



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE FITOPATOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM FITOPATOLOGIA

**IMAGENS HIPERESPECTRAIS E APRENDIZADO DE
MÁQUINA NA DETECÇÃO PRECOCE DO ESTRESSE
CAUSADO POR *Xanthomonas euvesicatoria* pv. *perforans* EM
SEMENTES E COTILÉDONES DE TOMATEIRO**

LUCIELLEN DA COSTA FERREIRA

Brasília - DF

2024

LUCIELLEN DA COSTA FERREIRA

**IMAGENS HIPERESPECTRAIS E APRENDIZADO DE MÁQUINA NA DETECÇÃO
PRECOCE DO ESTRESSE CAUSADO POR *Xanthomonas euvesicatoria* pv. *perforans*
EM SEMENTES E COTILÉDONES DE TOMATEIRO**

Tese apresentada à Universidade
de Brasília como requisito para a
obtenção do título de Doutor em
Fitopatologia pelo Programa de
Pós-graduação em Fitopatologia

Orientador

Prof. Maurício Rossato

Coorientador

Dra. Alice Maria Quezado Duval

Brasília - DF

2024

FICHA CATALOGRÁFICA

Ferreira, Luciellen da Costa

Imagens hiperespectrais e aprendizado de máquina na detecção precoce do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em sementes e cotilédones de tomateiro / Luciellen da Costa Ferreira

Brasília, 2024.

Número de páginas p.: 115.

Tese de doutorado. Programa de Pós-graduação em Fitopatologia, Universidade de Brasília, Brasília, DF.

- I. Palavras-chave: hiperespectro, mancha bacteriana, aprendizado de máquina, tomate, *Xanthomonas euvesicatoria* pv. *perforans*.
- II. Universidade de Brasília. PPG/FIT.
- III. Imagens hiperespectrais e aprendizado de máquina na detecção precoce do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em sementes e cotilédones de tomateiro

DEDICATÓRIA

Dedico ao meu amado filho Nicolas, que ainda cresce em meu ventre. Mesmo antes de conhecer seu rosto, você já é fonte de inspiração e força. Que sua chegada ao mundo conceda ainda mais significado à minha jornada.

AGRADECIMENTOS

Agradeço a Deus, cuja infinita bondade, misericórdia e amor são meu sustento. Nele encontro refúgio e fortaleza, fonte de minha esperança e fé.

Ao meu querido e amado esposo Claudio Cardoso Gomes, por sua fé, por todo cuidado, pela paciência, cumplicidade, apoio e principalmente por viver esse sonho ao meu lado.

Aos meus pais Edvan de Sousa Ferreira e Lucilene Alves da Costa de Sousa, por toda dedicação, amor, orações, ensinamentos, valores que me transmitiram, vocês me permitiram alcançar o inimaginável.

Aos meus irmãos e cunhada André Lucas da Costa Ferreira, Elise da Costa Ferreira e Ketelen Pereira de Moura Silva, por todo incentivo, amor e orações.

As minhas avós Ana Alves de Sousa e Nair Ferreira Gomes, por todo cuidado, amor e orações.

Ao meu orientador, Prof. Dr. Mauricio Rossato, pela dedicação, ensinamentos, paciência, compreensão, sua disponibilidade e acolhimento. Sou extremamente grata, pois, tenho orgulho em dizer que acompanhou minha trajetória acadêmica sendo meu orientador desde a graduação.

A minha co-orientadora, Dra Alice Maria Quezado Duval, pelo apoio na condução dos experimentos, paciência, ensinamentos e disponibilidade durante a execução dessa pesquisa.

A Dr^a. Ana Regia Neves por sua disponibilidade, ensinamentos e sugestões de melhorias na condução do trabalho.

À equipe da Embrapa Hortaliças que me acolheu com todo carinho e dedicação, no período de realização dos experimentos, em especial Luana, Fabiana e Sr. Pedro.

Ao Eng. Agrônomo Murilo Magalhães da Agristar do Brasil LTDA, por ter fornecido as sementes do híbrido de tomateiro HMX 7885, que fizeram parte dos experimentos dessa pesquisa.

Aos amigos que fiz ao longo da pós-graduação, cada um desempenhou um papel crucial no desenvolvimento dessa pesquisa, bem como tornaram os dias difíceis mais agradáveis e alegres: Ian Carlos, Alice, Leila, Julieth, Marcos, Erivaldo, Izaias, Isabella e em especial Angelica Alves, uma grande amiga que tive a oportunidade de conhecer, agradeço por todo apoio e amizade.

A todos os professores e ao Programa de Pós-Graduação em Fitopatologia da Universidade de Brasília (PPG-FIT).

Ao CNPq pela concessão da bolsa, à FAP/DF por financiar a presente pesquisa.

Trabalho realizado junto ao Departamento de Fitopatologia do Instituto de Ciências Biológicas da Universidade de Brasília, sob orientação do Prof. Dr. Mauricio Rossato, com apoio do Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) e da Embrapa Hortaliças (CNPq).

IMAGENS HIPERESPECTRAIS E APRENDIZADO DE MÁQUINA NA DETECÇÃO PRECOCE DO ESTRESSE CAUSADO POR *Xanthomonas euvesicatoria* pv. *perforans* EM SEMENTES E COTILÉDONES DE TOMATEIRO

LUCIELLEN DA COSTA FERREIRA

TESE APROVADA em 18/11/2024 por:

Prof^a. Dr^a. Marisa Álvares da Silva Velloso Ferreira
Examinador (Membro Interno)

Dr^a. Angela Mehta dos Reis
Examinador (Membro Externo)

Dr. Jayme Garcia Arnal Barbedo
Examinador (Membro Externo)

Dr^a. Adriane Wendland Ferreira
Examinador (Suplente)

Prof. Dr Mauricio Rossato
Orientador (Presidente)

Brasília - DF

2024

Sumário

LISTA DE TABELAS	iii
LISTA DE FIGURAS	iv
RESUMO GERAL	vi
GENERAL ABSTRACT.....	viii
INTRODUÇÃO GERAL	1
Objetivo Geral.....	9
Objetivos específicos	9
Justificativa	10
REFERÊNCIAS	11
CAPÍTULO 1	17
HYPERSPECTRAL IMAGING FOR THE DETECTION OF PLANT PATHOGENS IN SEEDS: RECENT DEVELOPMENTS AND CHALLENGES	17
Abstract.....	18
Introduction	19
Conventional methods for detecting plant pathogens in seeds.....	20
Hyperspectral imaging and artificial intelligence.....	22
Detection of plant pathogens in seeds by hyperspectral imaging.....	26
Final comments and perspectives	32
References	36
CAPÍTULO 2	44
DETECÇÃO DO ESTRESSE CAUSADO POR <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> EM COTILÉDONES DE TOMATEIRO POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA	44
RESUMO	44
1. INTRODUÇÃO	46
2. MATERIAL E MÉTODOS	49
2.1. Produção de tomateiro.....	49
2.2. Cepa de <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> (<i>Xep</i>).....	49
2.3. Inoculação	50
2.4. Captura de imagens hiperespectrais	50
2.5. Análise de dados espectrais.....	51

2.6.	Métodos de classificação.....	51
3.	RESULTADOS.....	53
3.1.	Inoculação	53
3.2.	Análise de dados espectrais da cv. Santa Cruz.....	54
3.3.	Classificação dos dados da cv. Santa Cruz.....	55
3.4.	Análise de dados espectrais HMX 7885.....	58
1.1.	Classificação dos dados HMX 7885	59
4.	DISCUSSÃO.....	62
5.	REFERÊNCIAS	68
CAPÍTULO 3		72
DETECÇÃO DO ESTRESSE CAUSADO POR <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> EM SEMENTES DE TOMATE POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA		72
RESUMO		72
1.	INTRODUÇÃO.....	74
2.	MATERIAL E MÉTODOS.....	78
2.1	Produção de sementes de tomateiro	78
2.2.	Cepa de <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> (<i>Xep</i>).....	78
2.3.	Inoculação	78
2.4.	Extração de sementes	79
2.5.	Inoculação artificial de sementes	79
2.6.	Captura de imagens hiperespectrais	80
2.7.	Análise de dados espectrais.....	80
2.8.	Métodos de classificação.....	81
3.	RESULTADOS.....	82
3.1.	Inoculação	82
3.2.	Análise de dados espectrais (Inoculação por pincel).....	83
3.3.	Análise de dados espectrais (Inoculação por injeção).....	85
3.4.	Análise de dados espectrais (Infiltração artificial de sementes).....	86
3.5.	Classificação dos dados.....	87
4.	DISCUSSÃO.....	90
5.	REFERÊNCIAS	95
6.	CONSIDERAÇÕES FINAIS.....	99

LISTA DE TABELAS

	Página
CAPÍTULO 1: HYPERSPECTRAL IMAGING FOR THE DETECTION OF PLANT PATHOGENS IN SEEDS: RECENT DEVELOPMENTS AND CHALLENGES	
Table 1. Compilation of publications focusing on the utilization of hyperspectral imaging for the detection of plant pathogens in seeds.....	27
CAPÍTULO 2: DETECÇÃO DO ESTRESSE CAUSADO POR <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> EM COTILÉDONES DE TOMATEIRO POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA	
Tabela 1. Desempenho dos modelos na classificação de dados de cotilédones cv. Santa Cruz.....	57
Tabela 2. Desempenho dos modelos na classificação de dados de cotilédones híbrido HMX 7885.....	61
CAPÍTULO 3: DETECÇÃO DO ESTRESSE CAUSADO POR <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> EM SEMENTES DE TOMATE POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA	
Tabela 1. Desempenho dos modelos na classificação de dados de sementes inoculadas por injeção em frutos.....	89
Tabela 2. Desempenho dos modelos na classificação de dados de sementes inoculadas por pincel em frutos.....	89
Tabela 3. Desempenho dos modelos na classificação de dados de sementes inoculadas à vácuo.....	89

LISTA DE FIGURAS

Página

CAPÍTULO 1: HYPERSPECTRAL IMAGING FOR THE DETECTION OF PLANT PATHOGENS IN SEEDS: RECENT DEVELOPMENTS AND CHALLENGES

Figure 1. Compilation of cited publications on hyperspectral imaging for the detection of plant pathogens in seeds. (A) Sample size used in each publication. (B) Method of inoculation for the production of infested/infect seeds. (C) Used methods for the confirmation or quantification of the pathogen within seeds. (D) The intervals of wavelength used for the hyperspectral image production, 400-1000 (VNIR), 700-2500 nm (NIR), 1000-2500 nm (SWIR)..... 31

Figure 2. Sequential stages involved in developing hyperspectral imaging for detecting plant pathogens within seeds. **1.** Production or acquisition of seeds naturally or artificially infected by inoculating seeds, flowers, or fruits. **2.** Image capture utilizing a hyperspectral camera. **3.** Data cube generation by integrating multiple wavelength readings. **4.** Validation of the infection status of each seed through various detection methods, such as culture media or molecular techniques. **5.** Background removal and selection of the region of interest. **6.** Selection of machine learning/deep learning models based on the seeds identified in step 4, taking into account the accuracy of each model. **7.** Testing the developed protocol and algorithm to validate the technique efficacy..... 35

CAPÍTULO 2: DETECÇÃO DO ESTRESSE CAUSADO POR *Xanthomonas euvesicatoria* pv. *perforans* EM COTILÉDONES DE TOMATEIRO POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA

Figura 1. Reflectância espectral de cotilédones de tomateiro (cv. Santa Cruz) sadios e em diferentes estágios de infecção da mancha bacteriana..... 55

Figura 2. Cotilédones de tomateiro (cv. Santa Cruz) em diferentes estágios de infecção da mancha bacteriana. (A) Diferença espectral normalizada. (B) Sensibilidade espectral normalizada..... 55

Figura 3. Desempenho dos modelos de *machine learning* na classificação de cotilédones de tomateiro (cv. Santa Cruz) sadios e infectados em diferentes estágios da mancha bacteriana..... 56

Figura 4. Reflectância espectral de cotilédones de tomateiro (híbrido HMX 7885) sadios e em diferentes estágios de infecção da mancha bacteriana.....	59
Figura 5. Cotilédones de tomateiro (híbrido HMX 7885) em diferentes estágios de infecção da mancha bacteriana. (A) Diferença espectral normalizada. (B) Sensibilidade espectral normalizada.....	59
Figura 6. Desempenho dos modelos de <i>machine learning</i> na classificação de cotilédones de tomateiro (híbrido HMX 7885) sadios e infectados em diferentes estágios da mancha bacteriana.....	60
CAPÍTULO 3: DETECÇÃO DO ESTRESSE CAUSADO POR <i>Xanthomonas euvesicatoria</i> pv. <i>perforans</i> EM SEMENTES DE TOMATE POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA	
Figura 1. Sintomas típicos de mancha bacteriana do tomateiro em frutos inoculados por pincel (A) Frutos imaturos 15 dias pós inoculação por pincel. (B) Frutos maduros.....	82
Figura 2. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com <i>Xep</i> pelo método de pincel em frutos. (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.....	84
Figura 3. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com <i>Xep</i> pelo método de injeção em frutos. (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.....	85
Figura 4. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com <i>Xep</i> pelo método de infiltração artificial (vácuo). (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.....	87
Figura 5. Desempenho dos modelos de <i>machine learning</i> na classificação de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas por diferentes métodos.....	88

RESUMO GERAL

FERREIRA, Luciellen da Costa. **Imagens hiperespectrais e aprendizado de máquina na detecção precoce do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em sementes e cotilédones de tomateiro.** 2024. Número de páginas (115 p). Tese (Doutorado em Fitopatologia) - Universidade de Brasília, Brasília, DF.

O tomate (*Solanum lycopersicum* L.) está entre as culturas de maior importância econômica mundial, sendo a segunda hortaliça mais produzida. A mancha bacteriana do tomateiro é uma doença limitante para a produção, possui como agente causal três espécies pertencentes ao gênero *Xanthomonas*. Em vista da importância econômica e social da cultura do tomate, dos danos e perdas ocasionados pela mancha bacteriana em campos de cultivo a nível mundial e o papel de sementes e mudas como fonte primária de inóculo, este estudo tem por objetivo desenvolver um método de diagnose precoce da mancha bacteriana em cotilédones e sementes de tomate. O estudo visa a aplicação de um sistema de imagem hiperespectral na faixa do visível e infravermelho próximo (VNIR) em conjunto com a análise de dados espectrais por aprendizado de máquina na detecção de alterações causadas pelo estresse devido a infecção por *Xanthomonas euvesicatoria* pv. *perforans* em tomateiro, bem como a aplicação da análise de reflectância espectral para evidenciar faixas do espectro que apresentem variações significativas em função da presença bacteriana na planta. Foi aplicada a técnica de aquisição de imagens hiperespectrais de sementes de tomateiro–livres do patógeno e infectadas por diferentes métodos de inoculação com uma cepa bacteriana de *X. euvesicatoria* pv. *perforans* (*Xep*), juntamente com a análise de dados usando algoritmos de aprendizado de máquina. Obteve-se um alto desempenho com o modelo SVM (Support Vector Machine), com uma acurácia >95% na classificação dos dados de sementes inoculadas por diferentes técnicas em comparação com as não-inoculadas. A abordagem de imagem hiperespectral e aprendizado de máquina também foi empregada na análise de cotilédones de tomateiro infectados por *Xep* em diferentes estágios da doença (assintomático, inicial e tardio). A implementação dessas técnicas resultou na obtenção de um alto desempenho do modelo SVM apresentando acurácia superior a 80% antes mesmo do surgimento dos sintomas, evidenciando excelente capacidade de classificar e distinguir dados de cotilédones de tomateiro saudáveis e infectados. A análise de reflectância espectral tanto com sementes quanto de cotilédones, gerou uma visão ampla das diferenças espectrais marcantes ao longo dos comprimentos de onda do VNIR, causadas em decorrência da infecção bacteriana. Os resultados deste estudo mostram que o uso de técnicas não destrutivas de imagens hiperespectrais e inteligência artificial para a diagnose da mancha

bacteriana do tomateiro são promissoras, e que podem contribuir para evitar introduções do patógeno em novas áreas, evitar surtos da doença, além de permitir um manejo precoce, diminuindo riscos de perdas na produção de mudas e em lavouras.

Palavras-chave: cotilédones, hiperespectro, inteligência artificial, reflectância, sementes, tomate, *Xanthomonas*.

Orientador – Prof. Maurício Rossato – Universidade de Brasília
Coorientador – Dra. Alice Maria Quezado Duval – Embrapa Hortaliças

GENERAL ABSTRACT

FERREIRA, Luciellen da Costa. **Hyperspectral imaging and machine learning for early detection of stress caused by *Xanthomonas euvesicatoria* pv. *perforans* in tomato seeds and cotyledons.** 2024. Number of pages (115 p). Thesis (Doctorate in Phytopathology) - University of Brasília, Brasília, DF.

The tomato (*Solanum lycopersicum* L.) is among the most economically important crops worldwide, being the second most produced vegetable. Bacterial spot of tomato is a limiting disease for production, caused by three species belonging to the genus *Xanthomonas*. Given the economic and social importance of tomato cultivation, the damage and losses caused by bacterial spot in cultivation fields worldwide, and the role of seeds and seedlings as the primary source of inoculum, this study aims to develop a method for early diagnosis of bacterial spot in tomato cotyledons and seeds. The study focuses on the application of a hyperspectral imaging system in the visible and near-infrared range (VNIR) combined with spectral data analysis using machine learning to detect changes caused by stress due to *Xanthomonas euvesicatoria* pv. *perforans* infection in tomato plants. Additionally, spectral reflectance analysis was applied to identify spectral ranges that exhibit significant variations associated with bacterial presence in the plant. Hyperspectral images of pathogen-free tomato seeds and seeds infected through different inoculation methods with a bacterial strain of *X. euvesicatoria* pv. *perforans* (*Xep*) were acquired and analyzed using machine learning algorithms. High performance was achieved with the Support Vector Machine (SVM) model, showing an accuracy of >95% in classifying data from seeds inoculated through different techniques compared to non-inoculated seeds. The hyperspectral imaging and machine learning approach was also employed to analyze tomato cotyledons infected with *Xep* at different stages of the disease (asymptomatic, initial, and late stages). The implementation of these techniques resulted in high performance of the SVM model, with accuracy exceeding 80% even before symptom onset, demonstrating excellent capability to classify and distinguish healthy and infected tomato cotyledon data. Spectral reflectance analysis of both seeds and cotyledons provided a comprehensive view of the significant spectral differences across VNIR wavelengths caused by bacterial infection. The results of this study demonstrate that the use of non-destructive hyperspectral imaging and artificial intelligence techniques for diagnosing tomato bacterial spot is promising. These techniques can help prevent the introduction of the pathogen into new areas, avoid disease outbreaks, and enable early management, thereby reducing the risk of losses in seedling production and in the field.

Keywords: artificial intelligence, cotyledons, hyperspectral, reflectance, seeds, tomato, *Xanthomonas*.

Advisor – Prof. Maurício Rossato – University of Brasília

Co-advisor – Dr. Alice Maria Quezado Duval – Embrapa Vegetables

INTRODUÇÃO GERAL

O gênero *Xanthomonas* compreende espécies patogênicas a mais de 400 hospedeiras botânicas de importância econômica como citros, brássicas, arroz, trigo, feijão, tomate, pimenta, mandioca e banana (Ryan et al. 2011; Timilsina et al. 2020). A mancha bacteriana é uma doença muito importante no cultivo de tomateiro, é causada por espécies distintas pertencentes ao gênero *Xanthomonas*: *X. vesicatoria*, *X. euvesicatoria* pv. *euvesicatoria*, *X. euvesicatoria* pv. *perforans* e *X. hortorum* pv. *gardneri* (Jones et al. 2004, Constantin et al. 2016, Morinière et al. 2020). A ocorrência da doença é favorecida sob condições de umidade elevada e temperaturas entre 22 e 27 °C (Lopes e Quezado-Soares 1997). Em infecções severas pode afetar significativamente a produção, ocasionando em perdas que podem variar entre 23% e 44% no rendimento de frutos, além de perdas indiretas em função da desfolha da planta (Bashan et al. 1985).

A doença pode afetar toda a parte aérea da planta. Nas folhas inicialmente são observadas lesões que possuem coloração marrom-escura e aspecto encharcado, à medida que a bactéria estabelece sua colonização, essas lesões podem coalescer e desenvolver aspecto de mancha (Ritchie 2000; Vallad et al. 2004). As lesões em frutos são circulares e ligeiramente elevadas, no decorrer da infecção as manchas se expandem e tornam-se marrom escuras, com aspecto áspero (Jones et al. 1991). Nas sépalas as lesões são amarronzadas e evoluem para áreas necróticas, enquanto no caule as lesões são estreitas, alongadas, de coloração castanho-claro e ásperas (Osdaghi et al. 2021). Além dos sintomas descritos, foi relatada a ocorrência de *X. euvesicatoria* pv. *perforans* em tomates cultivados em estufa na Itália, que apresentavam descoloração dos tecidos vasculares, amarelecimento, murcha e

lesões externas no caule, sintomatologia típica de necrose da medula por *Pseudomonas* spp. (Aiello et al. 2013).

É possível que os sintomas da mancha bacteriana em folhas de tomateiro sejam, por muitas vezes, confundidos com sintomas de doenças fúngicas como pinta preta (*Alternaria solani*) e mancha alvo (*Corynerspora cassiicola*). No entanto, no caso da mancha bacteriana, sob o microscópio de luz, é possível visualizar o exsudato bacteriano a partir de cortes de tecidos foliares contendo lesões. Os sintomas típicos de pinta preta e mancha alvo se desenvolvem em zonas concêntricas à medida que se estendem no limbo foliar, diferente de como ocorre na mancha bacteriana. Além disso a mancha alvo é caracterizada por lesões distribuídas uniformemente e de coloração mais clara (Strayer-Scherer et al. 2019).

O primeiro relato da mancha bacteriana ocorreu quase que simultaneamente na África do Sul e Estados Unidos em 1920. A bactéria foi inicialmente denominada *Bacterium vesicatorium* (Doidge 1920; Gardner e Kendrick 1921). Anos depois reclassificada como *Xanthomonas vesicatoria* (Dowson, 1939) e posteriormente classificada por Young et al. (1978) como *Xanthomonas campestris* pv. *vesicatoria*. As espécies de *Xanthomonas* causadoras da mancha bacteriana passaram por diversas reclassificações taxonômicas ao longo da história, baseadas em análises bioquímicas e filogenômicas. Atualmente as espécies que causam a doença são denominadas: *Xanthomonas euvesicatoria* pv. *perforans*, *Xanthomonas hortorum* pv. *gardneri*, *Xanthomonas vesicatoria* e *Xanthomonas euvesicatoria* pv. *euvesicatoria* (Vauterin et al. 1990; Stall et al. 1994; Jones et al. 1995; Jones et al. 2004; Potnis et al. 2015; Constantin et al. 2016; Timilsina et al. 2020; Morinière et al. 2020) Além da classificação complexa em diferentes espécies há também a divisão em raças patogênicas (Jones et al. 1995; Kebede et al. 2014; Potnis et al. 2015; Stall et al. 2009). As raças em *Xanthomonas* spp. no patossistema da mancha bacteriana são determinadas com base na reação de hipersensibilidade (HR) de genótipos de tomate em resposta a efetores do

sistema de secreção tipo III da bactéria, secretados em células da hospedeira, reconhecidos por genes de resistência específicos da planta (Stall et al. 2009).

A diversidade de cepas dentro das espécies de *Xanthomonas* que causam a mancha bacteriana possibilita maior adaptabilidade a diferentes regiões e climas, o que contribui para ampla distribuição mundial da doença (Osdaghi et al. 2021). Potnis et al. (2015) relataram a distribuição das espécies de *Xanthomonas* que causam a mancha bacteriana e elucidaram a dinâmica de ocorrência de cada espécie e raça em países produtores de tomate e pimenta. Araújo et al. (2017) determinaram a distribuição das espécies de *Xanthomonas* causando a mancha bacteriana em campos de cultivo de tomateiro no Brasil e evidenciaram a predominância de *X. euvesicatoria* pv. *perforans* raça T3 e *X. hortorum* pv. *gardneri*, apresentando maior prevalência *X. euvesicatoria* pv. *perforans*, mais precisamente correspondendo à 92% das cepas das regiões avaliadas.

O vento e a chuva desempenham um papel importante na dispersão de espécies de *Xanthomonas* a curtas e longas distâncias (Gottwald e Irey 2007; Champoiseau et al. 2009; Jha e Sonti 2009). Além disso, o patógeno pode sobreviver por curtos períodos em restos de cultura infectados, no solo, e de forma epifítica em plantas hospedeiras e plantas invasoras por semanas ou meses a depender das condições ambientais (Jones et al. 1986; Abrahamian et al. 2021). No entanto a disseminação por sementes é um fator determinante para ocorrência de epidemias causadas por *Xanthomonas*, incluindo a mancha bacteriana. As doenças transmitidas por sementes possuem estratégias de manejo muitas vezes limitadas e ineficazes, o que exige ainda mais métodos de detecção que auxiliem no estabelecimento de um controle efetivo (Gitaitis e Walcott 2007). As sementes atuam como fonte de inóculo primário para ocorrência de epidemias em casas de vegetação e no campo e podem ser responsáveis pela introdução de doenças em novas áreas ou países (Gitaitis e Walcott 2007). Além disso, podem levar a ocorrências da doença em viveiros de produção de mudas,

causando perdas na produção de mudas e propiciando que mudas infectadas sirvam como fonte de inóculo para a doença nas lavouras (Abrahamian et al. 2019). A ampla distribuição geográfica pode ser explicada pelo fato da principal forma de disseminação a longas distâncias das espécies causadoras da mancha bacteriana ser através das sementes (Kebede et al. 2014).

É de suma importância a compreensão da dinâmica de infecção em sementes realizada por *Xanthomonas* spp. no tomateiro. Até o momento a localização de *Xanthomonas* spp. em sementes de tomateiro não está totalmente elucidada. Sharma e Agrawal (2014) evidenciaram cepas de *X. euvesicatoria* colonizando tanto a parte externa quanto interna de sementes de tomateiro. A via de transmissão que a bactéria utiliza determina a localização do fitopatógeno nas sementes. Três rotas de transmissão de bactéria por sementes foram evidenciadas: translocação por meio do xilema da planta que leva a ocorrência da introdução do patógeno em sementes através do hilo; transmissão através dos órgãos florais, mais precisamente via pistilo, onde as bactérias deslocam do estigma por meio dos tecidos estilares até alcançar o óvulo, e por fim; contaminação externa das sementes em decorrência do contato com tecidos de frutos sintomáticos (Maude 1996).

Há hipóteses de que a transmissão para sementes seja principalmente em função da infecção bacteriana em órgãos reprodutivos, fato já identificado por diferentes pesquisadores em distintos patossistemas (Dutta et al. 2012; Darrasse et al. 2018; Van Der Wolf et al. 2019). Dutta et al. (2014) investigaram a dinâmica de colonização de *Xanthomonas euvesicatoria* em flores de pimenta e relataram que a inoculação em flores resultou em frutos sintomáticos, no entanto, utilizando a técnica de qPCR (*real-time* PCR) detectaram a bactéria em 39% dos lotes de sementes testados e recuperaram colônias viáveis de 35% desses lotes, tal resultado indica que as flores possam ser um local promissor para entrada do patógeno em sementes e que a infecção floral contribua para a infecção das sementes por *X. euvesicatoria*.

