RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 28/04/2028.

UNESP – UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE PRESIDENTE PRUDENTE
FACULDADE DE CIÊNCIAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CARTOGRÁFICAS

MAURO MORATA BORTOLOTO JUNIOR

DETECTION OF HUANGLONGBING (HLB) IN ORANGE TREES USING HYPERSPECTRAL SIGNATURES AND MACHINE LEARNING MODELS

Presidente Prudente – SP
2026

MAURO MORATA BORTOLOTO JUNIOR

DETECTION OF HUANGLONGBING (HLB) IN ORANGE TREES USING HYPERSPECTRAL SIGNATURES AND MACHINE LEARNING MODELS

Advisor: Prof. Ana Paula Marques Ramos
Co-advisor: Prof. Lucas Prado Osco

Master's thesis presented to the Graduate Program in Cartographic Sciences ("Programa de Pós-Graduação em Ciências Cartográficas") at São Paulo State University (UNESP), School of Technology and Sciences.

Area of Concentration: Exact and Earth Sciences.

Presidente Prudente – SP

2026

B739d

Bortoloto Junior, Mauro Morata

Detection of Huanglongbing (HLB) in Orange Trees Using Hyperspectral Signatures and Machine Learning Models / Mauro Morata Bortoloto Junior. -- Presidente Prudente, 2026

100 p. : il., tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientadora: Ana Paula Marques Ramos

Coorientadora: Lucas Prado Osco

1. hyperspectral remote sensing. 2. Greening. 3. precision agriculture. 4. machine learning. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

Research Impact

This research advances digital agriculture by integrating hyperspectral remote sensing and machine learning for Huanglongbing (HLB) detection in citrus crops. The results may support large-scale monitoring systems, improve phytosanitary decision-making, reduce economic impacts caused by the disease, and strengthen the sustainability of the Brazilian citrus production chain.

Impacto da Pesquisa

Esta pesquisa contribui para o avanço da agricultura digital ao integrar sensoriamento remoto hiperespectral e aprendizado de máquina na detecção de Huanglongbing (HLB) em citros. Os resultados podem subsidiar sistemas de monitoramento em larga escala, apoiar a tomada de decisão no manejo fitossanitário, reduzir impactos econômicos da doença e fortalecer a sustentabilidade da cadeia citrícola brasileira.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Presidente Prudente

CERTIFICADO DE APROVAÇÃO

Título: DETECTION OF HUANGLONGBING (HLB) IN ORANGE TREES USING HYPERSPECTRAL SIGNATURES AND MACHINE LEARNING MODELS

AUTOR: MAURO MORATA BORTOLOTO JUNIOR
ORIENTADORA: ANA PAULA MARQUES RAMOS
COORIENTADOR: LUCAS PRADO OSCO

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

Profa. Dra. ANA PAULA MARQUES RAMOS (Participaçao Presencial)
Departamento de Cartografia / UNESP / Câmpus de Presidente Prudente - FCT

Profa. Dra. MARIA DE LOURDES BUENO TRINDADE GALO (Participaçao Presencial)
Departamento de Cartografia / UNESP / Câmpus de Presidente Prudente - FCT

Prof. Dr. JOSÉ MARCATO JUNIOR (Participaçao Virtual)
Fundação Universidade Federal de Mato Grosso do Sul / Universidade Federal de Mato Grosso do Sul

Presidente Prudente, 28 de abril de 2026.

À minha esposa Herriet, meu filho Bernardo e meus pais Mauro e Elenice.

ACKNOWLEDGMENTS

The completion of this dissertation was made possible thanks to the support, encouragement, and collaboration of several institutions and individuals, to whom I express my sincere gratitude.

I would like to thank the Graduate Program in Cartographic Sciences (PPGCC) and São Paulo State University (UNESP) for the opportunity for academic training and for the infrastructure provided throughout this journey.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

I am grateful to the Coordination for the Improvement of Higher Education Personnel (CAPES) for the financial support granted, which was essential for the development of this research.

I would also like to thank the Brazilian Agricultural Research Corporation (Embrapa Instrumentation), especially Dr. Lúcio de Castro Jorge, for providing the dataset and for the technical and scientific support that made this study possible.

