagro.com.br/agricultura-de-precisao-no-manejo-de-nematoides/>>. Citado 2 vezes nas páginas 21 e 23.
- PAL, M. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, Taylor & Francis, v. 26, n. 2, p. 217–222, 2005. Citado na página 59.
- PEERBHAY, K. Y.; MUTANGA, O.; ISMAIL, R. Random forests unsupervised classification: The detection and mapping of solanum mauritianum infestations in plantation forestry using hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 8, n. 6, p. 3107–3122, 2015. Citado na página 41.
- PINHEIRO, J. B. *Nematoides*. 2022. <<https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/pimenta/producao/doencas-e-pragas/doencas/nematoides>>. Citado na página 17.
- PLANT, R. E. Site-specific management: the application of information technology to crop production. *Computers and electronics in agriculture*, Elsevier, v. 30, n. 1-3, p. 9–29, 2001. Citado 2 vezes nas páginas 21 e 23.
- PONZONI, F.; SHIMABUKURO, Y.; KUPLICH, T. Sensoriamento remoto da vegetação. atual e ampl. *São Paulo: Oficina de Textos*, 2012. Citado 2 vezes nas páginas 33 e 34.
- PONZONI, F. J.; KUPLICH, T. M.; RUDORFF, B. F. T. Calibration of remote sensing sensors. *Brazilian Archives of Biology and Technology*, Scielo Brasil, v. 50, n. 3, p. 429–443, 2007. Citado 2 vezes nas páginas 35 e 36.
- RIEDEL, W. E.; BLACKMER, T. M. Leaf reflectance spectra of cereal aphid-damaged wheat. *Crop Science*, Wiley Online Library, v. 39, n. 6, p. 1835–1840, 1999. Citado na página 18.
- SAARI, H. et al. Novel miniaturized hyperspectral sensor for uav and space applications. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Proceedings SPIE 7474, Sensors, Systems, and Next-Generation Satellites XIII*. [S.l.], 2009. v. 7474, p. 74741M–74741M–12. Citado 2 vezes nas páginas 32 e 33.

- SAHOO, R. N. et al. *Processing of hyperspectral remote sensing data*. New Delhi: Division of Agriculture Physics, Indian Agriculture Research Institute, 2013. Citado 2 vezes nas páginas 36 e 57.
- SARTONO, B.; SYAFITRI, U. D. Metode pohon gabungan: Solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal. In: *Forum Statistika dan Komputasi*. [S.l.: s.n.], 2010. v. 15, n. 1. Citado na página 59.
- SCHAER, B.; STRAUCH, C.; KNICKEL, K. Marketing sustainable agriculture: An analysis of the potential role of new food supply chains in sustainable rural development. sus-chain. *Sus-Chain, QLK5-CT-2002-01349*, 2002. Citado 2 vezes nas páginas 21 e 23.
- SEELAN, S. K. et al. Remote sensing applications for precision agriculture: A learning community approach. *Remote sensing of environment*, Elsevier, v. 88, n. 1-2, p. 157–169, 2003. Citado 2 vezes nas páginas 22 e 24.
- SHARMA, A. et al. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, IEEE, v. 9, p. 4843–4873, 2020. Citado na página 43.
- SILVA, L. Uma aplicação de árvores de decisão, redes neurais e knn para a identificação de modelos arma não sazonais e sazonais. *Rio de Janeiro. 145p. Tese de Doutorado-Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro*, 2005. Citado na página 38.
- SLATER, P.; HOUGH, M. A comparison of methods for atmospheric correction of landsat data. *Remote Sensing of Environment*, v. 23, p. 195–213, 1987. Citado na página 36.
- SMITH, G.; MILTON, E. The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*, Taylor & Francis, v. 20, n. 13, p. 2653–2662, 1999. Citado na página 36.
- TEILLET, P.; GUINDON, B.; GOODENOUGH, D. On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, v. 16, p. 83–93, 1990. Citado na página 36.
- THOME, K.; BIGGAR, S.; SLATER, P. Atmospheric correction algorithms for the thematic mapper: Evaluation and application. *Remote Sensing of Environment*, v. 78, p. 49–61, 2001. Citado na página 36.
- VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 1999. Citado na página 18.
- VAPNIK, V. N. *Statistical Learning Theory*. [S.l.]: Wiley-Interscience, 1998. Citado na página 41.
- WANG, H. et al. Mapping robinia pseudoacacia forest health conditions by using combined spectral, spatial, and textural information extracted from ikonos imagery and random forest classifier. *remote Sensing*, MDPI, v. 7, n. 7, p. 9020–9044, 2015. Citado na página 40.

WANG, X.; LI, Y.; LI, J. Estimation of soybean yield based on hyperspectral images using machine learning algorithms. In: SPRINGER. *International Conference on Computer and Computing Technologies in Agriculture*. [S.l.], 2020. p. 3–11. Citado 2 vezes nas páginas 60 e 78.

WANG, Y. et al. Evaluation of machine learning algorithms for predicting soybean yield with hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, v. 58, p. 139–149, 2017. Citado 2 vezes nas páginas 42 e 43.

WANG, Z. et al. Comparative study of machine learning algorithms for classification of soybean and weed species using high resolution aerial imagery. *Computers and Electronics in Agriculture*, v. 151, p. 79–89, 2018. Citado na página 42.

WHEELER, T. A.; KAUFMAN, H. W. Relationship of aerial broad band reflectance to *Meloidogyne incognita* density in cotton. *Journal of Nematology*, v. 36, p. 48–57, 2003. Citado na página 18.

XAVIER, L. C. M. et al. Potencial de imagens multiespectrais sentinel 2 na detecção de nematoides na cultura cafeeira. *Revista Brasileira de Geomática*, v. 7, n. 2, p. 095–114, 2019. Citado na página 19.

YI, P. et al. Estimating soybean yield using deep learning from uav-based multispectral images. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 10, n. 10, p. 1498, 2018. Citado na página 42.

ZHANG, M. et al. Detection of stress in tomatoes induced by late blight disease in california, usa, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 4, n. 4, p. 295–310, 2003. Citado na página 18.

ZHANG, N.; WANG, M.; WANG, N. Precision agriculture—a worldwide overview. *Computers and electronics in agriculture*, Elsevier, v. 36, n. 2-3, p. 113–132, 2002. Citado 2 vezes nas páginas 22 e 24.

ZHANG, X. et al. Determination of key canopy parameters for mass mechanical apple harvesting using supervised machine learning and principal component analysis (pca). *Biosystems engineering*, Elsevier, v. 193, p. 247–263, 2020. Citado na página 43.