Alguns estudos com diferentes espécies de *Xanthomonas* buscam elucidar a localização da bactéria em sementes, como é o caso de Van Der Wolf et al (2019) que evidenciaram a infecção de *X. campestris* pv. *campestris*, causadora da podridão negra das brássicas, no embrião e endosperma de sementes de couve, demonstrando que tais patógenos são capazes de estabelecer infecções profundas em sementes. A localização de *Xanthomonas citri* pv. *fuscans* em sementes de feijão, foi elucidada por Darrasse et al. (2018), que detectaram a presença da bactéria em diferentes regiões da semente incluindo o eixo embrionário. Estudos relacionados a transmissão de bactérias por sementes foram capazes de evidenciar a presença do patógeno no embrião de sementes. Na cultura da couve Van Der Wolf et al. (2019) evidenciaram a infecção profunda de sementes por *Xanthomonas campestris* pv. *campestris*, a bactéria foi visualizada em associação com o endosperma e embrião de sementes. No entanto, outros estudos mostram que patógenos bacterianos somente colonizam camadas externas ao embrião, endosperma, na parte interna e externa da testa, como no caso de Barak et al. (2002) que relataram a presença de *Xanthomonas campestris* pv. *vitians* em sementes de alface colonizando apenas os tecidos externos do órgão. Já no estudo de Dutta et al. (2012) foi demonstrado que a colonização da bactéria em sementes é dependente da via de infecção, avaliaram a inoculação de *Acidovorax citrulli* via pericarpo ovariano e pistilo na cultura da melancia, e evidenciaram que via pistilo a colonização em sementes ocorria principalmente no embrião, enquanto via pericarpo a infestação em sementes era externa. Tancos et al. (2013) constataram que independente da via de inoculação *Clavibacter michiganensis* subsp. *michiganensis* infectou os tecidos internos de sementes de tomate.

A disseminação da mancha bacteriana por mudas é um potencial risco para o cultivo de tomateiro, principalmente em viveiros de produção. A transferência das mudas de tomate para o campo geralmente é realizada com no mínimo 21 dias pós-semeio (Clemente e

Boiteux 2012). A ocorrência de infecções latentes causadas por *Xanthomonas* spp. em tomateiro pode contribuir para que a doença não seja detectada antes da chegada até o campo, em condições favoráveis ao desenvolvimento do patógeno pode levar a ocorrência de epidemias da mancha bacteriana em lavouras (Abrahamian et al. 2021). Em viveiros de produção de mudas de tomateiro a presença de uma única planta infectada pelo patógeno pode acarretar surtos da doença, devido a densidade de mudas, irrigação por aspersão e proximidade entre bandejas, o que favorece a disseminação da bactéria (Abrahamian et al. 2019; Abrahamian et al. 2021). Segundo um modelo epidemiológico previsto por Abrahamian et al. (2021), *X. euvesicatoria* pv. *perforans* pode ser disseminada a uma distância de mais de 1,5 m em 5 dias, em condições de altas concentrações de inóculo e temperatura em torno de 27°C. No caso de plantas assintomáticas, a disseminação do patógeno é favorecida, podendo assim ser transferido a longas distâncias de viveiros de mudas para lavouras.

Diversos fitopatógenos possuem a capacidade de dispersão por meio de materiais vegetais propagativos, como mudas, tubérculos, bulbos, bulbilhos e sementes, o que torna o processo de diagnose ainda mais crucial devido a necessidade de detecção do microrganismo antes do plantio. No decorrer dos anos, aliado ao desenvolvimento tecnológico, novas técnicas para diagnose de patógenos de plantas, incluindo fitobactérias, foram surgindo e sendo aperfeiçoadas, partindo desde a análise sintomatológica em planta, até as análises sorológicas, diagnose molecular e, mais recentemente, a detecção de fitopatógenos por imagens hiperespectrais e aplicação do aprendizado de máquina (*machine learning*) na análise de dados (Dutta et al. 2014; Feng et al. 2019).

As imagens hiperespectrais aliadas ao aprendizado de máquina automatizam a análise de dados de imagens, tornando mais eficiente a identificação dos comprimentos de onda de interesse e a aplicação prática das informações em larga escala. A utilização de

técnicas não destrutivas/invasivas como: espectroscopia de fluorescência, espectroscopia por VNIR (visible and near-infrared), imagem de fluorescência e imagens hiperespectrais, surgiram como mecanismo eficientes que podem contribuir para detecção de doenças de plantas (Golhani et al. 2018). Os sensores hiperespectrais tem sido bastante utilizado em estudos no âmbito da agricultura na análise de qualidade de sementes, solo e detecção de doenças (Demattê et al. 2016; Feng et al. 2019; Abdulridha et al. 2020).

As imagens hiperespectrais correspondem a uma tecnologia de sensoriamento remoto que capturam informações e possibilitam a análise ao longo de uma vasta gama de comprimentos de onda, permitindo a caracterização de objetos com base em suas assinaturas espectrais únicas (Stuart et al. 2019). Nos estudos com plantas as imagens hiperespectrais são amplamente empregadas no monitoramento de doenças, na previsão de rendimentos agrícolas, em análises de estresses e mudanças fisiológicas causadas por diferentes fatores (Wan et al. 2022). O aprendizado de máquina é um subcampo da inteligência artificial (IA) que possibilita que algoritmos aprendam e façam previsões ou auxiliem na tomada de decisão a partir da inserção de dados. O aprendizado de máquina busca desenvolver modelos que extraiam padrões em grandes volumes de dados e com isso efetuar funções de classificação, segmentação, regressão e predição (Chen et al. 2020). Há diferentes tipos de aprendizado de máquina, como no caso do aprendizado de máquina supervisionado que consiste em construir modelos de predição com base em um conjunto de dados rotulados. Esses dados costumam possuir uma referência, definida por humanos ou validada experimentalmente (Greener et al., 2022; Asnicar et al. 2024). Outra categoria é o aprendizado de máquina não supervisionado em que o sistema permite determinar padrões em dados não rotulados. Nesse caso é possível buscar por agrupamentos ou tendências em dados sem que seja fornecida uma referência prévia, um exemplo clássico dessa aplicação é a análise de *clusters* (Greener et al., 2022; Asnicar et al. 2024). Em determinadas

circunstâncias, as duas técnicas podem ser empregadas de forma conjunta, gerando assim o aprendizado de máquina semi-supervisionado, que pode favorecer a otimização de processos em situações em que a rotulação de dados é dispendiosa (Greener et al. 2022). Distintos modelos de aprendizado de máquina podem ser aplicados na extração de informações de interesse, exibindo diferentes níveis de desempenho de acordo com sua aplicabilidade e tipos de dados fornecidos (Omaye et al. 2024). Alguns algoritmos de aprendizado de máquina são usualmente aplicados no desenvolvimento de modelos de previsão de doenças em plantas (Ahmad et al. 2023). Como é o caso do SVM (Support Vector Machine), PLS-DA (Partial Least-Squares Discriminant Analysis), RF (Random Forest) e DT (Decision Tree) (Dua et al. 2020).

Há uma diversidade de estudos envolvendo a aplicação de imagens hiperespectrais aliada ao aprendizado de máquina na detecção de bactérias em sementes e mudas (Soni et al. 2022). A técnica vem sendo aplicada na detecção de fitobactérias em sementes, como no estudo de Lee et al. (2017) que identificaram a presença de *Acidovorax citrulli* em sementes de melancia utilizando a técnica de imagem hiperespectral raman, e evidenciaram uma acurácia de cerca de 75% na identificação de sementes infectadas pela bactéria. Baek et al. (2019) detectaram a presença de *Burkholderia glumae* em sementes de arroz infectadas através da aplicação de imagens hiperespectrais, utilizando dois ou três comprimentos de onda específicos alcançaram 90% de acurácia na detecção. Zhang et al. (2020) selecionaram sementes de arroz resistentes a praga bacteriana (*Burkholderia glumae*) por meio da aplicação de imagens hiperespectrais e construção de modelos por meio de aprendizado de máquina profundo baseado em rede neural convolucional.

O uso de imagens hiperespectrais e aprendizado de máquina na detecção da mancha bacteriana em mudas de tomateiro já foi relatado por Cunha et al. (2023), que desenvolveram um método de detecção da doença e obtiveram 90% de acurácia na diagnose utilizando o

modelo PLS-DA. Reis-Pereira et al. (2023) aplicaram as imagens hiperespectrais na detecção da mancha bacteriana em mudas de tomate e analisaram os dados utilizando dois algoritmos distintos de aprendizado de máquina obtendo acurácia de 74% na diagnose da doença. A aplicação de imagens hiperespectrais e aprendizado de máquina na identificação de doenças de plantas é uma técnica promissora, não destrutiva, que pode ser implementada a nível de campo e laboratório, sendo uma tecnologia revolucionária na nova era da agricultura.

Objetivo Geral

Este trabalho tem por objetivo desenvolver um método de detecção do estresse causado pela infecção por *Xanthomonas euvesicatoria* pv. *perforans* em sementes e mudas de tomateiro ainda em estágio cotiledonar utilizando imagens hiperespectrais e extração de dados por aprendizado de máquina.

Objetivos específicos

- Detecção do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em cotilédones de tomateiro utilizando imagens hiperespectrais e algoritmos de aprendizado de máquina;
- Detecção do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em sementes de tomateiro utilizando imagens hiperespectrais e modelos de predição de aprendizado de máquina;
- Buscar modelos preditivos de aprendizado de máquina que apresentem melhor desempenho na análise dos dados extraídos das imagens hiperespectrais de cotilédones e sementes de tomate infectados por *Xanthomonas euvesicatoria* pv. *perforans*.

Justificativa

- A produção de mudas é uma etapa primordial para a produção de tomateiro em larga escala, o que torna imprescindível a busca da diagnose precoce de doenças que acometem a cultura como é o caso da mancha bacteriana;
- A principal forma de disseminação a longas distâncias de *Xanthomonas euvesicatoria* pv. *perforans* em tomateiro ocorre por meio de sementes;
- Ainda não há métodos de diagnose por imagem espectral da mancha bacteriana em sementes de tomateiro, o que seria extremamente vantajoso em comparação aos métodos atualmente utilizados para a detecção que exigem mais etapas na realização, além de serem mais dispendiosos.

REFERÊNCIAS

- Abdulridha J, Ampatzidis Y, Qureshi J, Roberts P (2020) Laboratory and UAV-Based Identification and Classification of Tomato Yellow Leaf Curl, Bacterial Spot, and Target Spot Diseases in Tomato Utilizing Hyperspectral Imaging and Machine Learning. *Remote Sensing* 12(17): 2732. doi:10.3390/rs12172732
- Abrahamian P, Jones JB, Vallad GE (2019) Efficacy of copper and copper alternatives for management of bacterial spot on tomato under transplant and field production. *Crop Protection* 126: 104919.
- Abrahamian P, Sharma A, Jones JB, Vallad GE (2021) Dynamics and spread of bacterial spot epidemics in tomato transplants grown for field production. *Plant Disease* 105(3): 566–575. <https://doi.org/10.1094/PDIS-05-20-0945-RE/ASSET/IMAGES/LARGE/PDIS-05-20-0945-RET1.JPEG>
- Aiello D, Scuderi G, Vitale A, Firrao G, Polizzi G, Cirvilleri G (2013) A pith necrosis caused by *Xanthomonas perforans* on tomato plants. *European journal of plant pathology* 137: 29-41.
- Araújo ER, Costa JR, Ferreira MASV, Quezado-Duval AM (2017) Widespread distribution of *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. *Plant Pathology* 66(1): 159–168. doi: <https://doi.org/10.1111/PPA.12543>
- Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N (2024) Machine learning for microbiologists. *Nature Reviews Microbiology* 22(4): 191-205.
- Baek I, Kim MS, Cho BK, Mo C, Barnaby JY, McClung AM, Oh M (2019) Selection of Optimal Hyperspectral Wavebands for Detection of Discolored, Diseased Rice Seeds. *Applied Sciences* 9(5): 1027. <https://doi.org/10.3390/APP9051027>
- Ballvora A, Pierre M, van den Ackerveken G, Schronack S, Rossier O, Ganal M, Lahaye T, Bonas U (2001) Genetic mapping and functional analysis of the tomato Bs4 locus governing recognition of the *Xanthomonas campestris* pv. *vesicatoria* AvrBs4 protein. *Mol. Plant–Microbe Interact.* 14: 629–638.
- Barak JD, Koike ST, Gilbertson RL (2002) Movement of *Xanthomonas campestris* pv. *vitians* in the stems of lettuce and seed contamination. *Plant Pathol* 51:506–12.
- Bashan Y, Assouline I (1983) Complementary bacterial enrichment techniques for the detection of *Pseudomonas syringae* pv. *tomato* and *Xanthomonas campestris* pv. *vesicatoria* in infested tomato and pepper seeds. *Phytoparasitica* 11(3-4):187–193. doi:10.1007/bf02980690
- Bashan Y, Azaizeh M, Diab S, Yunis H, Okon Y (1985) Crop loss of pepper plants artificially infected with *Xanthomonas campestris* pv. *vesicatoria* in relation to symptom expression. *Crop Protection* 4: 77-84

- Cankar K, Kraigher H, Ravnika M, Rupnik M (2005) Bacterial endophytes from seeds of Norway spruce (*Picea abies* L. karst). *FEMS Microbiol Lett* 244:341–345. doi:10.1016/j.femsle.2005.02.008
- Champoiseau P, Rott P, Daugrois JH (2009) Epiphytic Populations of *Xanthomonas albilineans* and Subsequent Sugarcane Stalk Infection Are Linked to Rainfall in Guadeloupe. *Plant Dis* 93:339–346. <https://doi.org/10.1094/PDIS-93-4-0339>
- Chen RC, Dewi C, Huang SW, Caraka RE (2020) Selecting critical features for data classification based on machine learning methods. *Journal of Big Data* 7(1): 52.
- Clemente FMVT, Boiteux LS (2012) *Produção de Tomate para Processamento Industrial 1th Ed.* Embrapa, Brasil, Brasília DF
- Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, Cottyn B (2016) Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathology* 65: 792–806. <https://doi.org/10.1111/ppa.12461>
- Cunha VAG, Hariharan J, Ampatzidis Y, Roberts PD (2023) Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising remote sensing and artificial intelligence. *Biosystems Engineering* 234: 172-186.
- Darrasse A, Barret M, Cesbron S, Compant S, Jacques MA (2018) Niches and routes of transmission of *Xanthomonas citri* pv. *fuscans* to bean seeds. *Plant and Soil* 422(1–2): 115–128. <https://doi.org/10.1007/S11104-017-3329-3/FIGURES/4>
- Demattê JA, Fiorio PR, Araújo SR (2016) Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sensing* 2(8):1998. doi: 10.3390/rs2081998
- Doidge EM (1920) A tomato canker. *Annals of Applied Biology* 7:407-430. doi: 10.1111/j.1744-7348.1921.tb05528.x
- Dowson WJ (1939) On the systematic position and generic names of the Gram negative bacterial plant pathogens. *Zentralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten* 2(100): 177–193
- Dua Y, Kumar V, Singh RS (2020) Comprehensive review of hyperspectral image compression algorithms. *Optical Engineering* 59(9): 090902-090902.
- Dutta B, Avci U, Hahn MG, Walcott RR (2012) Location of *Acidovorax citrulli* in Infested Watermelon Seeds Is Influenced by the Pathway of Bacterial Invasion. *Phytopathology* 102(5):461-8. doi: <https://doi.org/10.1094/PHYTO-10-11-0286-R>
- Dutta B, Gitaitis R, Sanders H, Booth C, Smith S, Langston DB (2014) Role of blossom colonization in pepper seed infestation by *Xanthomonas euvesicatoria*. *Phytopathology* 104(3): 232–239. <https://doi.org/10.1094/PHYTO-05-13-0138-R>

- Dutta B, Gitaitis R, Smith S, Langston D Jr (2014) Interactions of Seedborne Bacterial Pathogens with Host and Non-Host Plants in Relation to Seed Infestation and Seedling Transmission. PLoS ONE 9(6): e99215. <https://doi.org/10.1371/journal.pone.0099215>
- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods* 15, 1–25. doi: 10.1186/s13007-019-0476-y
- Gardner MW, Kendrick JB (1921) Bacterial spot of tomato. *Journal of Agricultural*
- Gitaitis R, Walcott R (2007) The epidemiology and management of seedborne bacterial diseases. *Annual Review of Phytopathology* 45:371–397. doi:10.1146/annurev.phyto.45.062806.094321
- Glassner H, Zchori-Fein E, Yaron S, Sessitsch A, Sauer U, Compant S (2017) Bacterial niches inside seeds of *Cucumis melo* L. *Plant Soil*. doi:10.1007/s11104-017-3175-3
- Golhani K, Balasundram SK, Vadamalai G, Pradhan B (2018) A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture* 5(3):354-371. doi: 10.1016/j.inpa.2018.05.002
- Greener JG, Kandathil SM, Moffat L, Jones DT (2022) A guide to machine learning for biologists. *Nature reviews molecular cell biology* 23(1): 40-55.
- Jha G, Sonti RV (2009) Attack and defense in xanthomonas-rice interactions. *Proc indian Natn Sci Acad* 75:49–68.
- Jibrin MO, Potnis N, Timilsina S, Minsavage GV, Vallad GE, Roberts PD, Jones JB, Goss EM (2018) Genomic Inference of Recombination-Mediated Evolution in *Xanthomonas euvesicatoria* and *X. perforans*. *Applied and environmental microbiology* 84: e136–e118. doi: 10.1128/AEM.00136-18
- Jibrin MO, Timilsina S, Minsavage GV, Vallad GE, Roberts PD, Goss EM, Jones JB (2022) Bacterial Spot of Tomato and Pepper in Africa: Diversity, Emergence of T5 Race, and Management. *Frontiers in Microbiology* 13: 262. <https://doi.org/10.3389/FMICB.2022.835647/BIBTEX>
- Jibrin MO, Timilsina S, Potnis N, Minsavage GV, Shenge KC, Akpa AD, Alegbejo MD, Beed F, Vallad GE, Jones JB (2015) First Report of Atypical *Xanthomonas euvesicatoria* Strains Causing Bacterial Spot of Tomato in Nigeria. *Plant Disease* 99: 415-415. doi: 10.1094/PDIS-09-14-0952-PDN
- Jones J, Pohronezny K, Stall R, Jones J (1986) Survival of *Xanthomonas campestris* pv. *vesicatoria* in Florida on tomato crop residue, weeds, seeds, and volunteer tomato plants. *Phytopathology* 76:430-434. <https://doi.org/10.1094/Phyto-76-430>
- Jones JB (1991) “Bacterial Spot.” In *Compendium of Tomato Diseases*, edited by J. B. Jones et al. 27. APS Press.

- Jones JB, Bouzar H, Somodi GC, Stall RE, Pernezney K, El-Morsy G, Scott JW (1988) Evidence for the preemptive nature of tomato race 3 of *Xanthomonas campestris* pv. *vesicatoria* in Florida. *Phytopathology* 88: 33–38. doi: <https://doi.org/10.1094/PHYTO.1998.88.1.33>
- Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW (2004) Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Systematic and Applied Microbiology* 27:755-62
- Jones JB, Stall RE, Somodi GC, Bouzar H, Hodge NC (1995) A third tomato race of *Xanthomonas campestris* pv. *vesicatoria*. *Plant Disease* 79: 395–398.
- Jones RAC (2021) Global Plant Virus Disease Pandemics and Epidemics. *Plants* 10(2): 233. doi:10.3390/plants10020233
- Jones RAC, Coutts BA, Latham LJ, McKirdy SJ (2008) Cucumber mosaic virus infection of chickpea stands: temporal and spatial patterns of spread and yield-limiting potential. *Plant Pathology* 57(5):842–853. <https://doi.org/10.1111/j.1365-3059.2008.01838.x>
- Kebede M, Timilsina S, Ayalew A, Admassu B, Potnis N, Minsavage GV, Goss EM, Hong JC, Strayer A, Paret M, Jones JB, Vallad GE (2014) Molecular characterization of *Xanthomonas* strains responsible for bacterial spot of tomato in Ethiopia. *Eur. J. Plant Pathol* 140: 677–688. doi: <https://doi.org/10.1007/s10658-014-0497-3>
- Lee H, Kim M, Qin J, Park E, Song YR, Oh CS, Cho BK (2017) Raman Hyperspectral Imaging for Detection of Watermelon Seeds Infected with *Acidovorax citrulli*. *Sensors* 17(10): 2188. doi:10.3390/s17102188
- Lopes CA (2021) *Doenças do tomateiro* 3th Ed. Embrapa, Brasil, Brasília DF
- Lopes CA, Quezado-Soares AM (1997) *Doenças bacterianas das hortaliças: diagnose e controle*. Embrapa, Brasília
- Maude RB (1996) *Seedborne diseases and their control Principles & Practice*. CAB International, Oxon
- Minsavage GV, Dahlbeck D, Whalen MC, Kearny B, Bonas U, Staskawicz BJ, Stall RE (1990) Gene-for-gene relationships specifying disease resistance in *Xanthomonas campestris* pv. *vesicatoria*–pepper interactions. *Mol. Plant–Microbe Interact.* 3: 41–47. doi: 10.1094/MPMI-3-041
- Morinière L, Burlet A, Rosenthal ER, Nesme X, Portier P, Bull CT, Lavire C, Saux MFL, Bertolla F (2020) Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol et al., 2000 emend. Timilsina et al., 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al., 1995. *Systematic and Applied Microbiology* 43: 126087. doi: 10.1016/j.syapm.2020.126087

- Mullis KB, Faloona FA, Scharf S, Saiki RK, Chifre G, Erlich HA (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In: Cold Spring Harbor symposia on quantitative biology. Cold Spring Harbor Laboratory Press 263-273
- Omaye JD, Ogbuju E, Ataguba G, Jaiyeoba O, Aneke J, Oladipo F (2024) Cross-comparative review of Machine learning for plant disease detection: apple, cassava, cotton and potato plants. *Artificial Intelligence in Agriculture*.
- Osdaghi E, Jones JB, Sharma A, Goss EM, Abrahamian P, Newberry EA, Potnis N, Carvalho R, Choudhary M, Paret ML, Timilsina S, Vallad GE (2021) A centenary for bacterial spot of tomato and pepper. *Molecular Plant Pathology*, 22(12):1500-1519. <https://doi.org/10.1111/MPP.13125>
- Reis-Pereira RM, Santos FND, Tavares F, Cunha M (2023) Enhancing host-pathogen phenotyping dynamics: Early detection of tomato bacterial diseases using hyperspectral point measurement and predictive modeling. *Frontiers in Plant Science* 14: 1242201.
- Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, Vallad GE, Jones JB (2015) Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Molecular plant pathology* 16: 907–920. doi: 10.1111/mpp.12244
- Ritchie DF (2000) “Bacterial Spot of Pepper and Tomato.” *The Plant Health Instructor*. Retrieved from American Phytopathology Society. Updated 2007. <https://www.apsnet.org/edcenter/disandpath/prokaryote/pdlessons/Pages/Bacterialsport.aspx>.
- Ryan RP, Vorhölter FJ, Potnis N, Jones JB, Van Sluys MA, Bogdanove AJ, Dow JM (2011) Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews Microbiology* 9(5):344–355. <https://doi.org/10.1038/nrmicro2558>
- Sharma DK, Agrawal K (2014) Incidence and histopathological study of *Xanthomonas axonopodis* pv. *vesicatoria* in tomato (*Lycopersicon esculentum* Mill.) seeds. *Int J Agric Technol* 10:233–42.
- Soni A, Dixit Y, Reis MM, Brightwell G (2022) Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants. *Comprehensive Reviews in Food Science and Food Safety* 21(4): 3717–3745. doi: <https://doi.org/10.1111/1541-4337.12983>
- Stall RE, Beaulieu C, Egel D, Hodge NC, Leite RP, Minsavage GV, Bouzar H, Jones JB, Alvarez AM, Benedict AA (1994) Two genetically diverse groups of strains are included in *Xanthomonas campestris* pv. *vesicatoria*. *International Journal of Systematic Bacteriology* 44: 47–53. <https://doi.org/10.1099/00207713-44-1-47>
- Stall RE, Jones JB, Minsavage GV (2009) Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. *Annu. Rev. Phytopathol.* 47:265–284. doi: 10.1146/annurev-phyto-080508-081752

- Strayer-Scherer A, Liao Y-Y, Abrahamian P, Timilsina S, Paret M, Momol T, Jones J, Vallad GE (2019) Integrated Management of Bacterial Spot on Tomato in Florida. *EDIS* (6)8. <https://doi.org/10.32473/EDIS-PP353-2019>
- Stuart MB, McGonigle AJ, Willmott JR (2019) Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors* 19(14): 3071.
- Tancos MA, Chalupowicz L, Barash I, Manulis-Sasson S, Smart CD (2013) Tomato fruit and seed colonization by *Clavibacter michiganensis* subsp. *michiganensis* through external and internal routes. *Applied and Environmental Microbiology* 79(22): 6948–6957. doi: https://doi.org/10.1128/AEM.02495-13/SUPPL_FILE/ZAM999104858SO1.PDF
- Timilsina S, Potnis N, Newberry E A, Liyanapathirana P, Iruegas-Bocardo F, Branco FF, Goss EM, Jones JB (2020) *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology* 18(8): 415–427. <https://doi.org/10.1038/s41579-020-0361-8>
- Vallad GE, Pernezny KL, Momol T (2004) A Series on Disease in the Florida Vegetable Garden: Tomato. Gainesville: University of Florida Institute of Food and Agricultural Sciences. <https://edis.ifas.ufl.edu/pp121>.
- van der Wolf J, Kastelein P, da Silva Júnior TAF, Lelis FV, van der Zouwen P (2019) Colonization of siliques and seeds of rapid cycling Brassica oleracea plants by *Xanthomonas campestris* pv. *campestris* after spray-inoculation of flower clusters. *European Journal of Plant Pathology* 154: 445–461. <https://doi.org/10.1007/s10658-019-01668-4>
- Vauterin L, Swings J, Kersters K, Gillis M, Mew TW, Schroth MN, Palleroni NJ, Hildebrand DC, Stead DE, Civerolo L, Hayward AC, Maraite H, Stall RE, Vidaver AK, Bradbury JF (1990) Towards an improved taxonomy of *Xanthomonas*. *Int. J. Syst. Evol. Microbiol* 40: 312–316. <https://doi.org/10.1099/00207713-40-3-312>
- Wan L, Li H, Li C, Wang A, Yang Y, Wang P (2022) Hyperspectral sensing of plant diseases: Principle and methods. *Agronomy* 12(6): 1451.
- Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF (1978) A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand Journal of Agricultural Research* 21: 153–177. <https://doi.org/10.1080/00288233.1978.10427397>
- Zhang J, Yang Y, Feng X, Xu H, Chen J, He Y (2020) Identification of Bacterial Blight Resistant Rice Seeds Using Terahertz Imaging and Hyperspectral Imaging Combined With Convolutional Neural Network. *Frontiers in Plant Science* 11: 821. <https://doi.org/10.3389/FPLS.2020.00821/BIBTEX>

CAPÍTULO 1

HYPERSPECTRAL IMAGING FOR THE DETECTION OF PLANT PATHOGENS IN SEEDS: RECENT DEVELOPMENTS AND CHALLENGES

Luciellen da Costa Ferreira¹, Ian Carlos Bispo Carvalho¹, Lúcio André de Castro Jorge²,
Alice Maria Quezado-Duval³ and Maurício Rossato^{1*}

¹University of Brasília, Department of Plant Pathology, Brasília, Brazil, ²Embrapa
Instrumentação, São Carlos, Brazil, ³Embrapa Hortaliças, Brasília, Brazil

Published in Frontiers in Plant Science 15:1387925 (2024)

Abstract

Food security, a critical concern amid global population growth, faces challenges in sustainable agricultural production due to significant yield losses caused by plant diseases, with a multitude of them caused by seedborne plant pathogen. With the expansion of the international seed market with global movement of this propagative plant material, and considering that about 90% of economically important crops grown from seeds, seed pathology emerged as an important discipline. Seed health testing is presently part of quality analysis and carried out by seed enterprises and governmental institutions looking forward to exclude a new pathogen in a country or site. The development of seedborne pathogens detection methods has been following the plant pathogen detection and diagnosis advances, from the use of cultivation on semi-selective media, to antibodies and DNA-based techniques. Hyperspectral imaging (HSI) associated with artificial intelligence can be considered the new frontier for seedborne pathogen detection with high accuracy in discriminating infected from healthy seeds. The development of the process consists of standardization of methods and protocols with the validation of spectral signatures for presence and incidence of contaminated seeds. Concurrently, epidemiological studies correlating this information with disease outbreaks would help in determining the acceptable thresholds of seed contamination. Despite the high costs of equipment and the necessity for interdisciplinary collaboration, it is anticipated that health seed certifying programs and seed suppliers will benefit from the adoption of HSI techniques in the near future.