My sincere gratitude goes to my advisor, Prof. Dr. Ana Paula Marques Ramos, for her steadfast guidance, constant dedication, confidence in my work, and the knowledge she shared throughout my master's studies. Her academic excellence and commitment to science were fundamental to my development.

I also thank my co-advisor, Prof. Dr. Lucas Prado Osco, for generously sharing his expertise and knowledge with such willingness and dedication.

I extend my gratitude to the professors of the Graduate Program in Cartographic Sciences for the knowledge they shared and for their contribution to my scientific education.

I am thankful to the members of the examination committee, Prof. Dr. Maria de Lourdes Bueno Trindade Galo and Prof. Dr. Danilo Roberto Pereira, for their willingness to evaluate this work and for their valuable suggestions, critiques, and contributions, which certainly enriched this dissertation.

My thanks also go to my colleagues and friends at PPGCC for their companionship, exchange of experiences, and academic discussions that significantly contributed to the maturation of this research.

To my family, I offer my deepest gratitude for their unconditional support, constant encouragement, and understanding during the moments of absence dedicated to the completion of this work.

Finally, I would like to thank everyone who, directly or indirectly, contributed to the accomplishment of this dissertation. My most sincere thanks to all of you.

ABSTRACT

Huanglongbing (HLB) is one of the most severe diseases affecting citrus production, causing physiological alterations that impact leaf spectral response and compromise plant productivity. In this context, the objective of this study was to discriminate healthy and HLB-infected citrus plants using hyperspectral foliar data, with emphasis on identifying disease-sensitive spectral regions and evaluating the robustness of classification models under real field conditions. A total of 1,912 leaf samples from 89 plants were analyzed, with spectral measurements acquired in the 350–2,500 nm range. The data were subjected to spectral preprocessing, dimensionality reduction using Principal Component Analysis (PCA), and classification using supervised machine learning algorithms. Model robustness was assessed through external validation using spatially independent orchard blocks, ensuring a more realistic evaluation of generalization under field conditions. The results demonstrated a high discrimination capacity between healthy and infected plants, with performance exceeding 94% in the main evaluation metrics at the plant level and greater stability compared to the leaf scale. The most relevant spectral regions were concentrated in the red, red-edge, near-infrared, and shortwave infrared domains, consistent with changes in pigments, mesophyll structure, and leaf water status. The proposed approach remained robust when evaluated using spatially independent orchard blocks, highlighting the potential of hyperspectral data and machine learning for scalable HLB monitoring under field conditions.

keywords: hyperspectral remote sensing, Greening, precision agriculture, machine learning.

RESUMO

O Huanglongbing (HLB) é uma das doenças mais severas que afetam a citricultura, causando alterações fisiológicas que impactam a resposta espectral foliar e comprometem a produtividade das plantas. Nesse contexto, o objetivo deste estudo foi discriminar plantas cítricas sadias e infectadas por HLB utilizando dados foliares hiperespectrais, com ênfase na identificação de regiões espectrais sensíveis à doença e na avaliação da robustez de modelos de classificação sob condições reais de campo. Foram analisadas 1.912 amostras foliares provenientes de 89 plantas, com medições espectrais adquiridas na faixa de 350 a 2.500 nm. Os dados foram submetidos a etapas de pré-processamento espectral, redução de dimensionalidade por meio da Análise de Componentes Principais (PCA) e classificação utilizando algoritmos supervisionados de aprendizado de máquina. A robustez dos modelos foi avaliada por meio de validação externa utilizando blocos espacialmente independentes do pomar, proporcionando uma avaliação mais realista da capacidade de generalização em condições de campo.

Os resultados demonstraram elevada capacidade de discriminação entre plantas sadias e infectadas, com desempenho superior a 94% nas principais métricas de avaliação em nível de planta e maior estabilidade quando comparado à escala foliar. As regiões espectrais mais relevantes concentraram-se nas faixas do vermelho, red-edge, infravermelho próximo e infravermelho de ondas curtas, consistentes com alterações em pigmentos, estrutura do mesófilo e estado hídrico foliar. A abordagem proposta manteve-se robusta quando avaliada em blocos espacialmente independentes do pomar, evidenciando o potencial dos dados hiperespectrais e do aprendizado de máquina para o monitoramento escalável do HLB em condições reais de campo.