Keywords: artificial intelligence, bacteria, fungus, *machine learning*, nematode, phytopathogen, seedborne, virus.

Introduction

Food security is currently the focus of agricultural studies in search for sustainable production solutions that keep up with population growth. It is estimated that by 2050 the world will have approximately 10 billion inhabitants (Godfray et al., 2010; Tilman et al., 2011), which makes the demand for food a current challenge. One of the most important limitation for food production are plant diseases which can cause significant losses in productivity and food quality around the world (Savary et al., 2019). Fungal and bacterial pathogens can reduce crop yields by around 15% and plant viruses by a range of 3% to 7%, under favorable conditions such microorganisms can cause losses of up to 100% (Oerke and Dehne, 2004).

Around 90% of crops of economic and food importance are grown by seeds (Maude, 1996), therefore, the seed health is fundamental for agricultural production. Seeds are a primary source of plant pathogen contamination in crop fields and specific tools for identification and early diagnosis of pathogens in seeds are essential to mediate disease management in crops and avoid losses caused by them (Gitaitis and Walcott, 2007; Dutta et al., 2014). Some requirements are desired in the selection of seed health diagnosis methods, such as specificity; sensitivity; efficiency/speed; simplicity; a satisfactory cost-effectiveness; and reliability (Marcinkowska, 2002).

The application of molecular and serological detection methods, despite presenting specificity and sensitivity, are destructive, mostly limited to laboratory application, requiring qualified labor (Kumar and Gupta, 2020). In addition, it does not make it possible to monitor the progress of the disease in the plant or even the action of compounds applied to the crop to control the spread of the pathogen (Mishra et al., 2020). The step forward for a more efficient detection of plant pathogens, focused on quantification and distribution would be the use of hyperspectral images, which could be associated with artificial intelligence, for a faster and more reliable strategy. This article is a state-of-the-art review on the detection of plant

pathogens in seeds, emphasizing the recent advances with the use of hyperspectral images as a resource in the diagnosis of phytopathogens disseminated by seeds.

Conventional methods for detecting plant pathogens in seeds

Initially, the detection of pathogens in seeds was carried out using visual diagnosis methods. Some pathogens, mainly fungi, cause external symptoms in seeds that can be naked-eye visualized or with the aid of a microscope (Gaur et al., 2020). The presence of certain symptoms in seeds may be common among plant pathogens or even the signs may not be visible in the seed tissues (Kumar et al., 2020). Furthermore, the symptoms in seeds may also be of abiotic origins, such as environmental stresses and nutrition to the seed crops or mechanical injuries during the seed harvesting and processing (Bala, 2020).

Directed detection by pathogen cultivation by plaquing the seed on semi-selective or enriched culture media is one of the most used techniques for detecting plant pathogens in seeds (Mancini et al., 2016). Although efficient, they are restricted to the detection of only cultivable microorganisms. After the pathogen growth, identification can be performed by morphology, biochemical and/or molecular tests (Kumar et al., 2020).

ELISA (Enzyme-linked immunosorbent assays), a serological method which demands species-specific antibodies in order not to cause cross-reactions between related and unrelated species which could lead to false negative results (Dewey et al., 1997; Mancini et al., 2016). Moreover, these tests often yield false positives by detecting non-viable microorganisms. This impacts analysis interpretation and directly affects diagnostic accuracy (Walcott, 2003; Afouda et al., 2009).

Molecular techniques, such as nucleic acid amplification using DNA and RNA, offer enhanced sensitivity and specificity in pathogen detection. PCR (Polymerase chain reaction), a pivotal advancement in molecular biology by Mullis et al. (1986), has evolved into various

forms, including Nested PCR, BIO-PCR, qPCR (Quantitative Polymerase Chain Reaction), RT-qPCR (Reverse Transcription Quantitative Polymerase Chain Reaction), and LAMP (Loop-mediated isothermal amplification) (Massung et al., 1998; Notomi et al., 2000; Nazarenko et al., 2002; Bookout et al., 2006).

Conventional PCR is applied in several studies involving the detection of plant pathogens transmitted by seeds. However, these reproductive units are often infected by pathogens at a very low concentrations, which makes detection by PCR unfeasible due to insufficient amount of the target DNA. Based on this problem, BIO-PCR, proven to be effective for bacteria and fungi detection in seeds (Munkvold, 2009) as it allows the pathogen growth by incubation with semi-selective media followed by PCR (Schaad et al., 1995). Compared to conventional PCR, BIO-PCR is more advantageous due to the greater sensitivity, absence of PCR inhibitors (from plant or seed tissues), and detection of only viable cells, eliminating false positives (Marcinkowska, 2002). However, it may have higher costs due to the use of semi-selective media with expensive antibiotics, and delays from pathogen growth in the culture medium (Walcott, 2003; Schena et al., 2004).

Nested PCR is a very sensitive PCR modification, which can be completed in a 24-hour period (Chen et al., 2007). However, it demands additional time and reagents due the two rounds of PCR, and is more prone to contamination than conventional PCR (McCartney et al., 2003; Atkins and Clark, 2004). qPCR and RT-qPCR are other types of PCR which have been used for detecting seed-transmitted plant pathogens due to their sensitivity and speed. While effective, this technique may yield false positives from occasional spikes in background fluorescence or minimal DNA cross-contamination (Robène et al., 2015). Additionally, PCR-based analyses demand specialized and trained technical personnel and expensive equipment, usually being restricted to laboratory settings (Kaur et al., 2020). PCR techniques, including PCR, Nested-PCR, BIO-PCR, qPCR, RT-qPCR and others may be used as a multiplex, capable

of detecting simultaneously more than one seed-borne pathogens with greater sensitivity than conventional and individual PCR (Ha et al., 2009; Robène-Soustrade et al., 2010).

Derived from PCR, Loop-mediated isothermal amplification (LAMP) based technique efficiently detects plant pathogens in seeds (Choudhary et al., 2022), offering several advantages such as: it operates under isothermal conditions, allows direct visualization of amplification results, and can be applied in the field (Notomi et al., 2000). The LAMP assay is more sensitive and specific than PCR, mainly due to the addition of loop primers (Nagamine et al., 2002). However, the handling of samples when adding interleaved dyes, which allow visualization of the result, can influence the occurrence of false positives, due to the release of aerosols of target nucleic acid fragments into the environment (Hardinge and Murray, 2019). Also, as other molecular tests, it may detect not only the living microorganisms but also any remaining pathogen DNA within the sample.

Hyperspectral imaging and artificial intelligence

The human eye can detect only a small portion of the electromagnetic spectrum, distinguishing the spectral responses of images/objects in a restricted range of the visible spectrum between 400 nm and 700 nm (Zwinkels, 2015). To overcome this limitation, optical remote sensing has emerged as a technique capable of obtaining representative data in ranges beyond infrared, typically between 400 nm and 2500 nm (Oliveira et al., 2020). The possibility of capturing the spectral signal in a larger range allows the structuring of reflectance profiles and the detection of patterns that differentiate features between different targets far beyond the visual range (Shanmugapriya et al., 2019).

In recent decades, several studies have involved the application of remote sensing in agriculture both in field and laboratory. The use of optical remote sensing through hyperspectral images is an example (Demattê et al., 2010; Feng et al., 2019; Abdulridha et al., 2020b). These sensors measure the reflectance of an object, such as plant leaves, and identify any changes,

such as those caused by a plant pathogen, as any disturbance to the leaf region would alter the reflectance and direction of the light (Xie et al., 2008).

A HSI (hyperspectral imaging) combines conventional imaging system with spectroscopy. Unlike the RGB (red, green and blue) color system, HSI corresponds to a set of techniques capable of capturing wavelengths beyond red, green and blue (Li et al., 2014). HSI consists of a spectrograph capable of capturing reflectance in a wide range of the spectrum, such as visible (VIS), ultraviolet (UV), near-infrared (NIR), and short-wave infrared (SWIR) (Bock et al., 2010). Such images are three-dimensional (3D), formed by two spatial dimensions (x and y) and one spectral dimension (z). By combining conventional imaging with spectroscopy, HSI obtains complementary information from both fields (Mishra et al., 2017). An example of such a combination is that while spectroscopy collects data about plant physiology, conventional imaging system gathers information on plant structural dynamics (Montes et al., 2007; Bucksch et al., 2014).

Hyperspectral images have hundreds of wavelength bands close together in a spectral range, producing a dense colorfull cube of information, with spatial resolution capable of obtaining several pixels per target (Lowe et al., 2017). The spectral portion generally applied to the study of hyperspectral images of plants ranges from UV starting at ~250 nm to SWIR up to ~2500 nm. The most useful spectral band for plant analysis are the visible and near infrared (Lowe et al., 2017). In the range of 400-700 nm, it is possible to capture changes in leaf pigmentation, whereas at 700-1300 nm range, changes in the cellular structure of the mesophyll. Wider spectral ranges (1300-2500 nm) are needed, for example, for water content in a plant (Peñuelas and Filella, 1998).

The greatest advantage of the application of HSI is that they are techniques capable of detecting and differentiating a disease even in an asymptomatic plant (Martinelli et al., 2015; Whetton et al., 2018). The possibility of early detection through the use of hyperspectral sensors

can contribute to the early and efficient management of diseases, preventing spread in the field (Abdulridha et al., 2020a). Furthermore, hyperspectral imaging technology is non-invasive/destructive, which is advantageous when compared to molecular analysis, where the sample cannot be reused (Sankaran et al., 2010; Golhani et al., 2018).

The data generated is quite collinear and requires different statistical tools to extract information and model patterns (Mishra et al., 2017), such as for artificial intelligence. The application of artificial intelligence using machine learning and deep learning combined with hyperspectral images has been promising in several areas of agricultural management (Demattê et al., 2010; El-Mesery et al., 2019; Feng et al., 2019; Abdulridha et al., 2020b). Artificial intelligence has different definitions, succinctly, it refers to the ability of a machine to perform a function in a similar way to the human mind for “learning” and “problem solving”. Machine learning consists of computational algorithms that learn from input data and can perform classification or clustering tasks, which are ideal for finding patterns and trends in hyperspectral data (Dhakal et al., 2023).

Machine learning is a subfield of artificial intelligence and deep machine learning is a subset that integrates machine learning (Nguyen et al., 2019). Deep learning is excellent for analyzing and extracting useful data in large quantities or collected from different sources (Zhang et al., 2020). The analysis of hyperspectral images using deep learning with convolutional neural networks has emerged as a favorable methodology for evaluating and managing various conditions in agricultural crops, mainly aimed at disease diagnostics. Neural networks are designed to learn the spatial features that best classify or quantify a target/object, in the case of plant diseases, classifying and distinguishing, for example, healthy leaves from those infected by pathogens (Kattenborn et al., 2021).

In addition to deep learning, other algorithmic models based on machine learning are applied in hyperspectral data processing. Examples are: PCA (Principal Component Analysis)

and LDA (Linear Discriminant Analysis) used for extracting features of interest, similarities, and dissimilarities in the data (Dua et al., 2020; Hsieh and Kiang, 2020), and SVM (Support Vector Machine), for solving linear classification problems, which are widely employed for HSI data classification due to their ability to effectively separate heterogeneous samples on the mapped plane (Kale et al., 2017). Although SVM was designed to decipher linear datasets, its usage can be extended to non-linear data when combined with kernel methods (Melgani and Bruzzone, 2004).

Segmentation, an important step before delivering the data cube to test a model, is applied for noise removal from the images, extraction of useful and non-redundant features and establishing the relationship between similar data points in the sample. Therefore, enabling the selection of regions of interest (ROIs), thereby allowing for more accurate data classification. However, segmentation requires the use of methods that demonstrate satisfactory performance for ROI selection. There are algorithms that can assist in this aspect, each with its own peculiarities. Furthermore, some studies clarify which algorithms exhibit better performance according to the type of HSI data to be addressed (Grewal et al., 2023). For seeds, few publications describe the issues associated to those plant material, as the variance of reflectance on each seed, due surface deformities and specular reflection, increasing the difficulties of segmentation (Barbedo et al., 2015).

Another important step is the fusion of spectral and spatial data, which allows for the enhancement of spatial resolution in HSI images, while preserving spectral quality. However, the fusion process is complex, as there is typically a trade-off between preserving spatial data and spectral quality (Mookambiga and Gomathi, 2016). It is from this perspective that artificial intelligence has proven to be efficient in solving problems related to HSI data fusion, such as deep learning, which is ideal for extracting characteristics from complex and abundant data (Li

et al., 2022). Additionally, there is commercial software available that can integrate these steps, simplifying the process.

Detection of plant pathogens in seeds by hyperspectral imaging

The detection of plant pathogens in seeds using hyperspectral images associated with artificial intelligence mechanisms is described for different pathosystems (Table 1). A limited number of studies have been published specifically focusing on the detection of plant pathogens on seeds. The majority of these studies utilize prediction models for data processing, given that alterations on these organs may not be as visually discernible as those on fully grown plants, such as symptoms of necrosis or others. The low number of publications (22) and erratic distribution along the years reveals that it is still an open field to be explored, which probably concerns to the necessity of deep knowledge on the machine and algorithmic possibilities and their high costs.

Table 1. Compilation of publications focusing on the utilization of hyperspectral imaging for the detection of plant pathogens in seeds.

Seeds	Samp les	Pathogen	Inoculation Method*	Reference Method	Spectral Range	Model Classification	Accur acy	Reference
Peanut	-	<i>Aspergillus</i> spp.	Naturally infected	Visually confirmed	970-2570 nm	PCA	98.73 %	(Jiang et al., 2016)
Peanut	-	Diverse fungi species	Naturally infected, favorable conditions	Visually confirmed	967-2499 nm	SVM	>94%	(Qiao et al., 2017)
Peanut	600	<i>Aspergillus flavus</i>	Naturally infected, favorable conditions	Visually confirmed	400-1000 nm	CatBoost, GBDT, XGBoost, LightGBM	>97.42 %	(Wu et al., 2022)
Rice	926	<i>Villosiclava virens</i>	Artificially inoculated/Naturally infected	PCR	874.41-1734.91 nm	PLS-DA, SVM, ELM	>94%	(Wu et al., 2020)
Rice	210	<i>Aspergillus oryzae</i>	Artificially inoculated	Visually confirmed	400-1000 nm	PLSR	-	(Siripatrawan and Makino, 2015)
Rice	47570	<i>Fusarium</i> spp.	Artificially inoculated	Visually confirmed	874.41 - 1734.91 nm	Convolution Neural Networks (CNNs), PLS-DA, SVM	>90%	(Wu et al. 2024)
Oat	-	<i>Fusarium</i> spp.	Naturally infected	Visually confirmed	1000-2500 nm	PLSR, PLS-LDA	-	(Tekle et al., 2015)
Canola	900	<i>Aspergillus glaucus</i>	Artificially inoculated	-	1000-1600 nm	LDA, QDA, MDA	>90%	(Senthilkumar et al., 2015)
Choy Sum	1630	<i>Penicillium decumbens</i>	Artificially inoculated	Visually confirmed	400-1000 nm	SMOTE-siPLS-stacking	>99%	(Xie et al., 2024)
Corn	892	<i>Aspergillus</i> spp.	Naturally infected	Culture media	935-1700 nm	PW-PCA-SVM	100,00 %	(Chu et al., 2020)
Barley	3000	<i>Aspergillus glaucus</i> , <i>Penicillium</i> spp.	Artificially inoculated	-	1000-1600 nm	LDA, QDA, MDA	>80%	(Senthilkumar et al., 2016)
Pea, Bean, Chickpea, Lentil	1500	<i>Aspergillus flavus</i> , <i>Penicillium commune</i>	Artificially inoculated	-	900-1700 nm	LDA, QDA	>96%	(Karuppiyah et al., 2016)
Watermelon	96	Cucumber green mottle mosaic virus	Naturally infected	RT-qPCR	948-2016 nm	PLS-DA	83.30 %	(Lee et al., 2016a)
Watermelon	405	Cucumber green mottle mosaic virus	Naturally infected	RT-qPCR	950-2500 nm	LS-SVM	92%	(Seo et al., 2019)
Watermelon	336	<i>Acidovorax citrulli</i>	Artificially inoculated	Culture media	400-1000 nm	PLS-DA/LS-SVM	> 90%	(Lee et al., 2017b)
Watermelon	48	<i>Acidovorax citrulli</i>	Artificially inoculated	-	400-1800 nm	ANOVA	75%	(Lee et al., 2017a)
Corn	36	<i>Fusarium verticillioides</i>	Naturally infected	Visually confirmed	960-1662 nm	PLS-DA	77%	(Williams et al., 2010)
Wheat	120	<i>Fusarium</i> sp.	Naturally infected	Visually confirmed	1000-1700 nm	PLS-DA	100%	(Serranti et al., 2013)
Wheat	803	<i>Fusarium graminearum</i> , <i>Fusarium meridionale</i>	Naturally infected	Visually confirmed	528-1785 nm	LDA	91%	(Barbedo et al., 2015)

Wheat	21376	<i>Fusarium graminearum</i>	Naturally infected	Visually confirmed	938–1654 nm	LDA, PLS-DA	>92%	(Delwiche et al., 2019)
Wheat	1200	<i>Penicillium</i> spp., <i>Aspergillus glaucus</i> , <i>Aspergillus niger</i>	Artificially inoculated	-	1000-1600 nm	LDA, QDA, MDA	>95%	(Singh et al., 2007)
Wheat	800	<i>Fusarium graminearum</i>	Naturally infected	Visually confirmed	400-1000 nm	LDA	92%	(Shahin and Symons, 2011)

CNN (Convolutional neural network)

ELM (Extreme learning machine)

FDA (Factorial discriminant analysis)

LDA (Linear discriminant analysis)

LS-SVM (Least-squares support vector machines)

MDA (Multiple discriminant analysis)

PCA (Principal component analysis)

PLS-DA (Partial least squares-discriminant analysis)

PLSR (Partial least squares regression)

QDA (Qualitative data analysis)

SVM (Support vector machine)

SVR (Support vector regression)

“-“ Information not provided

*Naturally infected seeds also includes inoculated plants at flower of fruit development stages

The variables tested in these studies include sample size, the method for producing infected seed samples, the reference method for confirming the infection, as well as the wavelengths applied and algorithms utilized. The number of sample size among the publications is variable, from not informing the number of seeds applied in the study up to 47 thousand seeds (Figure 1A). The low number of seeds used to train and test the algorithm can superestimate the accuracy. Several publications utilize artificially inoculated seeds (40,9%), often through the infiltration of propagule suspensions, as positive controls for image generation (Figure 1B), while 54,5% used naturally infected seeds, including those for the detection of mycotoxin-producing fungi. Artificially inoculated or infiltrated seeds are commonly utilized to generate contaminated seed lots. However, this approach may introduce potential inaccuracies in the data, as the concentration of the pathogen could surpass levels typically observed in natural conditions. Wu et al. (2020) states that laboratory-inoculated seeds reflects accurately the detection of field-infected kernels. Nevertheless, such inoculated seeds might not express the impact of infection on the physiology, chemistry, and other compositions of the seed, which could be detected by hyperspectral imaging (HSI), as it can happen to soy seeds infected with *Fusarium verticillioides*, which may change flavonoids content (Pedrozo and Little, 2017). Further studies are required to validate this data. Additionally, the knowledge regarding the threshold limit of propagules on a seed for disease transmission, from germination to seedlings, potentially resulting in disease outbreaks, is crucial. This knowledge impairs the use of the method in properly detecting potentially harmful infected seeds (those that lead to disease transmission). For reference method, which is a method for confirmation of the pathogen presence within the seed, some publications (22.77%) did not use any kind of evaluation, while 54,5% assessed it visually (Figure 1C) and three publications (13,6%) used molecular methods. Concerning the wavelengths applied, VIS, NIR, VNIR, SWIR bands

(ranging from 400 to 2500 nm) were used in several studies, with a predominancy of NIR, occasionally in combination (Figure 1D).

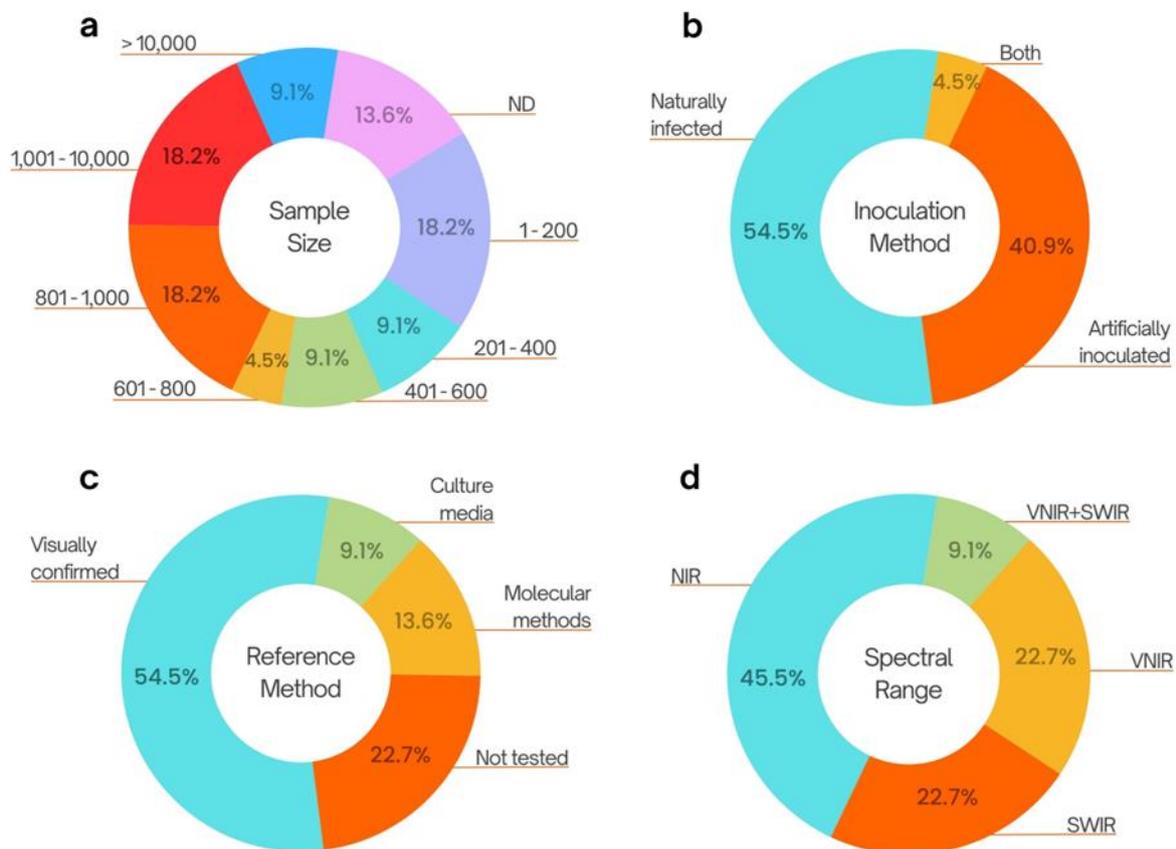


Figure 1. Compilation of cited publications on hyperspectral imaging for the detection of plant pathogens in seeds. **(A)** Sample size used in each publication. **(B)** Method of inoculation for the production of infested/infect seeds. **(C)** Used methods for the confirmation or quantification of the pathogen within seeds. **(D)** The intervals of wavelength used for the hyperspectral image production, 400-1000 (VNIR), 700-2500 nm (NIR), 1000-2500 nm (SWIR).

Most studies focused on the detection of fungi, particularly those associated with the production of mycotoxins, such as *Fusarium* and *Aspergillus* species (Table 1). Mycotoxin accumulation renders grains unfit for human and animal consumption. Few studies address the dispersion of fungi by seeds, which could introduce the pathogen and/or cause an outbreak on production fields. Notably, groups of pathogens such as bacteria and viruses received limited

attention, despite the fact that several important diseases transmitted by seeds fall within these categories. In the case of bacterial pathogens, Lee et al. (2017b) applied hyperspectral imaging in the near-infrared range with two models to classify healthy watermelon seeds and those infected with *Acidovorax avenae* subsp. *citrulli*. The accuracy of both PLS-DA and LS-SVM models was of 91.7% and 90.5%, respectively. Using the Raman hyperspectral imaging technique, they achieved 75% of classification accuracy (Lee et al., 2017a). On its turn, for viruses, using the PLS-DA model, Lee et al. (2016a) distinguished watermelon seeds infected by Cucumber green mottle mosaic virus (CGMMV) with an accuracy of 83.3% while Seo et al. (2019) obtained 92% accuracy using the LS-SVM (Least squares support vector machine) model.

For plant pathogenic fungi, numerous studies have investigated seed detection using hyperspectral images. Notably, some have demonstrated high accuracy in classification. For instance, Chu et al. (2020) achieved 100% accuracy using PCA and SVM models to classify hyperspectral data from both healthy and *Aspergillus* spp. infected corn seeds. Wu et al. (2022) utilized hyperspectral images in the NIR range to identify peanut seeds naturally infected with *Aspergillus flavus*, reaching an accuracy exceeding 97% with three distinct classification models. In Wu et al. (2024) study, the combination of hyperspectral data acquisition and sample classification using convolutional neural networks yielded an accuracy of over 98% in identifying rice seeds infected with *Fusarium* spp. Additionally, employing other learning models like PLS-DA and SVM resulted in accuracies surpassing 90%.

The detection of nematodes in seeds involving hyperspectral is still an open area, as no publication has been found, probably due to the fact that most plant nematode are transmitted by vegetative materials, such as potato tubers, with few species capable of infecting botanic seeds. Žibrat et al. (2021) and Lapajne et al. (2022) evaluated the use of hyperspectral imaging

and the application of different algorithmic models and obtained an accuracy of around 100% in the classification of infected seed potato tubers with *Meloidogyne luci*.

Final comments and perspectives

Due to its non-destructive nature, hyperspectral imaging might be applied to 100% of seeds without the need for sampling or destruction, as the case of molecular methods. Moreover, the utilization of hyperspectral imaging for detection does not preclude the use of molecular diagnosis as shown by Wu et al., 2020; instead, the integration of these methods enhances reliability and specificity in pathogen detection. The application of hyperspectral imaging diagnostics in seeds can expedite the detection of microorganisms, thereby lowering the chances of the introduction of new plant pathogens or variants in a site. However, much remains to be studied in order to enhance the reliability of this technology, aligning it with the standards observed in other scientific domains that already uses hyperspectral imaging.

Several difficulties in this area serve as deterrents for the adoption of this technology, including the high costs of equipment, complex software and methods. Furthermore, the involvement of a multidisciplinary team including plant pathologists, along with electronic engineers, physicists, and software engineers, is strongly recommended to deal with this new research frontier. Unfortunately, it can be observed in the recent published publications that plant pathologists are not present as part of their research teams. With the knowledge of epidemics, agricultural producing systems and seed-borne pathogen infection and detection, such specialist could shorten the period for technology application and adoption by the producing chain authors.

Concerning the generation of contaminated seed lots for the studies, prior to the development of the hyperspectral imaging protocol, assays should be carried out to simulate the most natural conditions of an infected seed. Alternatively, naturally infected seeds obtained through either natural or artificial infection of plants could be utilized. However, this alternative

approach introduces a different challenge: determining the method to validate the presence of the plant pathogen on a specific seed. Another issue is the lack of information regarding the spectral signature of the putative alterations caused to the seed by the presence of the pathogen and its successful transmission to the plant. Besides, sample standards need to be defined for each crop species.

Despite these aspects, the analyzed publications collectively highlighted important approaches that should be considered for the establishment of protocols for hyperspectral imaging detection technologies. Therefore, a seven-step scheme is proposed and recommended here as a template for future studies (Figure 2): obtention of contaminated seeds, preferably naturally infected; images capture; the data cube generation; assessment of the infection status of each seed; segmentation or extraction of the region of interest; model selection and classification based on sample size and accuracy; and finally, testing with a larger sample.