Palavras-chave**:** sensoriamento remoto hiperespectral; greening; agricultura de precisão; aprendizado de máquina.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

HLB – Huanglongbing

PCR – Polymerase Chain Reaction

CLas – *Candidatus Liberibacter asiaticus*

CLaf – *Candidatus Liberibacter africanus*

CLam – *Candidatus Liberibacter americanus*

ELISA – *Enzyme-Linked Immunosorbent Assay*

SWIR – Shortwave Infrared

NIR – Near Infrared

PCA – Principal Component Annalisys

LR – Logistic Regression

SVM – Support Vector Machines

RF – Random Forest

DT – Decision Tree

GB – Gradient Boosting

NB – Naïve Bayes

SMOTE – Synthetic Minority Over-sampling Technique

TPE – Tree-structured Parzen Estimator

EMBRAPA – Brazilian Agricultural Research Corporation

FUNDECITRUS – Fund for Citrus Protection

WOS – Web of Science

ANN – Artificial Neural Networks

kNN – k-Nearest Neighbors

PFD – Physiological Fruit Dropping

ML – Machine learning

DL – Deep learning

CLT – Central Limit Theorem

UAV – Unmanned Aerial Vehicle

TP – True Positive

FP – False Positive

TN – True Negative

FN – False Negative

1. PROBLEM STATEMENT

1.1 Introduction

Brazil holds a prominent position in global citrus production, ranking among the world's largest orange producers and as the leading exporter of processed orange juice, a sector of major economic importance to the national agribusiness. According to recent estimates from the Fund for Citrus Protection (FUNDECITRUS), Brazilian orange production for the 2025 season was estimated at 292.6 million 40.8-kg boxes, corresponding to approximately 13 million tons. This figure is lower than previous projections due to adverse climatic conditions, especially drought, and the impacts caused by Huanglongbing (HLB). Production is concentrated mainly in the state of São Paulo, the country's principal citrus-growing region, consolidating Brazil as a global reference in the citrus production chain (FUNDECITRUS, 2025; IBGE, 2022).

Despite this favorable scenario, citrus production is constantly affected by factors that compromise orchard productivity and longevity, among which phytosanitary diseases stand out. In this context, Huanglongbing is considered the most severe disease affecting citrus crops, being responsible for significant production losses in several producing regions. It is a systemic disease associated with bacteria of the genus *Candidatus Liberibacter*, whose main management strategy consists of eradicating infected plants to contain disease spread within orchards (BOVÉ, 2006; GOTTWALD, 2010).

One of the main challenges in HLB management is the difficulty of accurately identifying infected plants in the field. The disease's visual symptoms are often varied and can resemble those caused by nutritional deficiencies or other physiological stresses. Furthermore, the irregular distribution of the pathogen within the plant compromises the reliability of conventional diagnostic methods, since healthy and infected leaves may coexist in the same individual plant (GOTTWALD et al., 2007; BASSANEZI et al., 2011).

Conventional diagnostic methods, such as visual inspection and molecular analysis by Polymerase Chain Reaction (PCR), present important limitations. Visual inspection is subjective and ineffective in detecting asymptomatic infections, while PCR, although highly sensitive and specific, is costly and has low operational feasibility when applied on a large scale (LI et al., 2006). Thus, the need for laboratory analysis of many samples, combined with the spatial variability of the disease, makes phytosanitary monitoring expensive and difficult to implement at orchard scale.

In this context, remote sensing has emerged as a promising alternative for non-destructive assessment of plant phytosanitary status. Vegetation spectral reflectance is directly related to leaf biophysical and biochemical properties, such as pigment content, mesophyll structure, and water content (CURRAN, 1989; ASNER, 1998). Physiological alterations induced by HLB affect these parameters, producing detectable changes in foliar spectral signatures.

Among available technologies, hyperspectral remote sensing stands out because it allows continuous acquisition of hundreds of narrow spectral bands, enabling detection of subtle variations distributed throughout the electromagnetic spectrum. This characteristic makes hyperspectral data particularly suitable for discriminating between different phytosanitary conditions, especially in situations where physiological changes cannot be adequately captured by conventional multispectral sensors (THENKABAIL; LYON; HUETE, 2012; MAHLEIN et al., 2013).

Several studies have demonstrated the potential of remote sensing in plant disease detection, highlighting its ability to identify spectral changes before visual symptoms become apparent (MAHLEIN et al., 2013; ZHOU et al., 2015). However, interpretation of these spectral variations is not trivial, since disease-related signals often manifest subtly and are distributed across the electromagnetic spectrum, making their detection difficult through conventional univariate statistical methods.

In this scenario, machine learning techniques have been widely employed as tools for analyzing high-dimensional spectral data. Unlike traditional statistical approaches, supervised learning algorithms can explore complex and nonlinear multivariate relationships among spectral variables, enabling identification of discriminative patterns associated with disease presence (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; BELGIU; DRĂGUŢ, 2016). Recent studies have demonstrated the successful use of these methods in classifying phytosanitary conditions in different agricultural crops based on remote sensing data (ASHOURLOO et al., 2016; BECK et al., 2015; DELALIEUX et al., 2007).

Several studies have also demonstrated the potential of remote sensing applied to plant disease detection in citrus crops using terrestrial, aerial, and orbital platforms (SANKARAN et al., 2010; ZHOU et al., 2015). Nevertheless, despite recent advances, there is still no consensus regarding the most relevant spectral regions for discriminating against healthy and HLB-infected plants, and considerable variability is observed among results depending on the sensor used, the scale of analysis, and the methodologies employed.

Interpretation of these spectral data poses an additional challenge. Hyperspectral spectra present high dimensionality, strong collinearity between adjacent wavelengths, and often nonlinear relationships between reflectance and plant physiological status. Under these conditions, conventional statistical methods become limited in capturing complex patterns distributed across multiple bands simultaneously (BELLMAN, 1961; JOLLIFFE; CADIMA, 2016).

In this context, supervised machine learning techniques assume a central role. Unlike traditional univariate approaches, these algorithms can model complex multivariate relationships, exploring nonlinear interactions among spectral variables, and identifying discriminative patterns associated with disease presence (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; BELGIU; DRĂGUŢ, 2016). Recent studies have shown promising results with algorithms such as logistic regression, Random Forest, Support Vector Machines, and Gradient Boosting in the classification of phytosanitary conditions based on spectral data (ASHOURLOO et al., 2016; BECK et al., 2015; DELALIEUX et al., 2007).

Given the physiological complexity of Huanglongbing and the high dimensionality of hyperspectral data, important challenges remain regarding the identification of disease-related spectral signatures and their translation into robust classification models. In particular, the spectral regions most strongly associated with HLB-induced physiological alterations remain insufficiently understood, and the ability of machine learning algorithms to generalize under spatially independent field conditions requires further investigation.

1.2 General Objective

To characterize how physiological changes associated with Huanglongbing infection in orange trees are manifested in hyperspectral data, as well as to evaluate to what extent these changes can be detected by supervised machine learning algorithms under representative field conditions.

1.3 Specific Objectives

- To map the main approaches integrating remote sensing and artificial intelligence in the detection of HLB in orange trees;

- To determine variations in the spectral signatures of orange trees infected with HLB at both leaf and plant levels;
- To evaluate the potential of supervised machine learning algorithms to classify HLB-infected and non-infected orange trees at both leaf and plant levels;
- To identify spectral regions that discriminate physiological changes associated with HLB infection in orange trees;

1.4 Thesis Structure

This dissertation is organized into six chapters that progressively develop the theoretical, methodological, and analytical foundations required to investigate the potential of hyperspectral remote sensing and machine learning for Huanglongbing (HLB) detection in citrus plants. The structure was designed to establish a logical connection between the current state of knowledge, the experimental procedures adopted, and the scientific evidence generated throughout the study.

Chapter 1 – Problem Statement introduces the research context and presents the scientific and practical motivations underlying this work. The chapter discusses the importance of Brazilian citriculture, the economic and phytosanitary impacts of Huanglongbing, the limitations of conventional disease diagnosis methods, and the opportunities provided by hyperspectral remote sensing and machine learning technologies. The research objectives, hypotheses, and overall organization of the dissertation are also presented.