Hyperspectral imaging studies for seed detection are relatively recent, but they hold significant potential for the development of new protocols, particularly in the context of the prevailing era of robotics and artificial intelligence. The adoption of hyperspectral imaging detection technologies stands to offer substantial advantages in the production of healthy seeds, particularly in terms of enhanced reliability and practicality. This technology holds great promise for certification authorities and has the potential to bolster the confidence of farmers in this agricultural inputs, if a substantial amount of them can be checked for the presence of important seed-borne pathogens.

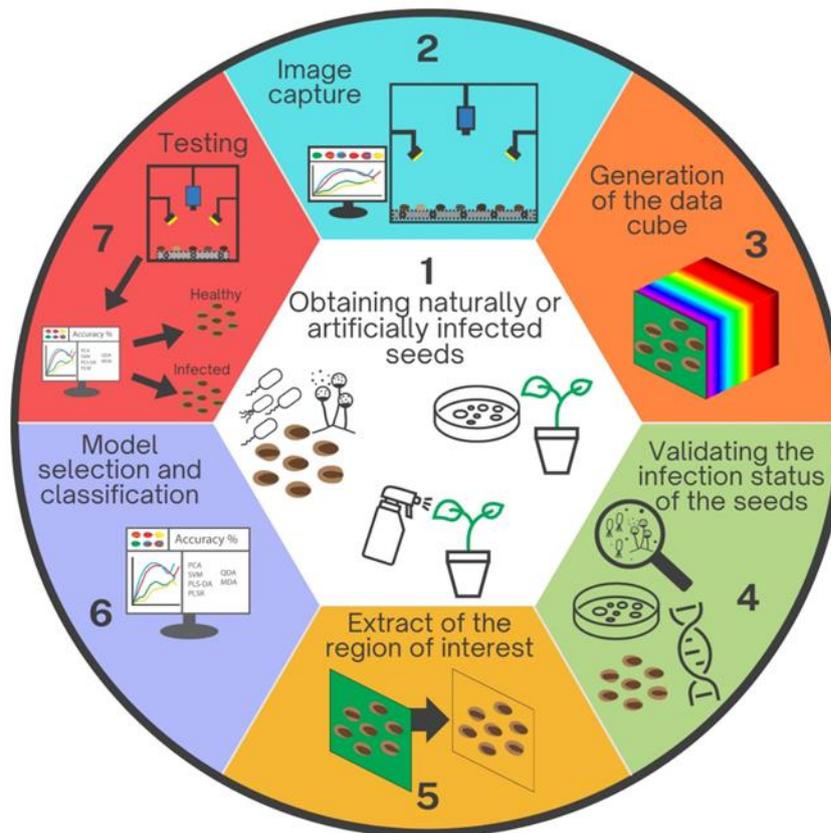


Figure 2. Sequential stages involved in developing hyperspectral imaging for detecting plant pathogens within seeds. **1.** Production or acquisition of seeds naturally or artificially infected by inoculating seeds, flowers, or fruits. **2.** Image capture utilizing a hyperspectral camera. **3.** Data cube generation by integrating multiple wavelength readings. **4.** Validation of the infection status of each seed through various detection methods, such as culture media or molecular techniques. **5.** Background removal and selection of the region of interest. **6.** Selection of machine learning/deep learning models based on the seeds identified in step 4, taking into account the accuracy of each model. **7.** Testing the developed protocol and algorithm to validate the technique efficacy.

Author contributions

LF: Data curation, Writing – original draft. IC: Writing – review & editing. LJ: Writing – review & editing. AQ: Writing – review & editing. MR: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research received funding from the Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF), under grant number 00193-00001058/2021-57.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., Roberts, P. (2020a). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precis. Agric.* 21, 955–978. doi: 10.1007/s11119-019-09703-4
- Abdulridha, J., Ampatzidis, Y., Qureshi, J., Roberts, P. (2020b). Laboratory and UAV-based identification and classification of tomato yellow leaf curl, bacterial spot, and target spot diseases in tomato utilizing hyperspectral imaging and machine learning. *Remote Sens.* 12 2732. doi: 10.3390/rs12172732
- Afouda, L., Wolf, G., Wydra, K. (2009). Development of a sensitive serological method for specific detection of latent infection of *Macrophomina phaseolina* in cowpea. *J. Phytopathol.* 157, 15–23. doi: 10.1111/j.1439-0434.2008.01453.x
- Atkins, S. D., Clark, I. M. (2004). Fungal molecular diagnostics: a mini review. *J. Appl. Genet.* 45, 3–15.
- Bala, A. (2020). “Non-parasitic seed disorders of major agricultural crops,” in *Seed-Borne Diseases of Agricultural Crops: Detection, Diagnosis & Management*. Eds. Kumar, R., Gupta, A. (Springer Nature Pte Ltd, Singapore), 809–820. doi: 10.1007/978-981-32-9046-4_28
- Barbedo, J., Tibola, C., Fernandes, M. (2015). Detecting *Fusarium* head blight in wheat kernels using hyperspectral imaging. *Biosyst. Eng.* 131, 65–76. doi: 10.1016/j.biosystemseng.2015.01.003
- Bock, C. H., Poole, G. H., Parker, P. E., Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *CRC. Crit. Rev. Plant Sci.* 29, 59–107. doi: 10.1080/07352681003617285
- Bookout, A. L., Cummins, C. L., Mangelsdorf, D. J., Pesola, J. M., Kramer, M. F. (2006). High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol.* 73, 1–15. doi: 10.1002/0471142727.mb1508s73
- Bucksch, A., Burrige, J., York, L. M., Das, A., Nord, E., Weitz, J. S., et al. (2014). Image-based high-throughput field phenotyping of crop roots. *Plant Physiol.* 166, 470–486. doi: 10.1104/pp.114.243519
- Chen, Y. Y., Conner, R. L., Gillard, C. L., Boland, G. J., Babcock, C., Chang, K. F., et al. (2007). A specific and sensitive method for the detection of *Colletotrichum lindemuthianum* in dry bean tissue. *Plant Dis.* 91, 1271–1276. doi: 10.1094/PDIS-91-10-1271
- Choudhary, P., Goswami, S. K., Chakdar, H., Verma, S., Thapa, S., Srivastava, A. K., et al. (2022). Colorimetric loop-mediated isothermal amplification assay for detection and ecological monitoring of *Sarocladium oryzae*, an important seed-borne pathogen of rice. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.936766
- Chu, X., Wang, W., Ni, X., Li, C., Li, Y. (2020). Classifying maize kernels naturally infected by fungi using near-infrared hyperspectral imaging. *Infrared Phys. Techn.* 105, 103242. doi: 10.1016/j.infrared.2020.103242

- Demattê, J. A. M., Fiorio, P. R., Araújo, S. R. (2010). Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sens.* 2, 1998–2016. doi: 10.3390/rs2081998
- Delwiche, S. R., Rodriguez, I. T., Rausch, S. R., Graybosch, R. A. (2019). Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging. *J. Cereal Sci.* 87, 18–24. doi: 10.1016/j.jcs.2019.02.008
- Dewey, F. M., Thornton, C. R., Gilligan, C. A. (1997). Use of monoclonal antibodies to detect, quantify and visualize fungi in soils. *Adv. Bot. Res.* 24, 275–308. doi: 10.1016/S0065-2296(08)60076-2
- Dhakar, K., Sivaramakrishnan, U., Zhang, X., Belay, K., Oakes, J., Wei, X., et al. (2023). Machine learning analysis of hyperspectral images of damaged wheat kernels. *Sensors* 23, 3523. doi: 10.3390/s23073523
- Dua, Y., Kumar, V., Singh, R. S. (2020). Comprehensive review of hyperspectral image compression algorithms. *Opt. Eng.* 59, 090902–090902. doi: 10.1117/1.OE.59.9.090902
- Dutta, B., Gitaitis, R., Smith, S., Langston, D. (2014). Interactions of seedborne bacterial pathogens with host and non-host plants in relation to seed infestation and seedling transmission. *PLoS One* 9, e99215. doi: 10.1371/journal.pone.0099215
- El-Mesery, H. S., Mao, H., Abomohra, A. E. F. (2019). Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors* 2019 19, 846. doi: 10.3390/s19040846
- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods* 15, 1–25. doi: 10.1186/s13007-019-0476-y
- Gaur, A., Kumar, A., Kiran, R., Kumari, P. (2020). “Importance of seed-borne diseases of agricultural crops: Economic losses and impact on society,” in *Seed-Borne Diseases of Agricultural Crops: Detection, Diagnosis & Management*. Eds. Kumar, R., Gupta, A. (Springer Nature Pte Ltd, Singapore), 3–23. doi: 10.1007/978-981-32-9046-4_1
- Gitaitis, R., Walcott, R. (2007). The epidemiology and management of seedborne bacterial diseases. *Annu. Rev. Phytopathol.* 45, 371–397. doi: 10.1146/annurev.phyto.45.062806.094321
- Godfray, H. C. J., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Nisbett, N., et al. (2010). The future of the global food system. *Philos. Trans. R. Soc Lond. B. Biol. Sci.* 365, 2769–2777. doi: 10.1098/rstb.2010.0180
- Golhani, K., Balasundram, S. K., Vadamalai, G., Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* 5, 354–371. doi: 10.1016/j.inpa.2018.05.002
- Grewal, R., Kasana, S. S., Kasana, G. (2023). Hyperspectral image segmentation: a comprehensive survey. *Multimed. Tools Appl.* 82, 20819–20872. doi: 10.1007/s11042-022-13959-w

- Ha, Y., Fessehaie, A., Ling, K. S., Wechter, W. P., Keinath, A. P., Walcott, R. R. (2009). Simultaneous detection of *Acidovorax avenae* subsp. *citrulli* and *Didymella bryoniae* in cucurbit seedlots using magnetic capture hybridization and real-time polymerase chain reaction. *Phytopathology* 99, 666–678. doi: 10.1094/PHYTO-99-6-0666
- Hardinge, P., Murray, J. A. H. (2019). Reduced false positives and improved reporting of loop-mediated isothermal amplification using quenched fluorescent primers. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-43817-z
- Hsieh, T. H., Kiang, J. F. (2020). Comparison of CNN algorithms on hyperspectral image classification in agricultural lands. *Sensors* 20, 1734. doi: 10.3390/s20061734
- Jiang, J., Qiao, X., He, R. (2016). Use of Near-Infrared hyperspectral images to identify moldy peanuts. *J. Food Eng.* 169, 284–290. doi: 10.1016/j.jfoodeng.2015.09.013
- Kale, K. V., Solankar, M. M., Nalawade, D. B., Dhumal, R. K., Gite, H. R. (2017). A research review on hyperspectral data processing and analysis algorithms. *Proc. Natl. Acad. sciences India section: Phys. Sci.* 87, 541–555. doi: 10.1007/s40010-017-0433-y
- Karuppiyah, K., Senthilkumar, T., Jayas, D. S., White, N. D. G. (2016). Detection of fungal infection in five different pulses using near-infrared hyperspectral imaging. *J. Stored Prod. Res.* 65, 13–18. doi: 10.1016/j.jspr.2015.11.005
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* 173, 24–49. doi: 10.1016/j.isprsjprs.2020.12.010
- Kaur, S. I., Kashyap, P. L., Kang, S. S., Sharma, A. (2020). “Detection and diagnosis of seed-borne viruses and virus-like pathogens,” in *Seed-Borne Diseases of Agricultural Crops: Detection, Diagnosis & Management*. Eds. Kumar, R., Gupta, A. (Springer Nature Pte Ltd, Singapore), 169–199. doi: 10.1007/978-981-32-9046-4_7
- Kumar, R., Gupta, A. (2020). *Seed-borne diseases of agricultural crops: Detection, diagnosis & management* (Springer Singapore). doi: 10.1007/978-981-32-9046-4
- Kumar, R., Gupta, A., Srivastava, S., Devi, G., Singh, V. K., Goswami, S. K., et al. (2020). “Diagnosis and detection of seed-borne fungal phytopathogens,” in *Seed-Borne Diseases of Agricultural Crops: Detection, Diagnosis & Management*. Eds. Kumar, R., Gupta, A. (Singapore: Springer Nature Pte Ltd), 107–142. doi: 10.1007/978-981-32-9046-4_5
- Lapajne, J., Knapič, M., Žibrat, U. (2022). Comparison of selected dimensionality reduction methods for detection of root-knot nematode infestations in potato tubers using hyperspectral imaging. *Sensors* 22, 367. doi: 10.3390/s22010367
- Lee, H., Kim, M. S., Lim, H. S., Park, E., Lee, W. H., Cho, B. K. (2016a). Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: Application to seeds of the “Sambok Honey” cultivar. *Biosyst. Eng.* 148, 138–147. doi: 10.1016/j.biosystemseng.2016.05.014

- Lee, H., Kim, M. S., Qin, J., Park, E., Song, Y. R., Oh, C. S., et al. (2017a). Raman hyperspectral imaging for detection of watermelon seeds infected with *Acidovorax citrulli*. *Sensors* 17, 2188. doi: 10.3390/s17102188
- Lee, H., Kim, M. S., Song, Y. R., Oh, C. S., Lim, H. S., Lee, W. H., et al. (2017b). Non-destructive evaluation of bacteria-infected watermelon seeds using visible/near-infrared hyperspectral imaging. *J. Sci. Food Agric.* 97, 1084–1092. doi: 10.1002/jsfa.7832
- Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., et al. (2022). Deep learning in multimodal remote sensing data fusion: A comprehensive review. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102926. doi: 10.1016/j.jag.2022.102926
- Li, L., Zhang, Q., Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078
- Lowe, A., Harrison, N., French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 13, 1–12. doi: 10.1186/s13007-017-0233-z
- Mancini, V., Murolo, S., Romanazzi, G. (2016). Diagnostic methods for detecting fungal pathogens on vegetable seeds. *Plant Pathol.* 65, 691–703. doi: 10.1111/ppa.12515
- Marcinkowska, J. Z. (2002). Methods of finding and identification of pathogens in seeds. *Plant Breed. Seed Sci.* 46, 31–48.
- Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. A review. *Advanced methods of plant disease detection. A review. Agron. Sustain. De.* 35, 1–25. doi: 10.1007/s13593-014-0246-1
- Massung, R. F., Slater, K., Owens, J. H., Nicholson, W. L., Mather, T. N., Solberg, V. B. (1998). Nested PCR assay for detection of granulocytic ehrlichiae. *J. Clin. Microbiol.* 36, 1090–1095. doi: 10.1128/JCM.36.4.1090-1095.1998
- Maude, R. (1996). *Seedborne diseases and their control: principles and practice* (Wallingford, UK: CAB International).
- McCartney, H. A., Foster, S. J., Fraaije, B. A., Ward, E. (2003). Molecular diagnostics for fungal plant pathogens. *Pest. Manage. Sci.* 59, 129–142. doi: 10.1002/ps.575
- Melgani, F., Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* 42, 1778–1790. doi: 10.1109/TGRS.2004.831865
- Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosyst. Eng.* 164, 49–67. doi: 10.1016/j.biosystemseng.2017.09.009
- Mishra, P., Polder, G., Vilfan, N. (2020). Close range spectral imaging for disease detection in plants using autonomous platforms: a review on recent studies. *Curr. Robot. Rep.* 1, 43–48. doi: 10.1007/s43154-020-00004-7

- Montes, J. M., Melchinger, A. E., Reif, J. C. (2007). Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* 12, 433–436. doi: 10.1016/j.tplants.2007.08.006
- Mookambiga, A., Gomathi, V. (2016). Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery. *Multidimens. Syst. Signal Process.* 27, 863–889. doi: 10.1007/s11045-016-0415-2
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273. doi: 10.1101/SQB.1986.051.01.032
- Munkvold, G. P. (2009). Seed pathology progress in academia and industry. *Annu. Rev. Phytopathol.* 47, 285–311. doi: 10.1146/annurev-phyto-080508-081916
- Nagamine, K., Hase, T., Notomi, T. (2002). Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Mol. Cell. Probes* 16, 223–229. doi: 10.1006/mcpr.2002.0415
- Nazarenko, I., Lowe, B., Darfler, M., Ikononi, P., Schuster, D., Rashtchian, A. (2002). Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. *Nucleic Acids Res.* 30, 37. doi: 10.1093/nar/30.9.e37
- Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., García, A. L., Heredia, I., et al. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artif. Intell. Rev.* 52, 77–124. doi: 10.1007/s10462-018-09679-z
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28, 63. doi: 10.1093/nar/28.12.e63
- Oerke, E. C., Dehne, H. W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Prot.* 23, 275–285. doi: 10.1016/j.cropro.2003.10.001
- Oliveira, M. R. R. D., Ribeiro, S. G., Mas, J. F., Teixeira, A. D. S. (2021). Advances in hyperspectral sensing in agriculture: a review. *Cienc. Agron.* 51, e20207739. doi: 10.5935/1806-6690.20200096
- Pedrozo, R., Little, C. R. (2017). *Fusarium verticillioides* inoculum potential influences soybean seed quality. *Eur. J. Plant Pathol.* 148, 749–754. doi: 10.1007/s10658-016-1127-z
- Peñuelas, J., Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 3, 151–156. doi: 10.1016/S1360-1385(98)01213-8
- Qiao, X., Jiang, J., Qi, X., Guo, H., Yuan, D. (2017). Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food Chem.* 220, 393–399. doi: 10.1016/j.foodchem.2016.09.119
- Robène, I., Perret, M., Jouen, E., Escalon, A., Maillot, M. V., Chabirand, A., et al. (2015). Development and validation of a real-time quantitative PCR assay to detect *Xanthomonas*

- axonopodis pv. allii from onion seed. *J. Microbiol. Methods* 114, 78–86. doi: 10.1016/j.mimet.2015.04.017
- Robène-Soustrade, I., Legrand, D., Gagnevin, L., Chiroleu, F., Laurent, A., Pruvost, O. (2010). Multiplex Nested PCR for Detection of *Xanthomonas axonopodis* pv. *allii* from Onion Seeds. *Appl. Environ. Microbiol.* 76, 2697. doi: 10.1128/AEM.02697-09
- Sankaran, S., Mishra, A., Ehsani, R., Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13. doi: 10.1016/j.compag.2010.02.007
- Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. doi: 10.1038/s41559-018-0793-y
- Schaad, N. W., Cheong, S. S., Tamaki, S., Hatziloukas, E., Panopoulos, N. J. (1995). A combined biological and enzymatic amplification (BIO-PCR) technique to detect *Pseudomonas syringae* pv. *phaseolicola* in bean seed extracts. *Phytopathology* 85, 243–248. doi: 10.1094/Phyto-85-243
- Schena, L., Nigro, F., Ippolito, A., Gallitelli, D. (2004). Real-time quantitative PCR: A new technology to detect and study phytopathogenic and antagonistic fungi. *Eur. J. Plant Pathol.* 110, 893–908. doi: 10.1007/s10658-004-4842-9
- Senthilkumar, T., Jayas, D. S., White, N. D. G. (2015). Detection of different stages of fungal infection in stored canola using near-infrared hyperspectral imaging. *J. Stored Prod. Res.* 63, 80–88. doi: 10.1016/j.jspr.2015.07.005
- Senthilkumar, T., Jayas, D. S., White, N. D. G., Fields, P. G., Gräfenhan, T. (2016). Detection of fungal infection and Ochratoxin A contamination in stored barley using near-infrared hyperspectral imaging. *Biosyst. Eng.* 147, 162–173. doi: 10.1016/j.biosystemseng.2016.03.010
- Seo, Y., Lee, H., Bae, H. J., Park, E., Lim, H. S., Kim, M. S., et al. (2019). Optimized multivariate analysis for the discrimination of cucumber green mosaic mottle virus-infected watermelon seeds based on spectral imaging. *Biosyst. Eng.* 44, 95–102. doi: 10.1007/s42853-019-00019-9
- Serranti, S., Cesare, D., Bonifazi, G. (2013). The development of a hyperspectral imaging method for the detection of *Fusarium*-damaged, yellow berry and vitreous Italian durum wheat kernels. *Biosyst. Eng.* 115, 20–30. doi: 10.1016/j.biosystemseng.2013.01.011
- Shahin, M. A., Symons, S. J. (2011). Detection of *Fusarium* damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Comput. Electron. Agric.* 75, 107–112. doi: 10.1016/j.compag.2010.10.004
- Shanmugapriya, P., Rathika, S., Ramesh, T., Janaki, P. (2019). Applications of remote sensing in agriculture - A review. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 2270–2283. doi: 10.20546/ijcmas.2019.801.238
- Singh, C. B., Jayas, D. S., Paliwal, J., White, N. D. G. (2007). Fungal detection in wheat using near-infrared hyperspectral imaging. *Trans. ASABE* 50, 2171–2176. doi: 10.13031/2013.24077

- Siripatrawan, U., Makino, Y. (2015). Monitoring fungal growth on brown rice grains using rapid and non-destructive hyperspectral imaging. *Int. J. Food Microbiol.* 199, 93–100. doi: 10.1016/j.ijfoodmicro.2015.01.001
- Tekle, S., Mage, I., Segtnan, V. H., Bjornstad, A. (2015). Near-infrared hyperspectral imaging of fusarium-damaged oats (*Avena sativa* L.). *Cereal Chem.* 92, 73–80. doi: 10.1094/CCHEM-04-14-0074-R
- Tilman, D., Balzer, C., Hill, J., Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108
- Walcott, R. R. (2003). Detection of seedborne pathogens. *Horttechnology* 13, 40–47. doi: 10.21273/HORTTECH.13.1.0040
- Whetton, R. L., Hassall, K. L., Waine, T. W., Mouazen, A. M. (2018). Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. *Biosyst. Eng.* 166, 101–115. doi: 10.1016/j.biosystemseng.2017.11.008
- Williams, P., Manley, M., Fox, G., Geladi, P. (2010). Indirect Detection of *Fusarium Verticillioides* in Maize (*Zea mays* L.) Kernels by near Infrared Hyperspectral Imaging. *J. Near Infrared Spectrosc.* 18, 49–58. doi: 10.1255/jnirs.858
- Wu, N., Jiang, H., Bao, Y., Zhang, C., Zhang, J., Song, W., et al. (2020). Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions. *Sens. Actuators B Chem.* 308, 127696. doi: 10.1016/j.snb.2020.127696
- Wu, N., Weng, S., Xiao, Q., Jiang, H., Zhao, Y., He, Y. (2024). Rapid and accurate identification of bakanae pathogens carried by rice seeds based on hyperspectral imaging and deep transfer learning. *Spectrochim Acta A Mol. Biomol. Spectrosc.* 311, 123889. doi: 10.1016/j.saa.2024.123889
- Wu, Q., Xu, L., Zou, Z., Wang, J., Zeng, Q., Wang, Q., et al. (2022). Rapid nondestructive detection of peanut varieties and peanut mildew based on hyperspectral imaging and stacked machine learning models. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1047479
- Xie, B., Chen, B., Ma, J., Chen, J., Zhou, Y., Han, X., et al. (2024). Rapid identification of Choy sum seeds infected with *Penicillium decumbens* based on hyperspectral imaging and stacking ensemble learning. *Food Anal. Methods* 17, 1–10. doi: 10.1007/s12161-024-02574-0
- Xie, Y., Sha, Z., Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *J. Plant Ecol.* 1, 9–23. doi: 10.1093/jpe/rtn005
- Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., Zhao, C. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sens* 12, 3188. doi: 10.3390/rs12193188
- Žibrat, U., Gerič Stare, B., Knapič, M., Susič, N., Lapajne, J., Širca, S. (2021). Detection of root-knot nematode *Meloidogyne luci* infestation of potato tubers using hyperspectral remote sensing and real-time pcr molecular methods. *Remote Sens* 13, 1996. doi: 10.3390/rs13101996

Zwinkels, J. (2015). "Light, electromagnetic spectrum," in Encyclopedia of Color Science and Technology. Ed. Luo, M. R. (Springer, Berlin, Heidelberg), 1–8. doi: 10.1007/978-3-642-27851-8_204-1

CAPÍTULO 2

DETECÇÃO DO ESTRESSE CAUSADO POR *Xanthomonas euvesicatoria* *pv. perforans* EM COTILÉDONES DE TOMATEIRO POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA

RESUMO

Vistorias e diagnose precoce da mancha bacteriana em mudas comerciais de tomate para processamento industrial são imprescindíveis para mitigar perdas causadas pela doença, além evitar que mudas infectadas sejam fonte de inóculo para a ocorrência de epidemias nas lavouras. A mancha bacteriana causada por *X. euvesicatoria pv. perforans* tem ocorrido em viveiros de produção de mudas de tomate para processamento industrial em diversas regiões de produção brasileiras. O emprego de modelagem preditiva com dados de reflectância espectral pode representar uma abordagem eficiente, ágil e de baixo custo para otimizar a diagnose de doenças de plantas. Este estudo teve por objetivo avaliar o potencial da aplicação da análise de dados hiperespectrais em conjunto com modelos de predição para a detecção da infecção por *Xanthomonas euvesicatoria pv. perforans* em mudas de tomateiro em diferentes estágios de desenvolvimento. Plântulas de tomateiro das cultivares Santa Cruz (OP, para consumo in natura) e HMX 7885 (F1, para processamento industrial) foram inoculadas em fase cotiledonar sob as mesmas condições ambientais. Mudas não inoculadas foram utilizadas como controle. Um sistema de sensor hiperespectral que obtém dados de imagem na faixa do visível e infravermelho próximo (VNIR) foi usado para aquisição dos dados espectrais de cotilédones sadios e infectados. Três modelos distintos de predição (Decision Tree - DT, Random Forest - RF e Support Vector Machine - SVM) foram aplicados na classificação de dados previamente normalizados, obtidos de cotilédones sadios e em diferentes estágios de infecção. O modelo SVM teve maior eficácia, atingindo um alto desempenho na classificação da mancha bacteriana, com acurácia de 96% para o conjunto de dados de ‘Santa Cruz’ e 88% para ‘HMX 7885’, ambos antes do aparecimento dos sintomas. Os resultados obtidos no estudo indicam que a aplicação de imagens hiperespectrais e aprendizado de máquina permite a detecção da infecção por este

importante patógeno antes do desenvolvimento dos sintomas, apresentando vantagem temporal em relação ao geralmente utilizado pelos viveiristas, a detecção visual durante as vistorias.

Palavras-chave: hiperespectro, tomate, *Xanthomonas*.

1. INTRODUÇÃO

O tomateiro (*Solanum lycopersicum* L.), é considerado a segunda hortaliça de maior importância no mundo, ao lado da cultura da batata. A China é destaque na produção de tomate, ocupando o primeiro lugar do *ranking* mundial com mais de 38 milhões de toneladas produzidas no ano de 2021, seguida dos EUA e Índia (FAOSTAT 2021). O Brasil está entre os dez maiores produtores de tomate do mundo, com uma produção de mais de 3 milhões de toneladas (FAOSTAT 2021). No ano de 2022 a estimativa da produção de tomate brasileiro foi de 3,8 milhões de toneladas em uma área plantada de 53.872 ha (IBGE 2022) O estado de Goiás é o maior produtor nacional, sendo responsável por cerca de 30% da produção do país, produzindo aproximadamente 1,1 milhões de toneladas, seguido pelos estados de São Paulo e Minas Gerais (IBGE 2022).

O tomate de consumo in natura geralmente é oriundo da agricultura familiar, enquanto o tomate para processamento industrial necessita de maior tecnificação, além de maior mão de obra temporária, principalmente nos períodos de colheita e seleção de frutos (Clemente e Boiteux 2012). O cultivo de tomate indústria em geral é o mais acometido pela mancha bacteriana, principalmente devido a irrigação por aspersão (pivô central) aplicada em grandes áreas de produção, favorecendo a ocorrência da doença (Klein-Gordon et al. 2021). A mancha bacteriana pode ser causada por até quatro espécies do gênero *Xanthomonas*: *X. vesicatoria*, *X. euvesicatoria* pv. *euvesicatoria*, *X. hortorum* pv. *gardneri* e *X. euvesicatoria* pv. *perforans* (Jones et al. 2004, Constantin et al. 2016, Morinière et al. 2020). Estudos apontam que a espécie com maior predominância no Brasil em ambos, campos de cultivo e mudas de tomate para processamento industrial, é a *X. euvesicatoria* pv. *perforans* (Araújo et al. 2012; Araújo et al. 2017; Neves, 2022). Além disso, a espécie se encontra amplamente distribuída a nível mundial em campos de cultivo de tomate (Potnis et al. 2015; Timilsina et al. 2020). (Jones et al. 2004,

Constantin et al. 2016, Morinière et al. 2020). Estudos apontam que a espécie com maior predominância em campos de cultivo de tomate indústria no Brasil é a *X. euvesicatoria* pv. *perforans* (Araújo et al. 2012; Araújo et al. 2017). Além disso, a espécie se encontra amplamente distribuída a nível mundial em campos de cultivo de tomate (Potnis et al. 2015; Timilsina et al. 2020). A mancha bacteriana causa graves prejuízos na produção de tomateiro desde a etapa de produção de mudas até o cultivo no campo. A doença ocorre em qualquer estágio da cultura, acometendo toda a parte aérea da planta, podendo causar perdas de até 50% na produtividade (Quezado-Duval et al. 2010).

A produção de mudas de tomateiro é uma etapa crucial da cadeia produtiva dessa hortaliça, visto que materiais propagativos podem ser fonte potencial de disseminação de diversas doenças na cultura, como é o caso da mancha bacteriana. O período latente, que corresponde ao intervalo entre a infecção causada pela bactéria e manifestação de sintomas visíveis na planta, pode influenciar na disseminação da mancha bacteriana em campos de produção de tomateiro (Abrahamian et al. 2019). Após a distribuição das mudas infectadas em campo, quando em condições favoráveis ao patógeno, podem ocorrer epidemias nas áreas produtoras de tomateiro. Uma única planta infectada pode resultar em um surto no viveiro, principalmente em função da proximidade entre bandejas, tipo de irrigação, geralmente aspersão, e densidade de mudas (Abrahamian et al. 2021). A espécie *X. euvesicatoria* pv. *perforans* pode ser disseminada a uma distância de mais de 1,5 m em cinco dias, especialmente em condições de temperatura em torno de 27°C e na presença de altas concentrações de inóculo. Quando presente em plantas que ainda não apresentam sintomas visíveis a doença pode ser dispersa a longas distâncias, de viveiros de mudas para lavouras (Abrahamian et al. 2019).

A detecção de doenças em plantas de forma não destrutiva é uma técnica emergente no monitoramento e diagnose de doenças em sistemas agrícolas. Com o avanço do sensoriamento remoto e da era computacional, plataformas autônomas possibilitam a aplicação de soluções de

manejo agrícola, que antes poderiam levar dias de experimentos (Virlet et al. 2017; Shakoor et al. 2017). A aplicação de imagens hiperespectrais é um exemplo de técnica não destrutiva/invasiva promissora, que vem sendo empregada na detecção de doenças em plantas na lavoura e em laboratório (Lee et al. 2016; Polder et al. 2019; Nguyen et al. 2021). As imagens hiperespectrais identificam a interação planta-patógeno por meio das mudanças fisiológicas, alterações na concentração de pigmentos fotossintéticos, conteúdo de água ou mesmo aumento dos metabólitos de resposta de defesa da planta, através de comprimentos de ondas específicos (Wahabzada et al. 2015; Stamford et al. 2023).

A inteligência artificial surgiu como uma ferramenta que contribui para a agilidade na análise de dados de forma robusta. O aprendizado de máquina é um campo da inteligência artificial, usualmente aplicado em estudos que focam a implementação de modelos de classificação de dados gerados com sensores hiperespectrais, em diversas áreas incluindo a agricultura (Dua et al. 2020). Diferentes estudos expõe o uso de imagens hiperespectrais e aprendizado de máquina na diagnose de doenças em tomateiro, trazendo mais celeridade e permitindo o encontro de padrões complexos. Abdulridha et al. (2020) usando o modelo de predição MLP (*Multilayer perceptron neural network*) na análise de dados obtidos por sensor hiperespectral, obtiveram uma acurácia de 99% na diagnose da mancha bacteriana em folhas de tomateiro, em diferentes condições experimentais (campo e laboratório). Cunha et al. (2023) empregaram o modelo PLS-DA na diagnose da mancha bacteriana em mudas de tomateiro e alcançaram uma acurácia de 90% na detecção da doença. O uso de imagens hiperespectrais e dois diferentes algoritmos, SVM e LDA (*Linear Discriminant Analysis*). apresentaram uma acurácia de 74% na diagnose da mancha bacteriana em plantas de tomate (Reis-Pereira et al. 2023).

Em vista da aplicabilidade de técnicas utilizando sensores hiperespectrais e aprendizado de máquina, esse estudo tem por objetivo elaborar um método de diagnose precoce da mancha

bacteriana em cotilédones de tomateiro, usando imagens hiperespectrais e análise de dados por modelos de predição de aprendizado de máquina.

2. MATERIAL E MÉTODOS

2.1. Produção de tomateiro

Os experimentos foram realizados no período de outubro de 2023 a fevereiro de 2024, na Embrapa Hortaliças, Gama, DF (Coordenadas). Foram utilizadas as cultivares Santa Cruz (OP, para consumo *in natura*) e HMX 7885 (F1, para processamento industrial). As mudas foram produzidas em casa de vegetação com cobertura de plástico contendo sistema de ventilação e de nebulização por microaspersores automatizados. A semeadura foi feita em bandejas de poliestireno de 168 células, e as plantas conduzidas até o estágio quatro folhas verdadeiras sob temperatura ambiente variando entre 25°C - 30°C. Foram semeadas pelo menos duas a três sementes por célula, sendo realizado o desbaste cerca de sete dias após a emergência das plântulas. A semeadura foi realizada em substrato comercial (Bioplant®) composto por turfa *Sphagnum* e fibra de coco e adicionado adubo de liberação lenta (Osmocote®) contendo NPK 14-14-14.

2.2. Cepa de *Xanthomonas euvesicatoria* pv. *perforans* (Xep)

A cepa utilizada foi a EH-2022-31, pertencente à coleção de trabalho da Embrapa Hortaliças, Brasília-DF, que corresponde a um isolado oriundo de folíolos sintomáticos de mudas comerciais de tomate para processamento. A identificação da espécie foi realizada por PCR multiplex usando *primers* específicos que distinguem *Xanthomonas* spp. responsáveis pela mancha bacteriana do tomateiro (Araújo et al. 2012).

2.3. Inoculação

A inoculação foi realizada na fase de cotilédones o que ocorreu sete dias pós-semeadura. Os cotilédones foram inoculados pelo método de pulverização da suspensão bacteriana na concentração de 10^8 UFC/mL calibrada previamente em espectrofotômetro ($OD_{600nm} = 0,3$) A quantidade de amostras de cotilédones foi de 450 plântulas do híbrido e 450 da cultivar. Foram realizadas duas repetições, totalizando 1.800 amostras de cotilédones, aproximadamente. As bandejas foram dispostas na bancada e pulverizadas de forma uniforme, até o escorrimento foliar, utilizando um pulverizador manual. Para o desenvolvimento de sintomas, as plantas foram mantidas em casa de vegetação durante 72 h sob nebulização intermitente (12 segundos a cada 1h). A temperatura registrada nesse período foi entre 28 e 30 °C. Os experimentos foram repetidos.

2.4. Captura de imagens hiperespectrais

A captura de imagens foi efetuada em uma câmera hiperespectral FX10e (Specim FX10e, Oulu, Finland) que opera na faixa do espectro eletromagnético VNIR (400-1000 nm). A câmera e LabScanner 40x20 (Specim FX10e, Oulu, Finland) foram calibrados quanto altura, foco e posição das luzes, usando o software Breeze (Prediktera AB), visando a captura de imagens com elevada qualidade. Antecedendo a captura das imagens, foram adquiridas as imagens de referência (branca/preta) com o obturador da câmera fechada. Foram obtidas imagens hiperespectrais de bandejas contendo as mudas apresentando apenas os dois cotilédones previamente inoculados (infectados) com a cepa bacteriana e como controle bandejas com cotilédones não-inoculados, sadios. A captura das imagens utilizando a câmera hiperespectral foi realizada a partir de 72 horas após inoculação até 168 h. Assim, a obtenção das imagens foi realizada em diferentes estágios de desenvolvimento da doença, considerando: estágio assintomático (72 h), estágio inicial (96 h e 120 h) e estágio tardio (144 h e 168 h). Durante cada um dos dias de avaliação uma imagem RGB convencional de cada uma das bandejas (sadia

e infectada) foi capturada para registro. As imagens hiperespectrais capturadas foram visualizadas em um computador acoplado a câmera, através do software Breeze que transforma os dados em pixels, formando uma imagem composta (*data cube*). Cada imagem continha blocos com 224 bandas (comprimentos de onda), com intervalos de 2,7 nm, na faixa de 400-1000 nm. Para adquirir a emissão espectral real de cada cotilédone, a influência dos padrões preto e branco foram subtraídos para obtenção da imagem calibrada (Kim et al. 2001). No *software* Breeze foi realizado a segmentação da imagem, seleção da região de interesse (ROI), bem como a normalização e padronização dos dados pelo método SNV (Standard Normal Variate). Os dados foram exportados para tabelas, com os dados de reflectância de cada banda em cada pixel para processamento posterior em outros softwares.

2.5. Análise de dados espectrais

Dois parâmetros foram empregados para avaliar em qual comprimento de onda verificou-se diferenças significativas, de forma a possibilitar a distinção de plantas saudáveis e doentes, bem como dos diferentes estágios de desenvolvimento da doença: (1) valor da diferença espectral, calculado pela diferença do valor médio da reflectância de cotilédones saudáveis e valor médio de reflectância de folhas infectadas pela bactéria em cada comprimento de onda; (2) valor de sensibilidade, calculado pela divisão do valor médio da reflectância de cotilédones doentes e cotilédones saudáveis, em cada comprimento de onda. As análises foram realizadas individualmente para cada cultivar. Os dados de reflectância normalizados dos cotilédones foram extraídos do *software* Breeze.

2.6. Métodos de classificação

Os dados hiperespectrais foram então analisados usando sistemas de classificação por aprendizado de máquina, avaliando diferentes estágios da doença (assintomático, inicial e tardio). O estágio assintomático se refere as plantas que são consideradas infectadas, porém não

apresentavam sintomas visíveis; no estágio inicial, considerou-se as pequenas lesões e encharcamentos; e no estágio tardio foram consideradas as plantas com necroses e coalescência de lesões. As análises foram empregadas de forma independente, usando os dados de reflectância espectral de cada estágio individualmente. Três modelos de classificação foram aplicados na análise dos dados: Support Vector Machine (SVM), Random Forest (RF) e Decision Tree (DT). O SVM é um modelo de classificação supervisionado que busca um hiperplano que melhor separe conjuntos de dados em classes distintas, com a maior margem de alcance (Cortes e Vapnik 1995). Os parâmetros de regularização (C) e kernel (gamma) foram ajustados. O parâmetro C controla o equilíbrio entre maximizar a margem de separação dos dados e minimizar os erros de classificação, enquanto o gamma estabelece o alcance da influência de cada exemplo de treinamento, influenciando o formato do hiperplano de separação. O RF é um modelo de classificação e regressão supervisionado baseado em um conjunto de múltiplas árvores de decisão, em que cada árvore é treinada em um subconjunto de dados aleatórios (Breiman 2001). Os dados de saída finais são obtidos pelo conjunto de predição de todas as árvores, por meio de votação para classificação ou média das previsões de todos os subconjuntos. Os parâmetros otimizados contemplaram o número de estimadores, a profundidade máxima das árvores e o número mínimo de amostras necessárias para dividir um nó e ser uma folha (ponto final de decisão do RF). O DT é um modelo de aprendizado supervisionado de classificação e regressão que divide os dados em distintos ramos baseado em condições e características, até alcançar uma decisão final. Cada nó interno corresponde a uma condição sobre um atributo, e cada ramo representa o resultado dessa condição, a predição ocorre nos nós folha (Quinlan 1986). No modelo DT, os parâmetros ajustados foram a profundidade máxima, o número mínimo de amostras necessários para dividir um nó, e o número mínimo de amostras para formar uma folha (ponto terminal do modelo DT).

O processamento dos dados por aprendizado de máquina foi realizado na plataforma Google Colab (Google Colaboratory), um ambiente de notebooks baseado em nuvem, que permite a execução da linguagem de programação *Python* diretamente no navegador. No estudo, o conjunto de dados de treinamento foi de 70% e o de teste de 30%, sendo inserido dados de entrada de plantas saudáveis e plantas doentes em diferentes estágios da doença, a abordagem foi aplicada para todos os modelos avaliados. Cada modelo de classificação teve seus hiperparâmetros ajustados durante o processo de avaliação dos resultados ainda na fase de treino. Os modelos foram avaliados quanto seu potencial diagnóstico pela Matriz de Confusão, pelos valores de *Accuracy* (acurácia) (Equação 1); *Precision* (precisão) (Equação 2); *F1-Score* (índice F1) (Equação 3) e *Recall* (revocação ou sensibilidade) (Equação 4). Essas métricas são baseadas em cálculos que avaliam a relação entre previsões positivas (*true positives*) e negativas corretas (*true negatives*), falsos positivos (*false positives*) e falsos negativos (*false negatives*) (Cunha et al. 2023; Reis-Pereira et al. 2023)

$$1. \text{ Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

$$2. \text{ Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$3. \text{ F1 Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$4. \text{ Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

3. RESULTADOS

3.1. Inoculação

A inoculação das plantas no estágio de cotilédones resultou em sintomas a partir do terceiro dia pós-inoculação (72 h) para ambas as cultivares. A cv. Santa Cruz mostrou-se, aparentemente, mais suscetível à mancha bacteriana, quando ao HMX 7885. A manifestação dos sintomas em ‘HMX 7885’ foi menos evidente nas primeiras 96 h, se intensificando após

esse período, enquanto na cultivar os sintomas de encharcamento já eram perceptíveis (96 h). Ao final da avaliação (168 h) as cultivares apresentavam sintomas mais severos com perfurações típicas da espécie bacteriana utilizada, bem como necrose foliar, mantendo-se a percepção de uma maior severidade em ‘Santa Cruz’.

3.2. Análise de dados espectrais da cv. Santa Cruz

A análise dos dados espectrais da cv. Santa Cruz de plantas sadias e infectadas, em diferentes estágios da doença nos comprimentos de ondas 400 a 936 nm, gerou três gráficos distintos. Para a reflectância espectral normalizada (Figura 1), ao longo dos comprimentos de ondas inferiores a 668 nm existe uma proximidade grande dos valores entre plantas sadias, assintomáticas, com sintomas iniciais e tardios, somente sendo visível uma diferença mais marcante após essa banda. Em especial para os sintomas tardios, bandas posteriores a 668 nm apresentaram maior divergências dos demais estádios. O aspecto da curva de diferença espectral normalizada entre os intervalos de 530 até 668 nm indicou uma divergência entre as três categorias, com outra variação próxima dos 750 nm (Figura 2A). Na sensibilidade espectral normalizada, entre as bandas 500 nm e 650 nm foi encontrada uma maior variação, com outro pico de divergência em 700 nm, tendendo à redução até o final dos comprimentos de onda analisados (Figura 2B).

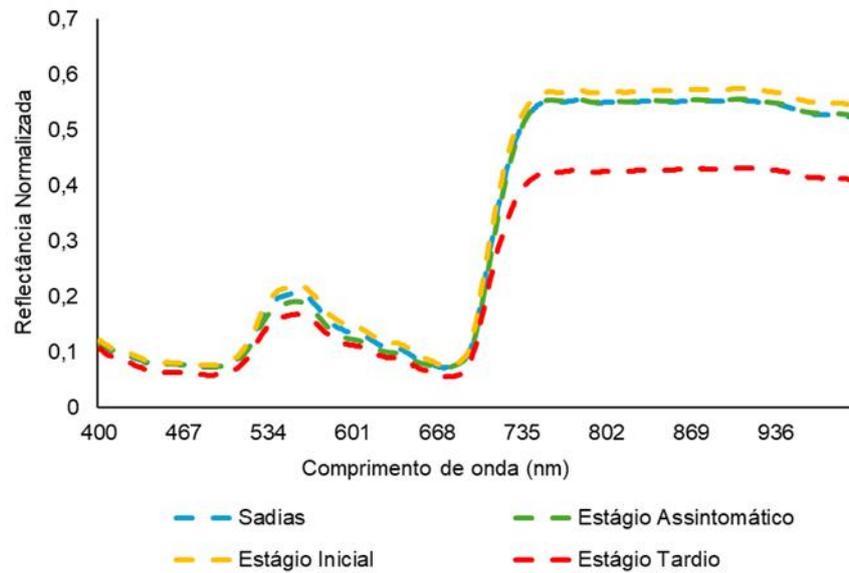


Figura 1. Reflectância espectral de cotilédones de tomateiro (cv. Santa Cruz) sadios e em diferentes estágios de infecção da mancha bacteriana.

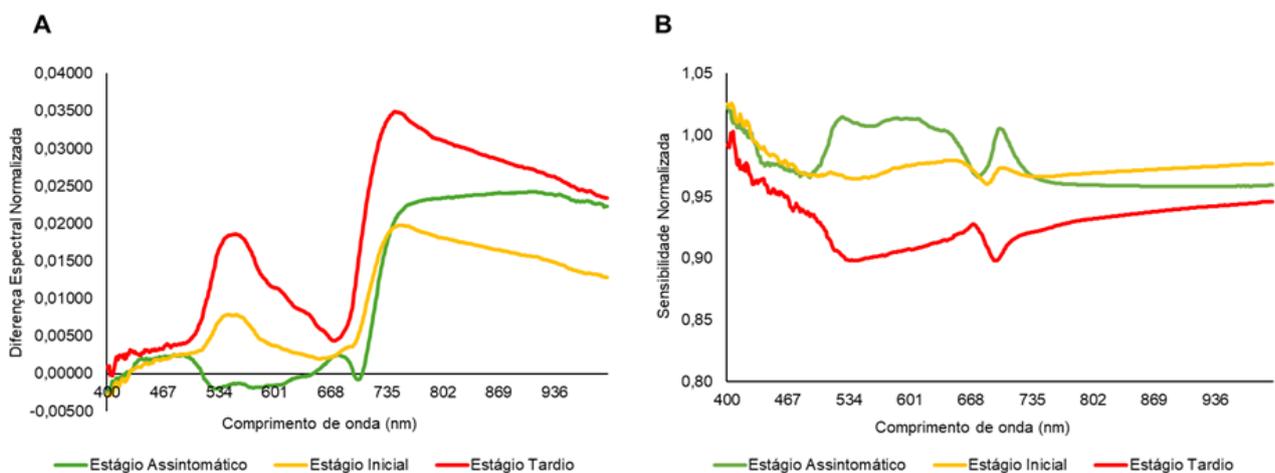


Figura 2. Cotilédones de tomateiro (cv. Santa Cruz) em diferentes estágios de infecção da mancha bacteriana. (A) Diferença espectral normalizada. (B) Sensibilidade espectral normalizada.

3.3. Classificação dos dados da cv. Santa Cruz

Dos modelos avaliados para classificação dos dados de cotilédones de tomateiro em diferentes estágios da mancha bacteriana, o modelo preditivo SVM foi o que apresentou melhor desempenho, com alta acurácia na classificação no estágio assintomático (96%), inicial (82%) e tardio (92,6%) (Figura 3). O modelo RF registrou acurácia de apenas 58% e o algoritmo DT de 57%, ambos no estágio assintomático. No estágio tardio a acurácia obtida por RF e DT foi

de 64% e 60%, respectivamente. Na análise de desempenho por meio das métricas *Precision*, *Recall* e *F1-Score* (Tabela 1) o modelo SVM manteve superioridade quando comparado aos outros modelos testados, evidenciando um *F1-Score* acima de 90% nos estágios assintomático e tardio. Os altos valores observados na *Precision* e *Recall* indicam que o modelo SVM pode classificar corretamente o conjunto de dados.

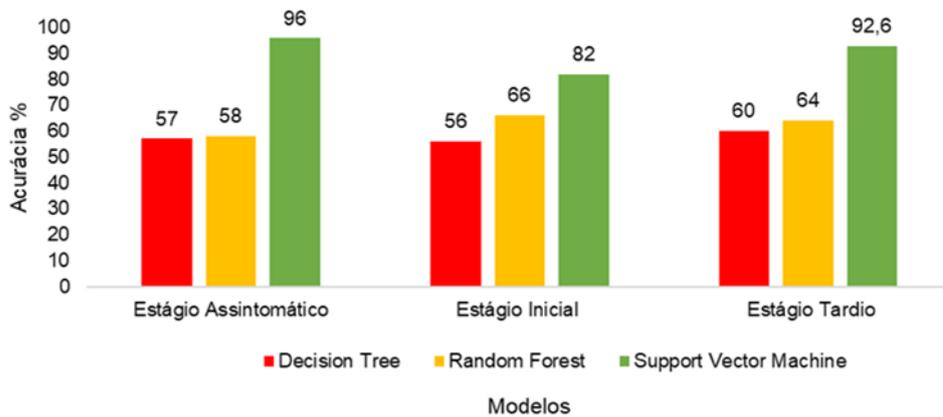


Figura 3. Desempenho dos modelos de *machine learning* na classificação de cotilédones de tomateiro (cv. Santa Cruz) sadios e infectados em diferentes estágios da mancha bacteriana.

Tabela 1. Desempenho dos modelos na classificação de dados de cotilédones cv. Santa Cruz

Estágios	Modelos	Classes	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Assintomático	DT	Infectada	48%	61%	53%
		Sadia	66%	54%	59%
	RF	Infectada	49%	50%	49%
		Sadia	64%	63%	64%
	SVM	Infectada	95%	95%	95%
		Sadia	96%	96%	96%
Inicial	DT	Infectada	61%	45%	52%
		Sadia	53%	68%	60%
	RF	Infectada	68%	66%	67%
		Sadia	64%	66%	65%
	SVM	Infectada	82%	84%	83%
		Sadia	82%	80%	81%
Tardio	DT	Infectada	59%	64%	61%
		Sadia	61%	56%	59%
	RF	Infectada	65%	60%	62%
		Sadia	63%	69%	66%
	SVM	Infectada	93%	91%	92%
		Sadia	92%	94%	93%

3.4. Análise de dados espectrais HMX 7885

Foram gerados três gráficos por meio da análise dos parâmetros de comportamento espectral avaliados para o híbrido HMX 7885, referente aos cotilédones sadios e infectados pela cepa bacteriana de *Xep* em diferentes estágios de infecção. A reflectância de cotilédones sadios e no estágio tardio de infecção apresentaram comportamento muito semelhante. No estágio assintomático e inicial os cotilédones exibiram uma baixa reflectância quando comparado as plantas sadias, sobretudo nos comprimentos de onda menores entre 400 nm e 600 nm e nos maiores 800-936 nm (Figura 4). A diferença espectral normalizada no estágio assintomático mostrou valores próximos de zero em todos os comprimentos de onda. O comportamento da curva no estágio inicial exibiu valores positivos entre 600 nm e 800 nm, enquanto no estágio tardio o comportamento espectral apresenta valores negativos em quase todo o espectro, com elevação por volta de 600 nm e 800 nm (Figura 5A). Na sensibilidade espectral valores acima de 1,0 apontam que cotilédones doentes refletem mais luz que os sadios e valores abaixo de 1,0 indicam o contrário, o estágio assintomático inicia com valores acima de 1,0 em comprimentos de onda curtos (400-467 nm), apontando maior reflectância nessa região do espectro. No entanto, houve uma evolução para valores próximos de 1,0 nos demais comprimentos de onda. No estágio inicial, a sensibilidade espectral apresentou valores ligeiramente acima de 1,0 ao longo do espectro. No estágio tardio os valores de sensibilidade mantiveram um comportamento acima de 1,0 em quase todo o espectro (Figura 5B).

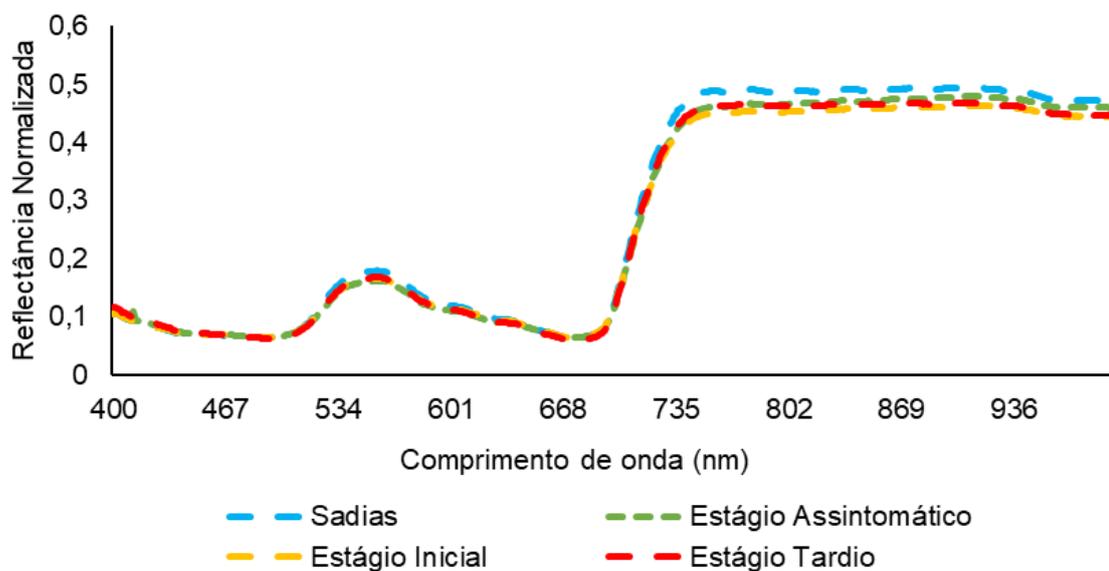


Figura 4. Reflectância espectral de cotilédones de tomateiro (híbrido HMX 7885) sadios e em diferentes estágios de infecção da mancha bacteriana

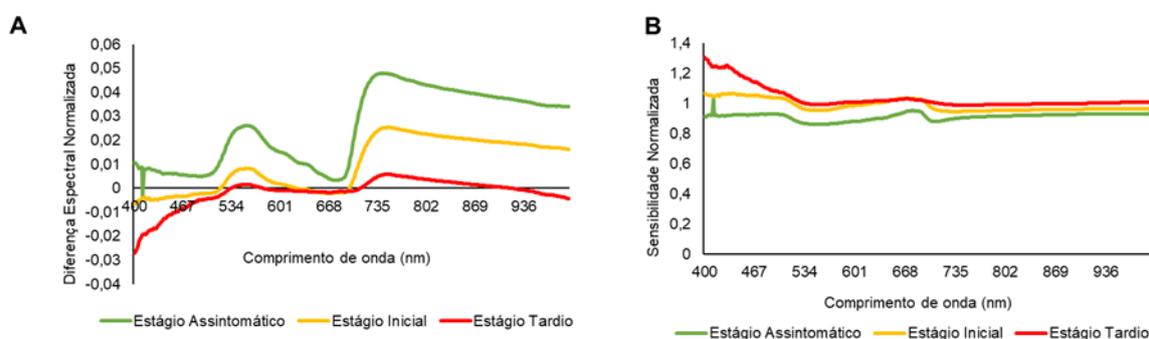


Figura 5. Cotilédones de tomateiro (híbrido HMX 7885) em diferentes estágios de infecção da mancha bacteriana. (A) Diferença espectral normalizada. (B) Sensibilidade espectral normalizada.

1.1. Classificação dos dados HMX 7885

A performance dos três algoritmos usados na classificação dos dados em diferentes estágios da mancha bacteriana no híbrido HMX 7885 indicou que o modelo de predição SVM

apresentou melhor desempenho na classificação de dados com uma acurácia de 88% no estágio assintomático e 93% no estágio tardio (Figura 8). Os algoritmos RF e DT registraram uma acurácia antes do surgimento de sintomas de 78% e 72%, respectivamente. Já no estágio mais avançado da doença a acurácia obtida no modelo RF foi de 70% e no DT 65%. Com base nos valores de *Precision*, *Recall* e *F1-Score* (Tabela 2), o modelo SVM também apresentou o melhor desempenho na classificação do conjunto de dados, evidenciando uma performance consistente com valores acima de 90% para *Precision* e *Recall*, nos estágios inicial e tardio.