Chapter 2 – Theoretical Background establishes the conceptual foundations necessary to support the research. Initially, the chapter addresses the etiology, epidemiology, physiological effects, and diagnostic challenges associated with HLB. Subsequently, the principles of hyperspectral remote sensing applied to vegetation analysis are discussed, emphasizing the interaction between electromagnetic radiation and plant tissues. Finally, the chapter presents the fundamentals of machine learning, highlighting its application to high-dimensional spectral datasets and its potential for disease detection and classification.

Chapter 3 – State of the Art: A Review Analysis of Remote Sensing and Machine Learning Applications in Huanglongbing Detection presents a systematic literature review and scientometric analysis aimed at characterizing the current scientific landscape of HLB detection through remote sensing techniques and artificial intelligence methods. This chapter identifies the most frequently used sensors, platforms, preprocessing strategies, and machine learning approaches, while also highlighting existing knowledge gaps and future research opportunities. The findings obtained in this review provide the scientific basis for the methodological decisions adopted in the experimental phases of this dissertation.

Chapter 4 – Materials and Methods describes the complete methodological framework adopted throughout the study. The chapter presents the study area, experimental design, spectral database, procedures for hyperspectral data acquisition, preprocessing techniques, statistical analyses, machine learning workflow, validation strategies, and performance metrics. Furthermore, this chapter introduces the two complementary experimental investigations conducted in this research. The first experiment focuses on identifying spectral regions associated with physiological alterations induced by HLB through statistical analyses performed at both leaf and plant levels. The second experiment investigates the ability of supervised machine learning algorithms to discriminate healthy and infected plants under spatially independent validation scenarios, assessing the robustness and generalization capability of predictive models under realistic field conditions.

Chapter 5 – Discussion integrates and interprets the results obtained throughout the dissertation. The findings from the literature review, statistical analyses, and machine learning experiments are jointly discussed to establish connections between disease physiology, spectral responses, and predictive modeling performance. Particular emphasis is given to the physiological interpretation of discriminative spectral regions, the advantages and limitations of different analytical scales, the influence of spatially independent validation strategies, and the implications of the results for operational HLB monitoring. This chapter also discusses the potential application of the identified spectral regions in the development of future multispectral and hyperspectral sensing systems for citrus disease detection.

Chapter 6 – Final Considerations summarizes the main scientific contributions of the dissertation and revisits the research objectives in light of the results obtained. The chapter highlights the relevance of integrating hyperspectral spectroscopy and machine learning for HLB detection, discusses the methodological and practical limitations of the study, and presents recommendations for future investigations, including the evaluation of additional environmental conditions, temporal monitoring approaches, and the development of operational remote sensing solutions for large-scale citrus disease surveillance.

Finally, the References section compiles all bibliographic sources cited throughout the dissertation, following the academic standards established by the Graduate Program in Cartographic Sciences.

2. THEORETICAL BACKGROUND

2.1 Huanglongbing: Pathophysiology, Symptoms, and Diagnostic Limitations

From an etiological perspective, HLB is caused by bacteria of the genus *Candidatus Liberibacter*, with the main species being *Candidatus Liberibacter asiaticus* (CLas),

6. FINAL CONSIDERATIONS

This dissertation advances the scientific understanding of Huanglongbing detection by demonstrating that physiological changes induced by the disease are consistently manifested in hyperspectral foliar reflectance and can be reliably detected through supervised machine learning methods.

The research contributes to three complementary dimensions: first, by consolidating the international methodological panorama regarding HLB detection through remote sensing; second, by identifying statistically robust spectral regions directly associated with disease-induced physiological alterations; and third, by validating predictive classification models under spatially independent field conditions, ensuring methodological robustness and practical relevance.

The results confirm that the most informative spectral domains are concentrated in the red, red-edge, near-infrared, and shortwave infrared regions, reinforcing their central importance in future sensor design and operational disease monitoring systems. Among the evaluated models, logistic regression proved especially effective, combining high predictive accuracy with interpretability and computational simplicity.

The study also demonstrates that plant-level aggregation offers superior analytical stability compared with leaf-level approaches, representing a more realistic scale for field implementation. This finding has direct implications for future deployment of automated monitoring systems in citrus orchards.