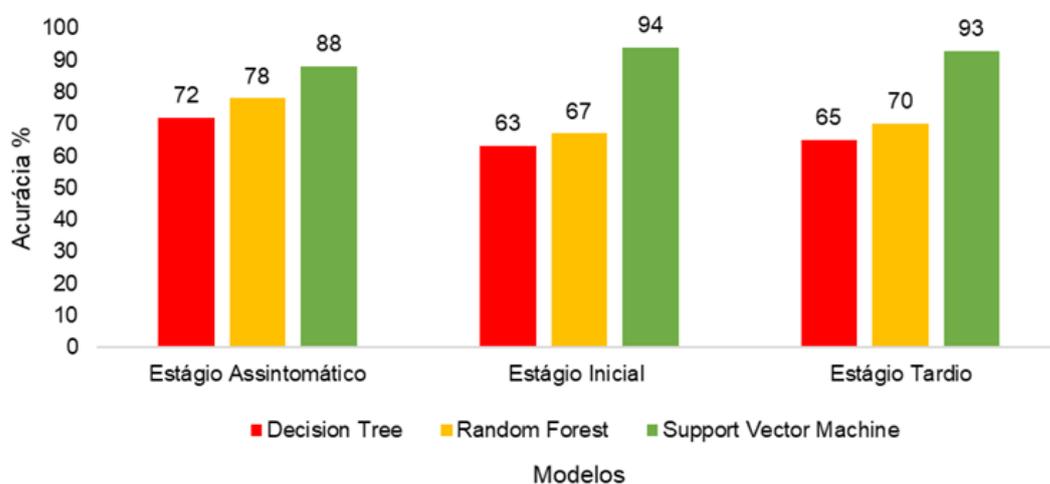


Figura 6. Desempenho dos modelos de *machine learning* na classificação de cotilédones de tomateiro (híbrido HMX 7885) sadios e infectados em diferentes estágios da mancha bacteriana.

Tabela 2. Desempenho dos modelos na classificação de dados de cotilédones híbrido HMX 7885.

Estágios	Modelos	Classes	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Assintomático	DT	Infected	75%	65%	70%
		Sadia	69%	78%	73%
	RF	Infected	77%	80%	78%
		Sadia	79%	76%	77%
	SVM	Infected	87%	90%	89%
		Sadia	90%	86%	88%
Inicial	DT	Infected	63%	78%	70%
		Sadia	63%	45%	53%
	RF	Infected	66%	82%	73%
		Sadia	70%	49%	57%
	SVM	Infected	94%	95%	94%
		Sadia	94%	92%	93%
Tardio	DT	Infected	66%	81%	73%
		Sadia	63%	43%	51%
	RF	Infected	68%	89%	77%
		Sadia	74%	44%	56%
	SVM	Infected	95%	93%	94%
		Sadia	91%	93%	92%

4. DISCUSSÃO

Compreender a dinâmica das infecções causadas por patógenos de planta é fundamental para a garantia da produtividade, bem como para a segurança alimentar. No presente estudo, avaliou-se a detecção precoce da mancha bacteriana do tomateiro em plântulas de tomate da cv. Santa Cruz (consumo in natura) e do híbrido HMX 7885 (processamento industrial), ainda no estágio cotiledonar, por meio do uso de sensor hiperespectral, juntamente com a classificação dos dados com modelos de predição de aprendizado de máquina. Além disso, foi aplicada a análise de reflectância espectral dos dados obtidos por meio do sensor hiperespectral, no intuito de evidenciar diferenças significativas na assinatura espectral em diferentes estágios da mancha bacteriana em cotilédones de tomateiro, tanto do híbrido, quanto da cultivar.

A reflectância espectral tem sido amplamente utilizada como uma técnica não destrutiva para monitorar a sanidade de plantas e detectar doenças de forma precoce (Mishra et al. 2020). No conjunto de dados de cotilédones da cv. Santa Cruz de tomateiro sadios e em diferentes estágios da mancha bacteriana observou-se que a reflectância variou significativamente com a progressão da doença. Os cotilédones sadios mostraram um perfil de reflectância elevado na faixa do infravermelho próximo - NIR (700-1000 nm), o que é esperado para plantas em condições fisiológicas estáveis, isso porque nessa faixa espectral ocorre o alto espalhamento de luz pelos tecidos do mesófilo foliar (Vasseur et al. 2022). No estágio assintomático, a curva de reflectância permaneceu similar à das plantas saudáveis, porém com divergências sutis ao longo do espectro. Isso reforça o potencial das imagens hiperespectrais na detecção precoce de doenças, revelando mudanças imperceptíveis ao olho humano (Terentev et al. 2022). Durante o estágio inicial da doença foi observado uma diminuição na reflectância em relação aos cotilédones sadios, o que foi ainda mais evidente no estágio tardio em que a curva de reflectância manteve um comportamento abaixo da curva dos cotilédones sadios. Isso ocorre

possivelmente devido ao avanço da infecção bacteriana causar um impacto na fisiologia da planta de forma estrutural ou a fatores associados a senescência (Marín-Ortiz et al. 2024; Vasseur et al. 2022). Para análise de plantas, as faixas de comprimento de onda mais úteis são os presentes no VNIR combinadas. Os comprimentos de onda de 400-700 nm podem capturar mudanças na pigmentação foliar e de 700-1300 nm na estrutura da célula do mesófilo, porém, para evidenciar mudanças no conteúdo de água da planta, são necessárias faixas que compreendem o espectro SWIR, infravermelho de ondas curtas (900-2500 nm) (Lowe et al. 2017

A curva de diferença espectral no estágio assintomático exibiu baixas variações, mantendo-se próxima de zero na maior parte do espectro. Esse comportamento foi semelhante ao da sensibilidade espectral em que o estágio assintomático exibiu valores próximos de 1,00 em todo o espectro, indicando poucas alterações na reflectância dos cotilédones infectados em comparação com os sadios. Isso sugere que as alterações na reflectância na fase assintomática são mínimas, indicando que as alterações fisiológicas causadas pela bactéria nesse estágio ainda causam pouco impacto no comportamento espectral, como visto por Cao et al. (2022), em que avaliaram a detecção precoce da queima bacteriana do arroz (*Xanthomonas oryzae*). Na diferença espectral observou-se um pico a partir de 668 nm, que coincide com o comprimento de onda associado à absorção de clorofila, esse aumento na diferença espectral pode indicar descoloração inicial ou degradação da clorofila, devido ao estresse causado pela presença da bactéria (Marín-Ortiz et al. 2024; Pengphorm et al. 2024). Na análise de sensibilidade espectral a curva na fase tardia da doença exibiu valores abaixo de 1,00. Esse comportamento implica a degradação dos cotilédones causada pelo avanço da infecção, levando a redução significativa da reflectância (Mahlein et al. 2012; Sterling e Rienzo 2022).

Observou-se que plantas sadias exibiram alta reflectância, mantendo valores próximos aos estágios iniciais e assintomáticos, com alterações mais evidentes na faixa do NIR. O declínio

na reflectância nos cotilédones nos estágios assintomático e inicial, pode ser atribuído às alterações fisiológicas e bioquímicas causadas pela presença da bactéria, que afetam diretamente a interação da luz com as folhas (Vasseur et al. 2022). No estágio tardio, a reflectância apresenta semelhanças com o comportamento espectral das plantas saudáveis. Sugerindo uma adaptação da planta à infecção, em que mecanismos de compensação são ativados a fim de restaurar a funcionalidade foliar. Isso pode ocorrer devido a planta ativar respostas de defesa ao patógeno, iniciando diversas reações bioquímicas e moleculares que contribuem para alterações espectrais (Reis-Pereira et al. 2024). As fitoalexinas são um exemplo que a sua produção pode estar ligada ao aumento na reflectância de plantas no espectro visível (Leucker et al. 2017). Além disso, a variação do comportamento espectral na planta hospedeira, em especial na faixa do visível, pode estar ligada a moléculas específicas produzidas pelo patógeno. Um exemplo disso, é a infecção bacteriana em tomateiro causada por *Pseudomonas syringae* pv. *tomato* que produzem uma fitotoxina específica, denominada coronatina, que induz alterações na fluorescência da clorofila, influenciando a absorção e dispersão de luz nos tecidos do tomateiro (Zhang et al. 2021). A sensibilidade espectral no estágio tardio evidenciou a normalização da sensibilidade espectral, com valores próximos de 1,00, o que sugere uma possível adaptação da planta a infecção. Exibindo uma provável tolerância ou equilíbrio entre os mecanismos de defesa da planta e os efeitos da interação com a bactéria (Reis-Pereira et al. 2024). Em contrapartida, no estágio tardio da doença os valores de diferença espectral foram consistentemente negativos, principalmente na faixa do visível, o que indica uma maior reflectância das plantas doentes nesse estágio, o que pode ter relação ao colapso celular e a necrose das folhas (Lowe et al. 2017).

Os resultados apresentados indicam que a infecção bacteriana nas plantas de tomate impacta significativamente as propriedades espectrais foliares, conforme relatado em estudos envolvendo a mancha bacteriana do tomateiro (Abdulridha et al. 2020) Essas descobertas

confirmam a utilidade da análise de dados de reflectância espectral na detecção da mancha bacteriana do tomateiro em diferentes estágios de desenvolvimento da doença. A possibilidade de detectar essas diferenças espectrais, especialmente em estágios iniciais, pode ser crucial para eficiência no manejo da doença, contribuindo para a aplicação de estratégias antes que a infecção alcance níveis mais severos. Além disso, os valores negativos de diferença espectral no estágio tardio da doença, podem servir como indicadores de gravidade da infecção (Mandal et al. 2023). O uso de sensor hiperespectral associado a exploração de dados por algoritmos de aprendizado de máquina e a análise dos padrões de reflectância são favoráveis a detecção precoce da mancha bacteriana em tomateiro e capazes de distinguir diferentes estágios da doença. Os dados obtidos são promissores, pois, as espécies causadoras da mancha bacteriana em tomate, podem permanecer de forma latente na planta, ocasionando em alterações imperceptíveis a olho nu, além do fato de que sua disseminação entre plantas pode não ser instantânea, ocorrendo ao longo do tempo em função de diversas condições como temperatura, umidade, presença de água livre, cepa bacteriana, concentração de inóculo e suscetibilidade da hospedeira (Abrahamian et al. 2021). O aprimoramento e adequação de modelos de predição para condições específicas de campo permanecem como um ponto chave para pesquisas futuras, objetivando a aplicação prática dessas técnicas em larga escala no monitoramento de lavouras.

Os modelos de predição utilizados nesse estudo demonstraram capacidade de classificar cotilédones de tomateiro sadios e doentes em diferentes estágios de desenvolvimento da mancha bacteriana. Cada algoritmo mostrou um desempenho distinto, o SVM evidenciou maior potencial de classificação de dados de cotilédones sadios e infectados mesmo antes do surgimento de sintomas (estágio assintomático) tanto para o híbrido quanto para a cultivar, com uma acurácia de 88% e 96%, respectivamente. Essa eficiência é condizente com aspectos conhecidos do SVM, como a capacidade de manipular dados de alta dimensão e a eficácia de classificação com margens claras de separação, mesmo diante de conjuntos de dados

complexos, frequentes em contextos agrícolas, em que os sinais de infecção causados por patógenos de plantas podem ser sutis e apresentar variabilidade (Pathak et al. 2022; Ali et al. 2024). O modelo Decision Tree e Random Forest exibiram uma performance inferior na classificação dos dados, em comparação com SVM. A simplicidade do modelo DT que constantemente favorece a interpretação e implementação, pode ser um fator limitante na sua eficiência em contextos em que as relações entre variáveis são mais complexas e menos delineadas, como é o caso da classificação de dados de doenças de plantas (Ali et al. 2024).

Estudos mostram o desempenho do SVM na análise de diferentes patossistemas envolvendo o tomateiro, como Zhao et al. (2024) que avaliaram diferentes modelos na classificação de plantas de tomate sadias e infectadas por *Cladosporium fulvum*. Entre os modelos testados o SVM isoladamente apresentou um baixo desempenho com uma acurácia de 69 %, já quando combinado ao PCA (Principal Component Analysis) obteve acurácia de 87,6 %. Diferentemente de Reis-Pereira et al. (2023) em que o SVM alcançou uma acurácia de 100 % e 74 %, na classificação da pinta bacteriana (*Pseudomonas syringae* pv. *tomato*) e mancha bacteriana (*Xanthomonas euvesicatoria*), respectivamente, antes mesmo do surgimento de sintomas. Esse mesmo estudo aplicando o modelo SVM evidenciou um valor de *F1-Score* de 76% avaliando conjunto de dados de mudas de tomateiro ainda sem sintomas da mancha bacteriana. Cunha et al. (2023) aplicando diferentes modelos na detecção da mancha bacteriana em folhas de tomate, obtiveram *F1-Score* de 74% no estágio da doença em que sintomas ainda não eram visíveis e no estágio mais avançado valores acima de 80%, aplicando o modelo SVM. Enquanto, Cen et al. (2022) combinando algoritmos genéticos e SVM, alcançaram *F1-Score* de 80% na classificação de dados de caules de tomate infectados pela mancha bacteriana em estágios iniciais da doença. Dessa forma, a escolha do modelo de aprendizado de máquina mais adequado pode diferir dependendo do estágio específico da doença, tipos de patógeno, espécie vegetal, ou mesmo entre cultivares, o que torna importante a adaptação dos modelos de acordo

ao contexto específico de cada aplicação (Rodriguez et al. 2024). O aprimoramento e adequação de modelos de predição para condições específicas de campo permanecem como um ponto chave para pesquisas futuras, objetivando a aplicação prática dessas técnicas em larga escala no monitoramento de lavouras. Assim, a integração das tecnologias hiperspectrais e do aprendizado de máquina visando sua aplicação prática na agricultura podem revolucionar a forma como doenças de plantas são detectadas oferecendo uma abordagem não destrutiva, sustentável e economicamente viável.

5. REFERÊNCIAS

- Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., Roberts, P. (2020a). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precis. Agric.* 21, 955–978. doi: 10.1007/s11119-019-09703-4
- Abrahamian P, Jones JB, Vallad GE (2019) Efficacy of copper and copper alternatives for management of bacterial spot on tomato under transplant and field production. *Crop Protection* 126: 104919.
- Abrahamian P, Sharma A, Jones JB, Vallad GE (2021) Dynamics and spread of bacterial spot epidemics in tomato transplants grown for field production. *Plant Disease* 105(3): 566–575. <https://doi.org/10.1094/PDIS-05-20-0945-RE/ASSET/IMAGES/LARGE/PDIS-05-20-0945-RET1.JPEG>
- Ali F, Razzaq A, Tariq W, Hameed A, Rehman A, Razzaq K, Ondrasek G (2024) Spectral Intelligence: AI-Driven Hyperspectral Imaging for Agricultural and Ecosystem Applications *Agronomy* 14(10): 2260.
- Araújo ER, Costa JR, Ferreira MASV, Quezado-Duval AM (2012) Simultaneous detection and identification of the *Xanthomonas* species complex associated with tomato bacterial spot using species-specific primers and multiplex PCR. *Journal of Applied Microbiology* 113(6): 1479–1490. <https://doi.org/10.1111/J.1365-2672.2012.05431.X>
- Araújo ER, Costa JR, Ferreira MASV, Quezado-Duval AM (2017) Widespread distribution of *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. *Plant Pathology* 66(1): 159–168. doi: <https://doi.org/10.1111/PPA.12543>
- Cao Y, Yuan P, Xu H, Martínez-Ortega JF, Feng J, Zhai Z (2022) Detecting asymptomatic infections of rice bacterial leaf blight using hyperspectral imaging and 3-dimensional convolutional neural network with spectral dilated convolution. *Frontiers in Plant Science* 13: 963170.
- Cen Y, Huang Y, Hu S, Zhang L, Zhang J (2022) Early detection of bacterial wilt in tomato with portable hyperspectral spectrometer. *Remote Sensing* 14(12): 2882.
- Clemente FMVT, Boiteux LS (2012) *Produção de Tomate para Processamento Industrial* 1th Ed. Embrapa, Brasil, Brasília DF
- CONAB - Companhia Nacional de Abastecimento (2021) *Compêndio de Estudos Conab*. 18:2448-3710 <http://www.conab.gov.br/>
- Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, Cottyn B (2016) Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathology* 65: 792–806. <https://doi.org/10.1111/ppa.12461>
- Cunha VAG, Hariharan J, Ampatzidis Y, Roberts PD (2023) Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising remote sensing and artificial intelligence. *Biosystems Engineering* 234: 172-186.

Dua Y, Kumar V, Singh RS (2020) Comprehensive review of hyperspectral image compression algorithms. *Optical Engineering* 59(9): 090902-090902.

FAOSTAT (2021) Crops and livestock products
<https://www.fao.org/faostat/en/#data/QCL/visualize>

IBGE - Instituto Brasileiro de Geografia e Estatística (2022) Levantamento Sistemático da Produção Agrícola Levantamento Sistemático da Produção Agrícola | IBGE

Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW (2004) Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Systematic and Applied Microbiology* 27:755-62

Klein-Gordon JM, Xing Y, Garrett KA, Abrahamian P, Paret ML, Minsavage GV, Strayer-Scherer AL, Fulton JC, Timilsina S, Jones JB, Goss EM, Vallad GE (2021). Assessing changes and associations in the *Xanthomonas perforans* population across Florida commercial tomato fields via a statewide survey. *Phytopathology*® 111(6):1029-1041.

Lee, H., Kim, M. S., Lim, H. S., Park, E., Lee, W. H., Cho, B. K. (2016a). Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: Application to seeds of the “Sambok Honey” cultivar. *Biosyst. Eng.* 148, 138– 147. doi: 10.1016/j.biosystemseng.2016.05.014

Leucker M, Wahabzada M, Kersting K, Peter M, Beyer W, Steiner U, Oerke EC (2016) Hyperspectral imaging reveals the effect of sugar beet quantitative trait loci on *Cercospora* leaf spot resistance. *Functional Plant Biology* 44(1): 1-9.

Lowe A, Harrison N, French AP (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant methods* 13(1): 80.

Mahlein AK, Steiner U, Hillnhütter C, Dehne HW, Oerke EC (2012) Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant methods* 8: 1-13.

Mandal N, Adak S, Das DK, Sahoo RN, Mukherjee J, Kumar A, Gakhar S (2023) Spectral characterization and severity assessment of rice blast disease using univariate and multivariate models. *Frontiers in Plant Science* 14: 1067189.

Marín-Ortiz JC, Botero-Fernández V, Zapata-Henao S, Hoyos-Carvajal LM (2024) Early detection of bacterial wilt in bananas caused by *Ralstonia solanacearum* using reflectance spectroscopy. *Journal of Plant Diseases and Protection* 131(2): 23-531.

Mishra P, Polder G, Vilfan N (2020) Close range spectral imaging for disease detection in plants using autonomous platforms: a review on recent studies. *Current robotics reports* 1: 43-48.

Morinière L, Burlet A, Rosenthal ER, Nesme X, Portier P, Bull CT, Lavire C, Saux MFL, Bertolla F (2020) Clarifying the taxonomy of the causal agent of bacterial leaf spot of

- lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol et al., 2000 emend. Timilsina et al., 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al., 1995. *Systematic and Applied Microbiology* 43: 126087. doi: 10.1016/j.syapm.2020.126087
- Nguyen C, Sagan V, Maimaitiyiming M, Maimaitijiang M, Bhadra S, Kwasniewski MT (2021) Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. *Sensors* 21(3): 742. doi: 10.3390/s21030742
- Pathak DK, Kalita SK, Bhattacharya DK (2022) Hyperspectral image classification using support vector machine: a spectral spatial feature based approach. *Evolutionary Intelligence* 1-15.
- Pengphorm P, Thongrom S, Daengngam C, Duangpan S, Hussain T, Boonrat P (2024). Optimal-Band Analysis for Chlorophyll Quantification in Rice Leaves Using a Custom Hyperspectral Imaging System. *Plants* 13(2): 259.
- Polder G, Blok PM, de Villiers HAC, van der Wolf JM, Kamp J (2019) Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Front Plant Sci* 10. doi: <https://doi.org/10.3389/fpls.2019.00209>.
- Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, Vallad GE, Jones JB (2015) Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Molecular plant pathology* 16: 907–920. doi: 10.1111/mpp.12244
- Quezado-Duval AM, Lopes CA (2010) Mancha-bacteriana: uma atualização para o sistema de produção integrada de tomate indústria. *Embrapa hortaliças* ISSN 1415-3033
- Reis-Pereira M, Santos FND, Tavares F, Cunha M (2023) Enhancing host-pathogen phenotyping dynamics: Early detection of tomato bacterial diseases using hyperspectral point measurement and predictive modeling. *Frontiers in Plant Science* 14: 1242201.
- Reis-Pereira M, Verrelst J, Tosin R, Rivera Caicedo JP, Tavares F, Neves dos Santos F, Cunha M (2024). Plant Disease Diagnosis Based on Hyperspectral Sensing: Comparative Analysis of Parametric Spectral Vegetation Indices and Nonparametric Gaussian Process Classification Approaches. *Agronomy* 14(3): 493.
- Rodríguez GVE, Izquierdo-Bueno I, Cantoral JM, Carbú M, Garrido C (2024) Artificial intelligence: A promising tool for application in phytopathology. *Horticulturae* 10(3): 197.
- Shakoor N, Lee S, Mockler TC (2017) High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Current Opinion in Plant Biology* 38: 184–192. doi: <https://doi.org/10.1016/J.PBI.2017.05.006>
- Stamford JD, Violet-Chabrand S, Cameron I, Lawson T (2023) Development of an accurate low cost NDVI imaging system for assessing plant health. *Plant Methods* 19(1): 1–19. doi: <https://doi.org/10.1186/S13007-023-00981-8/FIGURES/10>

- Sterling A, Di Rienzo JA (2022) Prediction of South American leaf blight and disease-induced photosynthetic changes in rubber tree, using machine learning techniques on leaf hyperspectral reflectance. *Plants* 11(3): 329.
- Terentev A, Dolzhenko V, Fedotov A, Eremenko D (2022) Current state of hyperspectral remote sensing for early plant disease detection: A review. *Sensors* 22(3): 757.
- Timilsina S, Potnis N, Newberry E A, Liyanapathirana P, Iruegas-Bocardo F, Branco FF, Goss EM, Jones JB (2020) *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology* 18(8): 415-427. <https://doi.org/10.1038/s41579-020-0361-8>
- Vasseur F, Cornet D, Beurier G, Messier J, Rouan L, Bresson J, Violle C. (2022) A perspective on plant phenomics: coupling deep learning and near-infrared spectroscopy. *Frontiers in Plant Science* 13: 836488.
- Virlet N, Sabermanesh K, Sadeghi-Tehran P, Hawkesford MJ (2017) Field scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Funct Plant Biol* 44:143–53. doi: <https://doi.org/10.1071/fp16163>
- Wahabzada M, Mahlein AK, Bauckhage C, Steiner U, Oerke EC, Kersting K (2015) Metro Maps of Plant Disease Dynamics-Automated Mining of Differences Using Hyperspectral Images. *PLOS ONE* 10(1): e0116902. doi: <https://doi.org/10.1371/JOURNAL.PONE.0116902>
- Zhang ZC, He B, Sun S, Zhang X, Li T, Wang HH, Geng XQ (2021) The phytotoxin COR induces transcriptional reprogramming of photosynthetic, hormonal and defence networks in tomato. *Plant Biology* 23: 69-79.
- Zhao X, Liu Y, Huang Z, Li G, Zhang Z, He X, Li Z (2024) Early diagnosis of *Cladosporium fulvum* in greenhouse tomato plants based on visible/near-infrared (VIS/NIR) and near-infrared (NIR) data fusion. *Scientific Reports* 14(1): 20176.

CAPÍTULO 3

DETECÇÃO DO ESTRESSE CAUSADO POR *Xanthomonas euvesicatoria* *pv. perforans* EM SEMENTES DE TOMATE POR IMAGEM HIPERESPECTRAL E APRENDIZADO DE MÁQUINA

RESUMO

A disseminação de patógenos por sementes contribui para a ocorrência de epidemias em campos de cultivo, sendo um fator limitante na produção de diversas culturas, como é o caso da mancha bacteriana do tomateiro. A detecção de patógenos ainda na semente permite evitar a introdução do microrganismo em uma área de cultivo e surtos da doença, favorecendo o manejo precoce, além de reduzir as perdas. O uso de sensores hiperespectrais, de modelos de predição de aprendizado de máquina e a análise de dados de reflectância espectral pode significar uma abordagem eficiente, rápida e econômica na diagnose de sementes infectadas por patógenos. Este trabalho teve por objetivo detectar e diferenciar sementes de tomateiro sadias e infectadas por *Xanthomonas euvesicatoria pv. perforans* usando diferentes métodos de inoculação, imagens hiperespectrais e modelos de aprendizado de máquina. Quatro métodos de inoculação foram aplicados, diretamente nas flores, injeção de frutos, pincelamento de frutos e de imersão de sementes em inóculo com aplicação de vácuo. As flores inoculadas não geraram frutos, enquanto injeção somente gerou frutos viáveis em concentrações mais baixas de inóculo. Frutos pincelados apresentaram os sintomas mais fortes. Em torno de 500 sementes por tratamento tiveram suas imagens registradas com a câmera hiperespectral na faixa de 400-1000 nm. Os dados espectrais adquiridos de sementes foram previamente normalizados pelo método SNV (Standard Normal Variate). A análise de dados de reflectância espectral de sementes de tomateiro sadias e infectadas por diferentes técnicas de inoculação, apontou diferenças significativas nas propriedades espectrais em função da infecção bacteriana. As inoculações por pincel e injeção apresentaram divergências entre sementes geradas de frutos inoculados e não inoculados, com o método à vácuo gerando variações maiores entre as sementes infiltradas e não infiltradas. Após processamento inicial das imagens, foram usados os modelos de aprendizado de máquina Decision Tree - DT, Random Forest -RF e Support Vector Machine - SVM. Enquanto os modelos DT e RF apresentaram baixa acurácia nos testes, o SVM mostrou alta acurácia na classificação de dados de sementes inoculadas pelo método de injeção em frutos e pincel, 97% e 98%, respectivamente. Os resultados alcançados apoiam o uso de imagens

hiperespectrais e aprendizado de máquina na detecção precoce do estresse causado por *Xanthomonas euvesicatoria* pv. *perforans* em sementes de tomateiro.

Palavras-chave: diagnose, hiperespectro, diagnose precoce, tomate, *Xanthomonas*.

1. INTRODUÇÃO

A produção de sementes de tomate é uma etapa crucial para a obtenção de cultivos livres de doenças. No Brasil a produção de sementes é regida por lei e normas que visam a manutenção da qualidade e sanidade de sementes, a Lei nº 10.711, de 5 de agosto de 2003 dispõe sobre o Sistema Nacional de Sementes e Mudas.

A sanidade de sementes de tomateiro pode ser afetada pela ocorrência de diferentes patógenos como fungos, vírus e bactérias. Os fitopatógenos fúngicos são maioria quando se trata da contaminação de sementes de tomate como é o caso de *Septoria lycopersici*, agente causador da septoriose, *Alternaria solani* que causa a pinta preta do tomateiro, *Sclerotinia sclerotiorum*, responsável pela doença conhecida como mofo branco, *Fusarium oxysporum* f. sp. *lycopersici* importante fitopatógeno do cultivo de tomate conhecido por ocasionar a murcha de fusarium e *Cladosporium fulvum* que causa a mancha de cladospório (Lopes et al. 2021). Os vírus Tomato mosaic virus (ToMV) e Tobacco mosaic virus (TMV), ambos pertencentes a família *Virgaviridae*, gênero *Tobamovirus*, são transmitidos de forma eficiente por meio de sementes de tomate (Inoue-Nagata e Lima 2021). As bactérias são potenciais agentes patogênicos na cultura, entre as bactérias fitopatogênicas já relatadas associadas a sementes de tomateiro pode-se citar *Clavibacter michiganensis* subsp. *michiganensis*, agente causal do cancro bacteriano, *Pseudomonas syringae* pv. *tomato* responsável pela pinta bacteriana em frutos e folhas de tomate e as espécies de *Xanthomonas* causadoras da mancha bacteriana do tomateiro (Lopes et al. 2021).