Despite these advances, some limitations remain. The experimental dataset is restricted to a single geographic area and one acquisition period, which limits broader environmental generalization. In addition, the absence of multi-season temporal validation prevents assessment of seasonal spectral variability effects on model transferability.

Future studies should therefore prioritize expanding sampling across different citrus-producing regions, incorporating multiple cultivars and environmental conditions, and evaluating temporal consistency across successive crop seasons. The integration of UAV-mounted multispectral sensors and transfer learning approaches also represents a promising direction for scaling the methodology toward operational phytosanitary surveillance. In conclusion, the integration of hyperspectral remote sensing and supervised machine learning constitutes a robust, interpretable, and scalable scientific approach for

HLB monitoring in citrus crops, offering strong potential to support precision agriculture strategies and improve phytosanitary management in Brazilian citriculture.

REFERENCES

BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, v. 88, n. 1, p. 7–37, 2006. DOI: https://doi.org/10.4454/jpp.v88i1.828.

GOTTWALD, T. R. Current epidemiological understanding of citrus Huanglongbing. *Annual Review of Phytopathology*, v. 48, p. 119–139, 2010. DOI: https://doi.org/10.1146/annurev-phyto-080508-081844.

COLETTA-FILHO, H. D. et al. First report of the causal agent of Huanglongbing (“Candidatus Liberibacter asiaticus”) in Brazil. *Plant Disease*, v. 88, n. 12, p. 1382–1382, 2004.

TEIXEIRA, D. C. et al. Candidatus Liberibacter americanus associated with citrus Huanglongbing in São Paulo State, Brazil. *International Journal of Systematic and Evolutionary Microbiology*, v. 55, n. 5, p. 1857–1862, 2005.

LOPES, S. A. et al. Detection and quantification of Candidatus Liberibacter americanus by real-time PCR. *Plant Disease*, v. 93, n. 9, p. 1045–1050, 2009.

WANG, N.; TRIVEDI, P. Citrus Huanglongbing: a newly relevant disease presents unprecedented challenges. *Phytopathology*, v. 103, n. 7, p. 652–665, 2013.

FUNDECITRUS. *Levantamento de incidência de Huanglongbing (HLB) no cinturão citrícola 2024*. Araraquara: Fundecitrus, 2024.

FUNDECITRUS. *Estimativa da safra de laranja 2024/25 do cinturão citrícola de São Paulo e Triângulo Mineiro*. Araraquara: Fundecitrus, 2024.

CASTRO, Carolina; GILES, F. *Update to Citrus Annual Report for Brazil*. Brasília: USDA Foreign Agricultural Service, 2025. Report No. BR2025-0002.

ABDULRIDHA, J.; BATUMAN, O.; AMPATZIDIS, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing*, v. 11, n. 11, p. 1373, 2019. DOI: https://doi.org/10.3390/rs11111373.

GOETZ, A. F. H.; VANE, G.; SOLOMON, J. E.; ROCK, B. N. Imaging spectrometry for Earth remote sensing. *Science*, v. 228, n. 4704, p. 1147–1153, 1985.

JENSEN, J. R. *Remote sensing of the environment: an earth resource perspective*. 2. ed. London: Pearson Education, 2014.

THENKABAIL, P. S.; LYON, J. G.; HUETE, A. (eds.). *Hyperspectral remote sensing of vegetation*. Boca Raton: CRC Press, 2012.

CURRAN, P. J. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, v. 30, n. 3, p. 271–278, 1989.

GATES, D. M.; KEEGAN, H. J.; SCHLETER, J. C.; WEIDNER, V. R. Spectral properties of plants. *Applied Optics*, v. 4, n. 1, p. 11–20, 1965.

KNIPLING, E. B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, v. 1, p. 155–159, 1970.

HORLER, D. N. H.; DOCKRAY, M.; BARBER, J. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, v. 4, n. 2, p. 273–288, 1983.

TUCKER, C. J. Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment*, v. 10, n. 1, p. 23–32, 1980.

VOGELMANN, T. C. Plant tissue optics. *Annual Review of Plant Physiology and Plant Molecular Biology*, v. 44, p. 231–251, 1993.

LICHTENTHALER, H. K. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods in Enzymology*, v. 148, p. 350–382, 1987.