As sementes são fonte de inóculo inicial potencial a ocorrência da mancha bacteriana e podem favorecer a entrada de diferentes raças de *Xanthomonas* spp. em novas regiões ou países. Araújo et al. (2017) relataram que possivelmente o surgimento da raça 4 de *X. euvesicatoria* pv. *perforans* no Brasil seja devido à importação de sementes infectadas. Na Flórida em 1992 houve uma repentina mudança da raça de *Xanthomonas* que prevalecia nos campos de produção

de pimenta, e isso foi explicado devido a introdução de sementes importadas nos campos de produção da cultura (Pohronezny et al. 1992).

No caso das bactérias fitopatogênicas em tomateiro a localização do patógeno em sementes já é bastante esclarecida para *Clavibacter michiganensis* subsp. *michiganensis*, que causa infecção sistêmica na planta e pode estar presente em sementes causando infecção tanto internamente quanto externamente, a depender da via de penetração inicial na hospedeira (Xu et al. 2010; Tancos et al. 2013). Em *Pseudomonas syringae* pv. *tomato* a infestação em sementes ocorre de forma superficial (Chai et al. 2020).

Para a mancha bacteriana do tomateiro, causada por diferentes espécies de *Xanthomonas*, a localização da colonização em sementes de tomate ainda não é totalmente esclarecida. Em um estudo realizado por Sharma et al. (2014) na Índia, detectaram através de microscopia a presença de *Xanthomonas euvesicatoria* em sementes obtidas de tomateiro naturalmente infectados. Essas sementes possuíam três diferentes níveis de sintomas e coloração, sementes com aspecto assintomático apresentaram a colonização externa superficial da bactéria, em sementes moderadamente escuras foi possível visualizar a presença de colônias bacterianas mais internamente no endosperma e em sementes com alto grau de infecção e totalmente escurecidas a colonização ocorreu até mesmo na região do embrião.

A obtenção de sementes naturalmente infectadas pode contribuir para estudos da dinâmica de infecção de patógenos bacterianos em sementes. Fato esse já relatado em diferentes estudos a partir da obtenção de sementes infectadas de campos de cultivo de tomateiro (Sharma et al. 2014; Chai et al. 2020). Além disso, métodos de inoculação que imitam infecções naturais causadas por bactérias a campo também podem favorecer no conhecimento da dinâmica de infecção em sementes, como no estudo de Dutta et al. (2014), onde realizaram a inoculação de flores de pimenta para recuperação de *Xanthomonas euvesicatoria* de sementes ao final do ciclo

de cultivo da planta, o que também foi realizado com *Xanthomonas campestris* pv. *campestris*, agente causal da podridão das brássicas (van Der Wolf et al. 2019).

A detecção de fitobactérias e demais fitopatógenos passou por diversas adaptações ao longo dos anos. Inicialmente a visualização de sintomas era o meio comumente aplicado para determinação de doenças em plantas, posteriormente aliado ao cultivo de microrganismos em meios de cultura específicos que possibilitava a análise morfológica de microrganismos cultiváveis (Bashan e Assoulina 1983). No âmbito da bacteriologia de plantas, os testes bioquímicos e sorológicos possibilitaram a detecção e determinação de diferentes espécies fitopatogênicas (Popović et al. 2010; Shahbaz et al. 2015). No entanto, essas metodologias podem apresentar falhas que em muitas ocasiões podem levar a erros de classificação e identificação. Desde a década de 80 com a descoberta da PCR por Mullis et al. (1986), a detecção molecular de fitopatógenos possibilitou maior eficiência e especificidade na diagnose de doenças em plantas. Ao longo dos anos a PCR passou por diversas adaptações dando origem à diferentes técnicas de amplificação de ácido nucleico, amplamente aplicadas na detecção de fitopatógenos como é o caso da qPCR, RT-qPCR e LAMP (Notomi et al. 2000; Nazrenko et al. 2002; Bookout et al. 2006). Os métodos moleculares e sorológicos como ELISA (*Enzyme-linked immunosorbent assay*), apesar de possibilitarem uma detecção específica de fitopatógenos, são métodos destrutivos, que inviabilizam o reaproveitamento de amostras (Mishra et al. 2020).

A diagnose precoce de doenças em tempo real utilizando sensores não destrutivos, como no caso das imagens hiperespectrais, tem avançado desde a última década, sendo uma possibilidade que pode favorecer diretamente os cultivos agrícolas reduzindo perdas por fitopatógenos (Sankaran et al., 2010; Martinelli et al. 2015). A diagnose de doenças em sementes causadas por patógenos de plantas envolvendo a aplicação de imagens hiperespectrais é realidade em inúmeros estudos. A detecção de fungos em sementes por hiperespectro foi

demonstrado por Daskalov et al. (2019) na detecção de *Fusarium* sp. na cultura do milho. Em vírus de planta, a detecção por hiperespectro também é algo bastante utilizado atualmente, como feito por Lu et al. (2018), que por imagens hiperespectrais desenvolveram um modelo para detecção do Tomato yellow leaf curl vírus em tomateiro. Gu et al. (2019) utilizaram as imagens hiperespectrais aliadas ao aprendizado de máquina na detecção do Tomato spotted wilt virus em tabaco. Lee et al. (2016) utilizaram a técnica de infravermelho próximo e imagens hiperespectrais para detecção de Cucumber green mottle mosaic virus em sementes de melancia e obtiveram 83,3 % de acurácia na distinção de sementes infectadas e sadias utilizando o sistema de hiperespectro. A detecção de fitobactérias em sementes por imagem hiperespectral juntamente com o aprendizado de máquina vem sendo aplicada em diferentes patossistemas, como na detecção de *Acidovorax citrulli* em melancia e *Xanthomonas oryzae* pv. *oryzae* em arroz (Lee et al. 2017; Baek et al. 2019; Zhang et al. 2020).

As imagens hiperespectrais aliadas ao aprendizado de máquina, automatizam a análise de dados de imagens, tornando mais eficiente a identificação dos comprimentos de onda de interesse e a aplicação prática das informações em larga escala. Como Li et al. (2017) que realizaram a diagnose de sementes de melancia infectadas por *Acidovorax avenae* subsp. *citrulli*, utilizando imagens hiperespectrais e predição por algoritmos de aprendizado de máquina, evidenciando uma acurácia de 91,7 % e 90,5 % usando os modelos PLS-DA e LS-SVM. A aplicação de sensores hiperespectrais associado a algoritmos de inteligência artificial em laboratórios ou em unidades de tratamento de sementes podem auxiliar na detecção precoce de fitopatógenos e contribuir para aplicação de estratégias de manejo adequada. O objetivo principal do presente capítulo foi desenvolver e avaliar metodologias para a detecção e diferenciação de sementes de tomateiro sadias e infectadas pela bactéria *Xanthomonas euvesicatoria* pv. *perforans*, empregando técnicas de inoculação em frutos e sementes e analisando imagens hiperespectrais para identificar assinaturas espectrais características das

sementes infectadas. Adicionalmente, o trabalho busca explorar a aplicação de modelos de *l* para classificar e distinguir sementes infectadas de sadias, com foco na acurácia e precisão de cada método de inoculação e análise espectral.

2. MATERIAL E MÉTODOS

2.1 Produção de sementes de tomateiro

Plantas da cv. Santa Cruz foram produzidas em casa de vegetação na Embrapa Hortaliças, Gama, DF. Foram cultivadas a partir de sementes em bandejas de 128 células de poliestireno até a idade fisiológica de quatro folhas verdadeiras. O plantio das mudas foi feito para vasos de 5 L onde ficaram durante todo o período do experimento. A adubação foi realizada a cada dez dias até o término do experimento e diante necessidade. As plantas foram mantidas em condições de casa de vegetação (25°C - 30°C).

2.2. Cepa de *Xanthomonas euvesicatoria* pv. *perforans* (*Xep*)

Foi utilizada a cepa de *Xep* EH-2022-31 da coleção de Bactérias Fitopatogênicas de Hortaliças da Embrapa Hortaliças obtida de folíolos de mudas comerciais sintomáticas. A identificação da espécie da EH-2022-31 foi realizada por meio de PCR multiplex utilizando primers específicos que diferenciam *Xanthomonas* spp. associadas a mancha bacteriana do tomateiro (Araújo et al. 2012).

2.3. Inoculação

A inoculação em frutos da cv. Santa Cruz para produção de sementes de tomateiro infectadas foi realizada utilizando três metodologias distintas. 1. Frutos de tomateiro ainda imaturos (8 a 26 mm de diâmetro) produzidos em casa de vegetação foram inoculados externamente com pincel n°2 na concentração de 1×10^8 UFC/mL da cepa *Xep* EH-2022-31 e mantidos em câmara úmida por cerca de 72 h. Como controle negativo foram pincelados frutos

com água destilada estéril. Após a inoculação foi avaliada a incidência de sintomas em frutos e quando maduros colhidos para extração de sementes. 2. Utilizando uma seringa hipodérmica (1 mL) foi aplicado 100 μ L do inóculo da cepa bacteriana (EH-2022-31) na concentração de 1×10^2 UFC/mL. A injeção do inóculo foi realizada na parte superior do fruto diretamente atingindo a placenta. Foi avaliada a incidência de sintomas ao longo dos dias. Após maturação dos frutos foi realizada a colheita e extração de sementes (da Silva et al. 2013). 3. Aplicação de 10 μ L de inóculo com concentração de 1×10^6 UFC/mL da cepa *Xep* EH-2022-31 na ponta dos pistilos de flores recém-abertas das plantas de tomateiro. Foram avaliadas a quantidade de flores que permaneceram nas plantas e que geraram fruto.

2.4. Extração de sementes

A extração foi realizada baseada no protocolo estabelecido pela International Seed Federation. Inicialmente as sementes foram retiradas dos frutos e acondicionadas em beckers cobertos com papel filtro para fermentação em temperatura ambiente, no intuito de retirar a mucilagem envolta na semente (Clemente e Boiteux 2012). Após 48 horas as sementes passaram por lavagem e foram expostas a secagem e posteriormente acondicionadas em temperatura de 4 °C até a captura da imagem hiperespectral.

2.5. Inoculação artificial de sementes

Para a infiltração da cepa *Xep* (EH-2022-31) em sementes sadias de tomateiro cv. Santa Cruz, a suspensão bacteriana foi obtida a partir de cultura pura cultivada em meio NA, mantida em câmara de crescimento a 28 °C/72 h. Após o crescimento bacteriano, sementes sadias foram inoculadas artificialmente na concentração de inóculo de 10^8 UFC mL. As sementes foram imersas na suspensão bacteriana e aplicado um vácuo (20 Hg) durante 5 minutos (Xu et al. 2010; Srivastava et. al. 2021). Após liberação do vácuo a suspensão foi drenada das sementes,

em seguida colocadas sob papel filtro para secagem ao ar por 24 horas. Posteriormente foram armazenadas em sacos de papel na temperatura de 4°C até a captura da imagem.

2.6. Captura de imagens hiperespectrais

Para a captura de imagens foi utilizada a câmera hiperespectral Specim FX10e. O modelo FX10e opera na luz visível e infravermelho próximo, na faixa espectral de 400-1000 nm. A câmera foi acoplada com base na altura ideal e focada manualmente, o posicionamento das luzes foi ajustado com o *software* Breeze (Prediktera AB). De cada um dos tratamentos (injeção, pincel e vácuo) foram utilizadas 500 sementes sadias e 500 infectadas. As sementes foram dispostas sob a esteira do LabScanner 40x20 (Specim) para a captura da imagem, como controle foram utilizadas sementes confirmadamente sadias da mesma cultivar, para distinguir possíveis alterações. Anterior a captura de imagens foram adquiridas as imagens de referência (branca/preta) com a tampa da câmera fechada. A imagem capturada foi visualizada em um computador acoplado a câmera, através do *software* Breeze que irá transformar os dados em pixels, formando uma imagem composta (*hiper cube*). Cada imagem a ser obtida é composta por um bloco de imagem de 224 bandas com intervalos de 2,7 nm no intervalo de 400 até 1000 nm na dimensão espectral. Para adquirir a emissão espectral real de cada semente, a influência dos padrões preto e branco foram retirados para obtenção da imagem calibrada (Kim et al. 2001). A segmentação da imagem, seleção de ROI e normalização e padronização dos dados pelo método SNV (Standard Normal Variate) foi realizada no *software* Breeze.

2.7. Análise de dados espectrais

Para a análise que identifique regiões do espectro que apresentem distinções significativas entre sementes sadias e doentes da cv. Santa Cruz, dois parâmetros foram aplicados: (I) diferença espectral, calculada pela diferença do valor médio da reflectância de sementes sadias e valor médio de reflectância de sementes infectadas pela cepa bacteriana de *Xep* EH-2022-31,

em cada comprimento de onda; (II) valor de sensibilidade, calculado pela divisão do valor médio da reflectância de sementes infectadas e sementes sadias, em cada comprimento de onda. Os dados de reflectância normalizados de sementes foram obtidos no software Breeze.

2.8. Métodos de classificação

O processamento de dados por aprendizado de máquina foi efetuado no Google Colab (Google Colaboratory), uma plataforma de notebooks em nuvem que permite executar códigos em *Python* diretamente no navegador. Foram usados três diferentes modelos de classificação: Support Vector Machine (SVM), Random Forest (RF) e Decision Tree (DT) (Cortes e Vapnik 1995; Breiman 2001; Quinlan 1986). Para o SVM os parâmetros de regularização (C) e kernel (gamma) foram ajustados. O parâmetro C controla o equilíbrio entre maximizar a margem de separação dos dados e minimizar os erros de classificação, enquanto o gamma estabelece o alcance da influência de cada exemplo de treinamento, afetando o formato do hiperplano de separação. No modelo RF os parâmetros otimizados incluíram o número de estimadores, a profundidade máxima das árvores e o número mínimo de amostras necessárias para dividir um nó e ser uma folha (ponto final de decisão do modelo). No modelo DT, os parâmetros calibrados foram a profundidade máxima, o número mínimo de amostras necessários para dividir um nó, e o número mínimo de amostras para formar uma folha (ponto terminal da árvore de decisão). No estudo, o conjunto de dados de treinamento foi de 70% e o de teste de 30%, sendo inserido dados de entrada de sementes sadias e sementes infectadas pela cepa bacteriana de *Xep* (EH-2022-31), por diferentes métodos de inoculação. O desempenho dos modelos foi avaliado na matriz de confusão, pelos valores de *Accuracy* (acurácia) (Equação 1); *Precision* (precisão) (Equação 2); *F1-Score* (índice F1) (Equação 3) e *Recall* (revocação ou sensibilidade) (Equação 4). Essas métricas são fundamentadas na observação de verdadeiros positivos (*true positives*), verdadeiros negativos (*true negatives*), falsos positivos (*false positives*) e falsos negativos (*false negatives*) (Cunha et al. 2023; Reis-Pereira et al. 2023)

1. Accuracy = $\frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$
2. Precision = $\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$
3. F1 Score = $2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
4. Recall = $\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$

3. RESULTADOS

3.1. Inoculação

A injeção em frutos não resultou em sintomas típicos da espécie bacteriana, notou-se em alguns frutos o escurecimento interno, porém, não afetou a extração de sementes. A inoculação por pincel resultou em sintomas típicos da mancha bacteriana (Jones et al. 1991), que surgiram 15 dias pós inoculação e continuaram a progredir ao longo da maturação de frutos (Figura 1). As sementes provenientes da extração dos frutos inoculados por pincel, por injeção e de forma artificial não apresentaram nenhuma alteração aparente. A inoculação pelas flores não resultou em frutos já que todas abortaram.



Figura 1. Sintomas típicos de mancha bacteriana do tomateiro em frutos inoculados por pincel (A) Frutos imaturos 15 dias pós inoculação por pincel. (B) Frutos maduros.

3.2. Análise de dados espectrais (Inoculação por pincel)

Os dados foram obtidos na faixa do espectro VNIR, entre 400 nm e 936 nm. A reflectância espectral normalizada (Figura 2A) mostra uma tendência de uma maior reflectância ao longo do espectro medido para as sementes sadias. Nas sementes inoculadas o comportamento da curva inicia de forma similar às sadias em torno de 400 nm, ocorrendo um distanciamento entre as curvas mais evidente por volta dos 600 nm, as sementes inoculadas mantêm um comportamento uniforme de reflectância abaixo das sadias após os comprimentos de onda mais baixos. A diferença espectral normalizada (Figura 2B) apresenta um comportamento onde a curva inicia em 400 nm com uma diferença espectral em torno de zero e aumenta progressivamente até atingir um pico por volta de 668 nm. Após esse pico a diferença espectral diminui continuamente até o final do espectro medido, aproximando novamente de zero. A sensibilidade espectral normalizada (Figura 2C) expressa um comportamento em que a curva inicia com uma queda em torno de 534 nm, indicando menor sensibilidade espectral nesse comprimento, com um valor abaixo de 1,0. Após essa faixa a curva tende a subir até o final do espectro medido (936 nm).

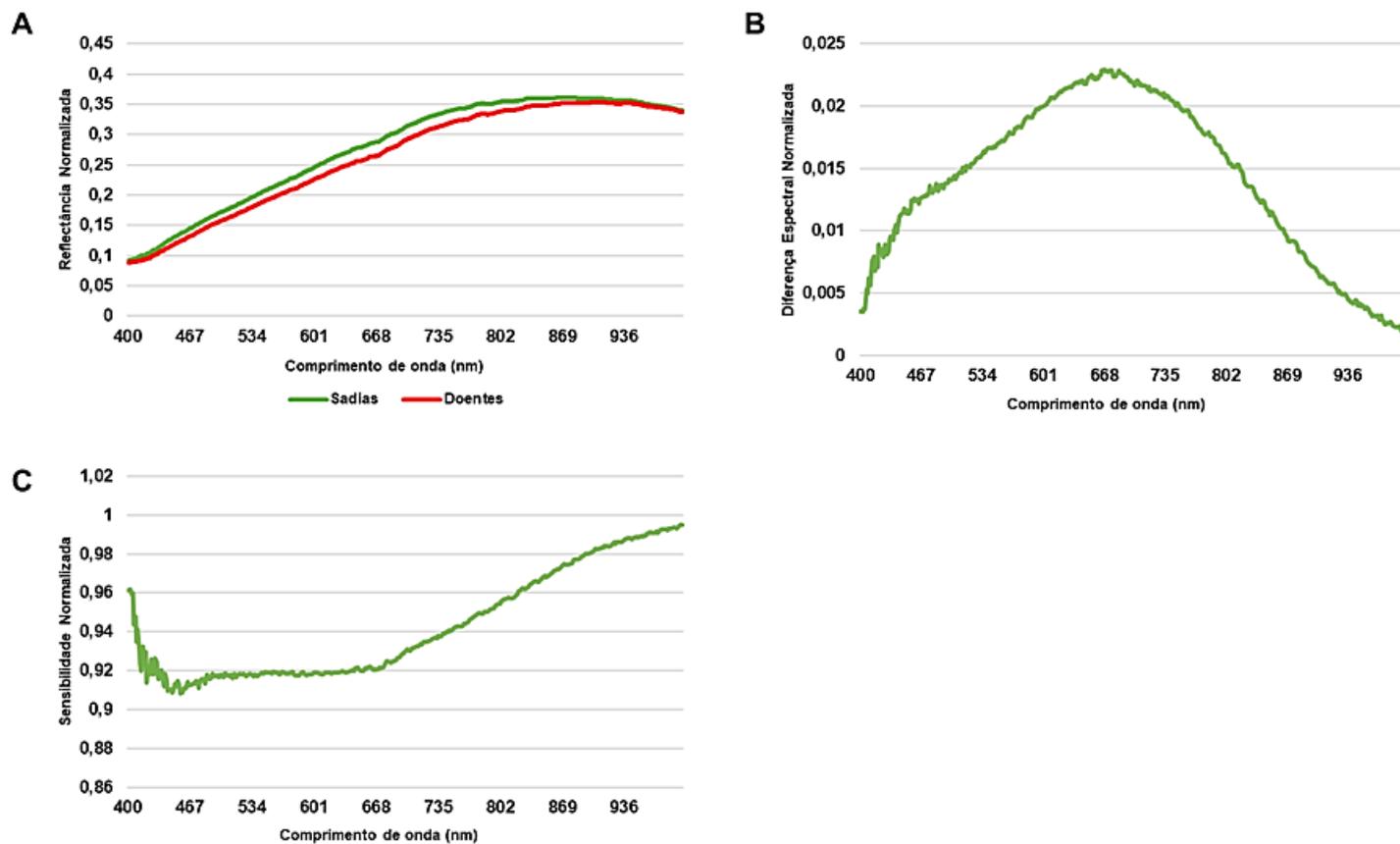


Figura 2. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com *Xep* pelo método de pincel em frutos. (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.

3.3. Análise de dados espectrais (Inoculação por injeção)

Na reflectância espectral normalizada (Figura 3A) ambas as curvas que representam as sementes sadias e sementes inoculadas apresentam um aumento na reflectância proporcional ao aumento dos comprimentos de onda, mostrando maior reflexão da luz em comprimentos de onda longos. Além disso, é observado que a curva das sementes sadias mantém um comportamento ligeiramente constante acima da curva das sementes inoculadas, ao longo de todo o espectro medido. O comportamento da diferença espectral normalizada (Figura 3B) apresenta um pico em torno de 668 nm, onde a diferença espectral entre sementes sadias e inoculadas é mais evidente, após essa faixa a curva apresenta uma diminuição na diferença espectral até o fim do espectro avaliado. A sensibilidade espectral normalizada (Figura 3C) exibe um comportamento inicial (400- 534 nm) acima de 1,0, indicando nessa faixa que as sementes infectadas refletem mais luz do que as sementes sadias. Após 534 nm ocorre um declínio na sensibilidade espectral, mais evidente em 668 nm atingindo o ponto mínimo de sensibilidade com o valor em torno de 0,98. Após esse ponto a sensibilidade tende a aumentar até o final do espectro medido.

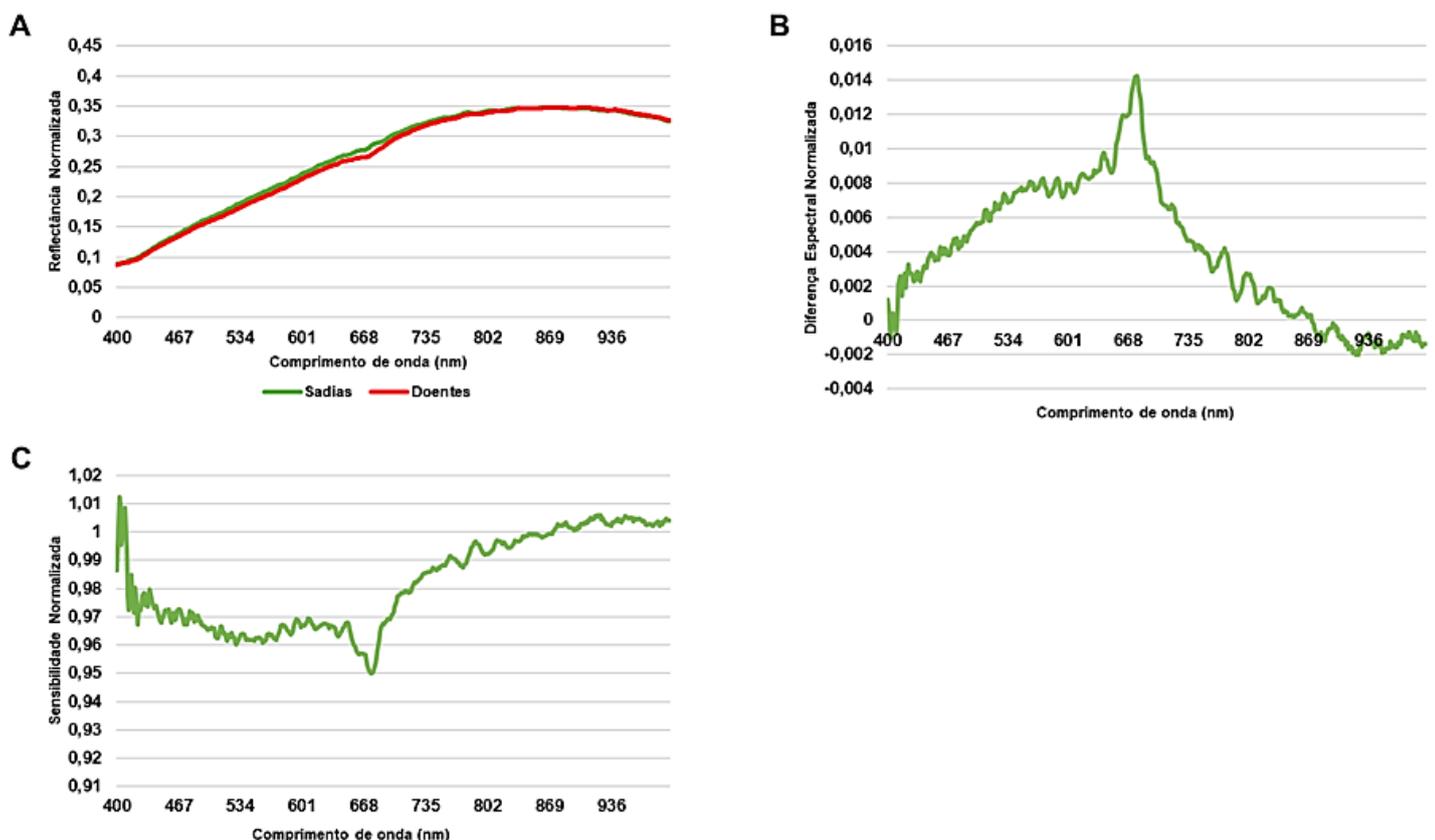


Figura 3. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com *Xep* pelo método de injeção em frutos. (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.

3.4. Análise de dados espectrais (Infiltração artificial de sementes)

Na reflectância espectral normalizada (Figura 4A) as curvas de sementes sadias e sementes inoculadas tendem a aumentar proporcional ao aumento dos comprimentos de onda. Além disso, é evidenciado que a curva das sementes inoculadas mantém um comportamento constante acima da curva das sementes sadias ao longo de todo o espectro medido, mostrando maior reflectância de sementes inoculadas em relação as sadias. A diferença espectral normalizada (Figura 4B) apresenta valores negativos em todos os comprimentos de onda, o que infere que as sementes inoculadas possuem uma reflectância média superior à das sementes sadias. A curva de diferença espectral tende a descer a partir de 400 nm, alcançando um pico de diferença negativa em torno de 467 nm, após esse pico a curva mostra uma ligeira elevação e posteriormente diminui nos comprimentos de onda mais longos, permanecendo negativa. O comportamento inicial da sensibilidade espectral normalizada (Figura 4C) mostra uma elevação nos comprimentos de onda mais baixos, próximo de 400 nm, indicando que nessa faixa a reflectância de sementes inoculadas tende a ser mais alta que das sadias. Conforme o comprimento de onda aumenta a sensibilidade espectral apresenta diminuição, ficando próxima de 1,0 em torno de 534 nm e constantemente em torno de 1,0 até 936 nm. Inferindo que após 534 nm a reflectância entre sementes sadias e inoculadas apresentam similaridade.

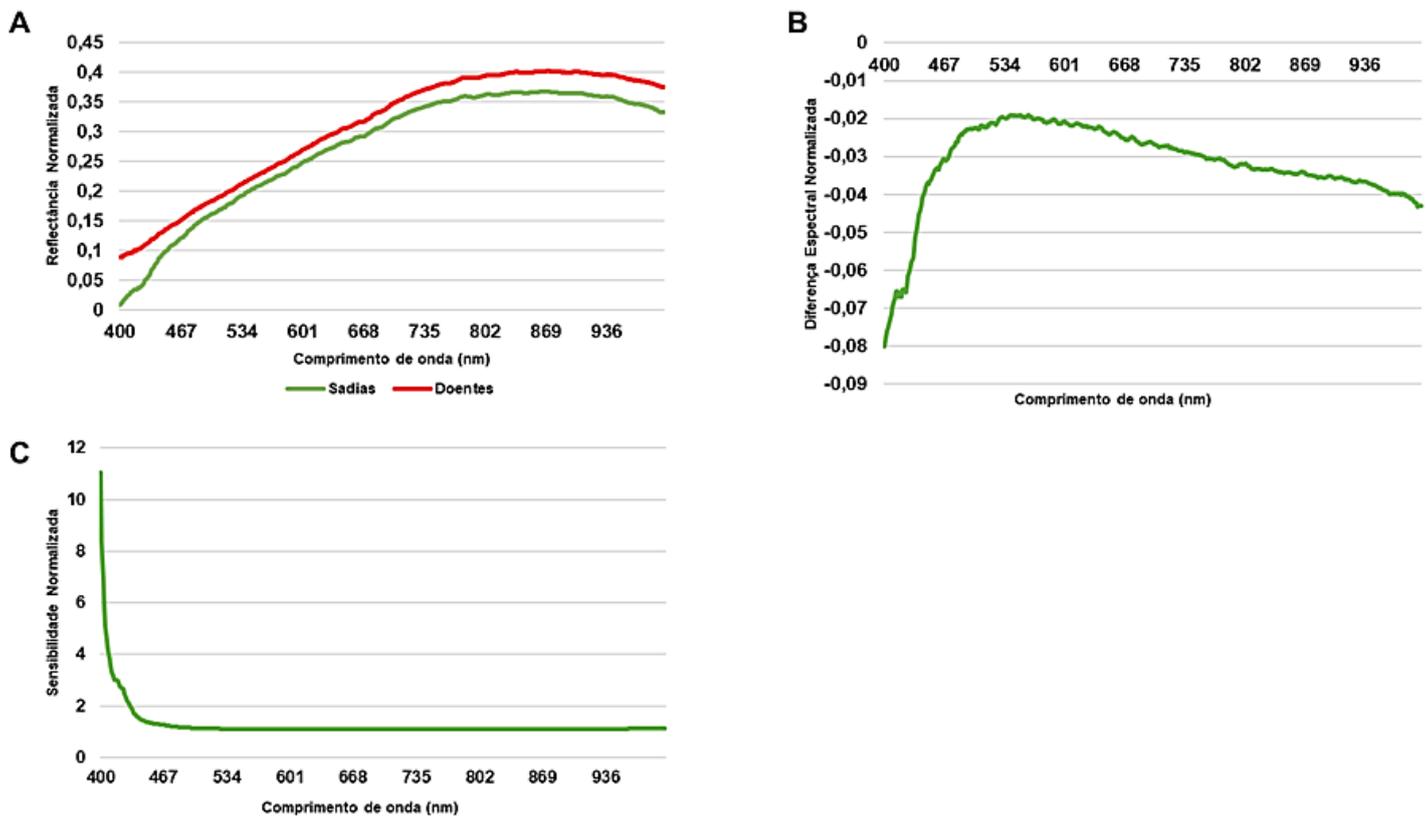


Figura 4. Assinatura espectral de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas (doentes) com *Xep* pelo método de infiltração artificial (vácuo). (A) Reflectância espectral normalizada (B) Diferença espectral normalizada. (C) Sensibilidade espectral normalizada.

3.5. Classificação dos dados

Os modelos de predição Support Vector Machine e Random Forest apresentaram desempenho satisfatório na classificação de dados de sementes de tomateiro cv. Santa Cruz sadias e infectadas pela *Xep* com diferentes métodos de inoculação (injeção do inóculo em frutos, contato do pincel com inóculo em frutos e infiltração artificial de sementes) (Figura 5). Na técnica de injeção do inóculo em frutos o modelo SVM mostrou acurácia de 97 % na classificação de dados de sementes sadias e infectadas, seguido de RF com 71 % e DT com 62 %. Na inoculação de frutos por pincel a classificação de dados de sementes sadias e infectadas apresentou uma acurácia de 98 % no SVM, 78 % no RF e 74 % no DT. Na infiltração da bactéria em sementes usando o método a vácuo os três modelos apresentaram 100 % de acurácia na classificação de sementes sadias e infectadas. Na análise de desempenho por meio das métricas *Precision*, *Recall* e *F1-Score*, o modelo RF exibiu um desempenho consistente na classificação

dos dados de sementes inoculadas pelo método de injeção em frutos (Tabela 1) e pincel em frutos (Tabela 2), no entanto, o SVM foi superior em ambos os métodos apresentando valores acima de 97% no *F1-Score*. Em contrapartida, na classificação do conjunto de dados de sementes infiltradas pelo método à vácuo (Tabela 3) todos os modelos avaliados exibiram um desempenho de 100% nas medições de *Precision*, *Recall* e *F1-Score*.

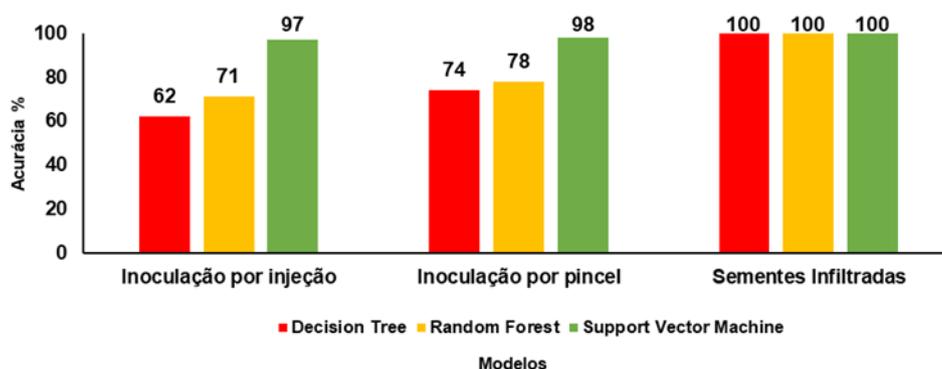


Figura 5. Desempenho dos modelos de *machine learning* na classificação de sementes de tomateiro (cv. Santa Cruz) sadias e inoculadas por diferentes métodos.

Tabela 1. Desempenho dos modelos na classificação de dados de sementes inoculadas por injeção em frutos.

Modelos	Classes	Precision	Recall	F1 score
DT	Infectada	65%	53%	59%
	Sadia	60%	71%	65%
RF	Infectada	72%	69%	71%
	Sadia	70%	73%	71%
SVM	Infectada	100%	94%	97%
	Sadia	95%	100%	97%

Tabela 2. Desempenho dos modelos na classificação de dados de sementes inoculadas por pincel em frutos.

Modelos	Classes	Precision	Recall	F1 score
DT	Infectada	75%	72%	74%
	Sadia	72%	76%	74%
RF	Infectada	80%	76%	78%
	Sadia	77%	80%	78%
SVM	Infectada	97%	99%	98%
	Sadia	99%	97%	98%

Tabela 3. Desempenho dos modelos na classificação de dados de sementes inoculadas à vácuo.

Modelos	Classes	Precision	Recall	F1 score
DT	Infectada	100%	100%	100%
	Sadia	100%	100%	100%
RF	Infectada	100%	100%	100%
	Sadia	100%	100%	100%
SVM	Infectada	100%	100%	100%
	Sadia	100%	100%	100%

4. DISCUSSÃO

Em termos epidemiológicos, a transmissão de microrganismos fitopatogênicos por sementes representa um importante meio de dispersão de patógenos, contribuindo para surtos de doenças em plantas. Esclarecer a dinâmica de infecção de sementes e propor métodos eficientes de detecção de patógenos é primordial para a produção agrícola, favorecendo a mitigação de perdas econômicas e produtivas.

No presente estudo, avaliou-se a detecção da mancha bacteriana do tomateiro em sementes de tomate da cv. Santa Cruz, inoculadas por diferentes métodos (pincel, injeção e infiltração) com *Xep*, aplicando sensores hiperespectrais em conjunto com a classificação de dados por aprendizado de máquina. Além disso, foi realizada a análise de dados de reflectância espectral, com o objetivo de expor diferenças significativas entre sementes sadias e inoculadas pela bactéria.

Os modelos de predição utilizados apresentaram variações na eficácia de classificação de sementes de tomateiro sadias e inoculadas conforme cada método de inoculação aplicado, indicando a relevância da escolha do modelo adequado de acordo com cada cenário específico de condições do material vegetal. O modelo SVM apresentou superioridade comparado aos modelos DT e RF, na classificação de dados de sementes inoculadas por injeção em frutos e pincel em frutos, com acurácia de 97 % e 98 %, respectivamente. A superioridade do modelo SVM, pode ser explicada pela sua capacidade de criar hiperplanos ótimos que separam de forma eficiente as classes em espaços de alta dimensão (Pathak et al. 2022; Ali et al. 2024). Esse resultado, sugere que o SVM pode ser mais apropriado para distinção de sementes de tomateiro sadias e inoculadas pela bactéria. A alta acurácia na detecção de doenças em sementes usando o modelo SVM já foi relatada por Qiao et al. (2017) analisando a infecção natural de sementes de amendoim infectadas por diferentes espécies fúngicas. Além desse, Seo et al. (2019)

realizaram a classificação de sementes de melancia sadias e infectadas naturalmente com Cucumber green mottle mosaic virus e obtiveram acurácia acima de 90 % usando o modelo LS-SVM em conjunto.

Na metodologia de infiltração artificial de sementes com a cepa bacteriana de *Xep*, todos os modelos apresentaram 100% de acurácia na classificação de sementes sadias e inoculadas. A infiltração de sementes garante a eficiência da inoculação do patógeno, o que pode ser um facilitador na uniformidade da infecção (Wu et al. 2020). Por outro lado, a inoculação artificial de sementes pode inferir em resultados tendenciosos, já que não expressam a realidade ou nem mesmo se aproximam de uma infecção natural de sementes em campo (Ferreira et al. 2024). A escolha do método de inoculação, que diretamente afeta a distribuição espacial do patógeno nas amostras é crucial. Wu et al. (2020) mostraram a influência da inoculação artificial e infecção natural de sementes de arroz com o falso carvão (*Villosiclava virens*), onde evidenciaram que na faixa do NIR sementes inoculadas artificialmente apresentavam um comportamento de reflectância menor do que as sementes naturalmente infectadas no campo e inferiram que o fator poderia ser em decorrência da menor concentração da suspensão esporos em sementes inoculadas em laboratório.

Nos estudos envolvendo a análise da infecção causadas por patógenos em sementes utilizando aprendizado de máquina, a avaliação do desempenho dos modelos é exibida no geral pela comparação dos valores de acurácia de acordo com cada modelo testado (Lee et al. 2017; Qiao et al. 2017; Wu et al. 2024). Entretanto, um estudo de discriminação de sementes de tomate pertencentes a diferentes cultivares, buscou avaliar a performance de diferentes modelos na classificação dos dados utilizando *Precision*, *Recall* e *F1-Score*, além da acurácia, evidenciando o melhor desempenho para *Random Forest* e *Multilayer Perceptron* (Ropelewska e Jan Piecko 2022)

A análise da reflectância espectral de sementes sadias e inoculadas pelos métodos de injeção e pincel, exibiram maior reflectância das sementes sadias ao longo de boa parte do espectro medido, o que pode inferir em alguma alteração química da semente em decorrência da infecção bacteriana (Baek et al. 20219). Existem muitos trabalhos com análises de sementes, porém a maioria com foco maior no uso do NIR para estudos de viabilidade ou identificação de variedades, por meio da análise de conteúdo de proteínas, lipídeos, teor de fibras. No entanto, a interpretação desses dados por *machine learning* pode ser complexa, levando a algoritmos de calibração de “caixa preta”, que não podem ser correlacionadas com os dados, levando à erros em análises em tempo real (Reddy et al. 2022). No estudo de Shrestha et al. (2017) envolvendo a classificação de sementes viáveis e não viáveis de tomateiro com uso do NIR e a aplicação do modelo PLS-DA, evidenciaram que comprimentos de onda (1910 -1980 nm), que fornecem informações químicas da água e de proteínas, são essenciais para a discriminação entre sementes viáveis e não viáveis.

Na análise de diferença espectral de sementes sadias e inoculadas pelos métodos de injeção e pincel evidenciou-se um pico por volta de 668 nm. A sensibilidade espectral mostra um aumento após a mesma faixa espectral em ambos os métodos de inoculação. Esse padrão pode ser particularmente útil, pois, indica o ponto em que ocorrem alterações mais evidentes na reflectância entre sementes sadias e inoculadas. Essa observação é sustentada por estudos que exploraram a análise de dados de reflectância para avaliar a sanidade de sementes identificando padrões específicos que indicam estresse bióticos e abióticos (Feng et al. 2019).

Em contraste aos métodos mais “naturais” de inoculação usados, as sementes inoculadas por infiltração à vácuo exibem valores superiores em comparação com as sadias ao longo do espectro. Esse comportamento também foi evidenciado em um estudo de Lee et al. (2017) em

que realizaram a inoculação artificial de sementes de melancia para avaliar a infecção causada por *Acidovorax avenae* subsp. *citrulli* e observaram que as sementes infectadas exibiam maior reflectância que as sementes sadias, em especial na região do visível. A diferença espectral entre as sementes sadias e doentes revelou uma maior reflectância em sementes inoculadas, especialmente em comprimentos de onda menores (400 nm - 467 nm), o que pode ser atribuído à formação de compostos bacterianos específicos na superfície das sementes ou a modificações estruturais induzidas pela infecção bacteriana, afetando a dispersão e absorção da luz (Rudrappa et al. 2008; Carezzano et al. 2023). Esse fato pode ser explicado pelo método de inoculação artificial que propicia uma alta carga bacteriana sob as sementes, com uma distribuição relativamente maior do inóculo quando comparado a métodos de inoculação menos invasivos, influenciando diretamente nas propriedades de reflectância das sementes (Ferreira et al. 2024). O padrão da curva de sensibilidade espectral mostrou que as sementes sadias e inoculadas têm reflectâncias semelhantes em comprimentos de onda mais altos, sugerindo que a influência da infecção sobre a reflectância pode ser mais limitada a faixas menores do espectro.

Os resultados expostos sugerem que a combinação de imagens hiperespectrais com análise de dados por aprendizado de máquina, em especial com o modelo SVM, pode ser útil como uma ferramenta de diagnose não destrutiva, favorecendo significativamente o monitoramento da doença em campo. Entre os métodos de inoculação para a geração de protocolos de diagnose por imagem hiperespectral, a infiltração artificial pode não refletir completamente a complexidade de infecções naturais causadas por *Xep* em sementes, o que é uma importante limitação da técnica a ser considerada.

Portando, recomenda-se a continuação da pesquisa para refinar as técnicas de detecção e explorar sua aplicabilidade de forma mais ampla, em diferentes condições de inoculação ou em sementes infectadas naturalmente em campos de tomateiro com surtos da mancha bacteriana, visando a otimização dessa tecnologia para aplicação em um contexto prático seja em viveiros

de produção de mudas, em unidades de tratamento de sementes ou mesmo em fazendas de produção de tomate.

5. REFERÊNCIAS

- Ali F, Razzaq A, Tariq W, Hameed A, Rehman A, Razzaq K, Ondrasek G (2024) Spectral Intelligence: AI-Driven Hyperspectral Imaging for Agricultural and Ecosystem Applications *Agronomy* 14(10): 2260.
- Araújo ER, Costa JR, Ferreira MASV, Quezado-Duval AM (2017) Widespread distribution of *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. *Plant Pathology* 66(1): 159–168. doi: <https://doi.org/10.1111/PPA.12543>
- Baek I, Kim MS, Cho BK, Mo C, Barnaby JY, McClung AM, Oh M (2019) Selection of Optimal Hyperspectral Wavebands for Detection of Discolored, Diseased Rice Seeds. *Applied Sciences* 9(5): 1027. <https://doi.org/10.3390/APP9051027>
- Bashan Y, Assouline I (1983) Complementary bacterial enrichment techniques for the detection of *Pseudomonas syringae* pv. *tomato* and *Xanthomonas campestris* pv. *vesicatoria* in infested tomato and pepper seeds. *Phytoparasitica* 11(3-4):187–193. doi:10.
- Bookout AL, Cummins CL, Mangelsdorf DJ, Pesola JM, Kramer MF (2006) High-Throughput Real-Time Quantitative Reverse Transcription PCR. *Current Protocols in Molecular Biology* 73:15.8.1-15.8.28. <https://doi.org/10.1002/0471142727.MB1508S73>
- Carezzano ME, Paletti Rovey MF, Cappellari LDR, Gallarato LA, Bogino P, Oliva ML M, Giordano W (2023) Biofilm-forming ability of phytopathogenic bacteria: a review of its involvement in plant stress. *Plants* 12(11): 2207.
- Chai A, Ben H, Shi Y, Xie X, Li L, Li B (2020) Quantification of Viable Cells of *Pseudomonas syringae* pv. *tomato* in Tomato Seed Using Propidium Monoazide and a Real-time PCR Assay. *Plant Disease*. doi:10.1094/pdis-11-19-2397-re
- Clemente FMVT, Boiteux LS (2012) *Produção de Tomate para Processamento Industrial* 1th Ed. Embrapa, Brasil, Brasília DF
- Daskalov PI, Kirilova E, Daskalov P, Mancheva V, Georgieva T (2019). Comparative study on the application of visible and near infrared hyperspectral imaging for *Fusarium* disease assessment of corn seeds. *Wseas Transactions on Electronics* 10.
- Dutta B, Gitaitis R, Sanders H, Booth C, Smith S, Langston DB (2014) Role of blossom colonization in pepper seed infestation by *Xanthomonas euvesicatoria*. *Phytopathology* 104(3): 232–239. <https://doi.org/10.1094/PHYTO-05-13-0138-R>
- Feng L, Zhu S, Liu F, He Y, Bao Y, Zhang C (2019) Hyperspectral imaging for seed quality and safety inspection: A review. *Plant methods* 15: 1-25.
- Ferreira LDC, Carvalho ICB, Jorge LADC, Quezado-Duval AM, Rossato M (2024). Hyperspectral imaging for the detection of plant pathogens in seeds: recent developments and challenges. *Frontiers in Plant Science* 15: 1387925.
- Gu Q, Sheng L, Zhang T, Lu Y, Zhang Z, Zheng K, Hu H, Zhou H (2019) Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique

- and machine learning algorithms. *Computers and Electronics in Agriculture* 167: 105066. doi: <https://doi.org/10.1016/J.COMPAG.2019.105066>
- Inoue-Nagata AK, Lima MF (2021) Doenças causadas por vírus. In: Lopes CA (Eds.) Doenças do tomateiro. Embrapa, Brasil, Brasília DF
- Lee H, Kim MS, Lim HS, Park E, Lee WH, Cho BK (2016) Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: Application to seeds of the “Sambok Honey” cultivar. *Biosystems Engineering* 148: 138-147. doi: <https://doi.org/10.1016/J.BIOSYSTEMSENG.2016.05.014>
- Lee H, Kim MS, Song YR, Oh CS, Lim HS, Lee WH (2017). Non-destructive evaluation of bacteria-infected watermelon seeds using visible/near-infrared hyperspectral imaging. *J. Sci. Food Agric.* 97: 1084–1092. doi: 10.1002/jsfa.7832
- Lopes CA (2021) Doenças do tomateiro 3th Ed. Embrapa, Brasil, Brasília DF
- Lu J, Zhou M, Gao Y, Jiang H (2018) Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves. *Precision Agriculture* 19(3): 379–394. doi: <https://doi.org/10.1007/S11119-017-9524-7/TABLES/3>
- Martinelli F, Scalenghe R, Davino S, Panno S, Scuderi G, Ruisi P, Villa P, Stroppiana D, Boschetti M, Goulart LR, Davis CE, Dandekar AM (2015) Advanced methods of plant disease detection. A review. *Agron Sustain Dev* 35:1–25. doi: <https://doi.org/10.1007/s13593-014-0246-1>
- Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosyst. Eng.* 164, 49–67. doi: 10.1016/j.biosystemseng.2017.09.009
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273. doi: 10.1101/SQB.1986.051.01.032
- Nazarenko, I., Lowe, B., Darfler, M., Ikononi, P., Schuster, D., Rashtchian, A. (2002). Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. *Nucleic Acids Res.* 30, 37. doi: 10.1093/nar/30.9.e37
- Notomi T, Okayama, H, Masubuchi, H, Yonekawa T, Watanabe K, Amino N. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28, 63. doi: 10.1093/nar/28.12.e63
- Pathak DK, Kalita SK, Bhattacharya DK (2022) Hyperspectral image classification using support vector machine: a spectral spatial feature based approach. *Evolutionary Intelligence* 1-15.
- Pohronezny K, Stall RE, Canteros BI, Kegley M, Datnoff LE, Subramanya R (1992) Sudden shift in the prevalent race of *Xanthomonas campestris* pv. *vesicatoria* in pepper fields in southern Florida. *Plant Disease* 76: 118–20. doi: 10.1094/PD-76-0118

- Popović T, Balaz J, Nikolic Z, Starovi M, Gavrilovi V, Aleksi G, Vasi M, Živkovi S (2010) Detection and identification of *Xanthomonas axonopodis* pv. *phaseoli* on bean seed collected in Serbia. *African Journal of Agricultural Research* 5:2730–2736
- Qiao X, Jiang J, Qi X, Guo H, Yuan D (2017). Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food Chem.* 220: 393–399. doi: 10.1016/j.foodchem.2016.09.119
- Reddy P, Guthridge KM, Panozzo J, Ludlow EJ, Spangenberg GC, Rochfort SJ (2022) Near-infrared hyperspectral imaging pipelines for pasture seed quality evaluation: An overview. *Sensors* 22(5): 1981.
- Rego CHQ, França-Silva F, Gomes-Junior FG, Moraes MHDD, Medeiros ADD, Silva CBD (2020) Using multispectral imaging for detecting seed-borne fungi in cowpea. *Agriculture* 10(8): 361.
- Ropelewska E, Piecko J (2022) Discrimination of tomato seeds belonging to different cultivars using machine learning. *European Food Research and Technology* 248(3): 685-705.
- Rudrappa T, Biedrzycki ML, Bais HP (2008) Causes and consequences of plant-associated biofilms. *FEMS microbiology ecology* 64(2): 153-166.
- Sankaran, S., Mishra, A., Ehsani, R., Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13. doi: 10.1016/j.compag.2010.02.007
- Seo Y, Lee H, Bae HJ, Park E, Lim HS, Kim MS (2019) Optimized multivariate analysis for the discrimination of cucumber green mosaic mottle virus-infected watermelon seeds based on spectral imaging. *Biosyst. Eng* 44: 95–102. doi: 10.1007/s42853-019-00019-9
- Shahbaz MU, Mukhtar T, Begum N (2015) Biochemical and Serological Characterization of *Ralstonia solanacearum* Associated with Chilli Seeds from Pakistan. *International Journal of Agriculture & Biology* 17(1).
- Sharma DK, Agrawal K (2014) Incidence and histopathological study of *Xanthomonas axonopodis* pv. *vesicatoria* in tomato (*Lycopersicon esculentum* Mill.) seeds. *Int J Agric Technol* 10:233–42.
- Shrestha S, Deleuran LC, Gislum R (2017) Separation of viable and non-viable tomato (*Solanum lycopersicum* L.) seeds using single seed near-infrared spectroscopy. *Computers and Electronics in Agriculture* 142: 348-355.
- Tancos MA, Chalupowicz L, Barash I, Manulis-Sasson S, Smart CD (2013) Tomato fruit and seed colonization by *Clavibacter michiganensis* subsp. *michiganensis* through external and internal routes. *Applied and Environmental Microbiology* 79(22): 6948–6957. doi: https://doi.org/10.1128/AEM.02495-13/SUPPL_FILE/ZAM999104858SO1.PDF
- van der Wolf J, Kastelein P, da Silva Júnior TAF, Lelis FV, van der Zouwen P (2019) Colonization of siliques and seeds of rapid cycling *Brassica oleracea* plants by

Xanthomonas campestris pv. *campestris* after spray-inoculation of flower clusters. *European Journal of Plant Pathology* 154: 445–461. <https://doi.org/10.1007/s10658-019-01668-4>

- Wu N, Jiang H, Bao Y, Zhang C, Zhang J, Song W (2020) Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions. *Sens. Actuators B Chem* 308: 127696. doi: 10.1016/j.snb.2020.127696
- Wu, N., Weng, S., Xiao, Q., Jiang, H., Zhao, Y., He, Y. (2024). Rapid and accurate identification of bakanae pathogens carried by rice seeds based on hyperspectral imaging and deep transfer learning. *Spectrochim Acta A Mol. Biomol. Spectrosc.* 311, 123889. doi: 10.1016/j.saa.2024.123889
- Xu X, Miller SA, Baysal-Gurel F, Gartemann KH, Eichenlaub R, Rajashekara G (2010) Bioluminescence imaging of *Clavibacter michiganensis* subsp. *michiganensis* infection of tomato seeds and plants. *Applied and Environmental Microbiology* 76(12): 3978–3988. doi: <https://doi.org/10.1128/AEM.00493-10/ASSET/BCBF4754-443C-4A19-B7D4-F90BC0928A38/ASSETS/GRAPHIC/ZAM9991010230005.JPEG>
- Zhang J, Yang Y, Feng X, Xu H, Chen J, He Y (2020) Identification of Bacterial Blight Resistant Rice Seeds Using Terahertz Imaging and Hyperspectral Imaging Combined With Convolutional Neural Network. *Frontiers in Plant Science* 11: 821. <https://doi.org/10.3389/FPLS.2020.00821/BIBTEX>

6. CONSIDERAÇÕES FINAIS

O estudo destacou a viabilidade e eficácia do uso de imagens hiperespectrais combinados com modelos de aprendizado de máquina para a detecção precoce da mancha bacteriana em cotilédones e sementes de tomateiro, avaliando uma cultivar de polinização aberta, OP (Santa Cruz) e uma híbrida (HMX 7885). O modelo de predição SVM (Support Vector Machine), demonstrou capacidade superior na distinção e classificação de dados tanto de cotilédones sadios e infectados pela cepa bacteriana de *Xep* (EH-2022-31), quanto das sementes sadias e inoculadas por diferentes métodos. Tal desempenho sugere um grande potencial de aplicação desse modelo de forma prática no contexto da mancha bacteriana na produção de tomate. Os resultados mostram que enquanto o modelo SVM se destacou por sua robustez, os modelos RF (Random Forest) e DT (Decision Trees), apresentaram desempenhos inferiores, enfatizando a importância da seleção criteriosa de algoritmos para a análise de dados hiperespectrais com base em cada patossistema. Além disso, a análise detalhada do comportamento espectral por meio da análise de dados de reflectância espectral permitiu identificar mudanças fenotípicas significativas em sementes e cotilédones infectados que podem ser primordiais para o desenvolvimento de sistemas de alerta precoce da mancha bacteriana. A variabilidade nas repostas à infecção entre a cultivar e o híbrido, bem como o comportamento espectral de cada um, ressalta a necessidade de otimizar as técnicas de análise hiperespectral para se adaptarem às características específicas de cada planta, mesmo dentro de uma mesma espécie vegetal. Assim, o aprimoramento dos modelos, juntamente com a otimização das análises hiperespectrais, pode contribuir ainda mais para a precisão e a aplicabilidade dessas tecnologias no campo. Por fim, a combinação de imagens hiperespectrais e análises de dados avançadas apresenta uma promissora ferramenta de diagnose não destrutiva. Essa abordagem possibilita intervenções mais rápidas e precisas, aperfeiçoando o manejo da doença antes mesmo do aparecimento de sintomas visíveis. Portanto, recomenda-se a continuação da pesquisa para

refinar essas técnicas de detecção e expandir sua aplicabilidade em diferentes condições de inoculação e em condições de campo, visando maximizar o potencial desta tecnologia em práticas agrícolas sustentáveis e economicamente viáveis.