MAHLEIN, A.-K. Plant disease detection by imaging sensors: parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, v. 100, n. 2, p. 241–251, 2016. DOI: https://doi.org/10.1094/PDIS-03-15-0340-FE.

MAHLEIN, A.-K. et al. Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, v. 128, p. 21–30, 2013.

SANKARAN, S.; MISHRA, A.; EHSANI, R.; DAVIS, C. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, v. 72, n. 1, p. 1–13, 2010. DOI: https://doi.org/10.1016/j.compag.2010.02.007.

USHA, K.; SINGH, B. Potential applications of remote sensing in horticulture: a review. *Scientia Horticulturae*, v. 153, p. 71–83, 2013. DOI: https://doi.org/10.1016/j.scienta.2013.01.008.

ZHANG, M.; QIN, Z.; LIU, X.; USTIN, S. L. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, v. 13, n. 4, p. 601–607, 2011.

ZHOU, J.; PENG, Y.; ZHANG, Y.; DU, X. Hyperspectral reflectance data mining for detecting HLB citrus trees. *Remote Sensing*, v. 7, n. 6, p. 7482–7504, 2015.

ZHOU, R. et al. Spectral reflectance characteristics of citrus leaves infected by Huanglongbing. *Computers and Electronics in Agriculture*, v. 112, p. 98–106, 2015.

LI, W.; DUAN, Y.; LI, D.; LEE, W. S.; YANG, C. Spectral characteristics and detection of citrus Huanglongbing infected leaves. *Transactions of the ASABE*, v. 55, p. 711–720, 2012.

SANKARAN, S.; MISHRA, A.; MAJA, J. M.; EHSANI, R. Visible–near infrared spectroscopy for detection of Huanglongbing in citrus orchards. *Plant Disease*, v. 97, p. 1619–1626, 2013.

LI, L. et al. Detection of citrus Huanglongbing using visible and near infrared spectroscopy. *Journal of Food, Agriculture & Environment*, v. 12, n. 2, p. 446–451, 2014.

TANG, Y. et al. Early detection of Huanglongbing in citrus using hyperspectral images and machine learning. *Biosystems Engineering*, v. 194, p. 95–107, 2020.

HUANG, W. et al. Identifying optimal spectral bands for detecting citrus Huanglongbing using hyperspectral data. *Computers and Electronics in Agriculture*, v. 194, p. 106747, 2022.

GUPTA, H. et al. Machine learning based detection of Huanglongbing using hyperspectral imagery. *Computers and Electronics in Agriculture*, v. 169, p. 105240, 2020.

MITCHELL, T. M. *Machine learning*. New York: McGraw-Hill, 1997.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. New York: Springer, 2009.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An introduction to statistical learning: with applications in R*. New York: Springer, 2013.

BISHOP, C. M. *Pattern recognition and machine learning*. New York: Springer, 2006.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. *Classification and regression trees*. New York: Chapman & Hall, 1984.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, p. 273–297, 1995.

VAPNIK, V. N. *The nature of statistical learning theory*. New York: Springer, 1995.

CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016. p. 785–794.

KE, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017. p. 3146–3154.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, v. 374, n. 2065, 2016.

JOLLIFFE, I. T. *Principal component analysis*. 2. ed. New York: Springer, 2002.

BELLMAN, R. E. *Adaptive control processes: a guided tour*. Princeton: Princeton University Press, 1961.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, v. 45, n. 4, p. 427–437, 2009.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.

POWERS, D. M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37–63, 2011.

HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, p. 1263–1284, 2009.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann, 1995. p. 1137–1143.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. 3. ed. Hoboken: Wiley, 2013.

LUNDERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*. 2017. p. 4765–4774.

AKIBA, T. et al. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2019. p. 2623–2631.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

RINNAN, Å.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, v. 28, n. 10, p. 1201–1222, 2009.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, v. 43, n. 5, p. 772–777, 1989.

MARTENS, H.; NAES, T. *Multivariate calibration*. Chichester: Wiley, 1989.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, v. 36, n. 8, p. 1627–1639, 1964.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, v. 39, n. 3, p. 491–500, 1985.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

WORKMAN, J.; WEYER, L. *Practical guide to interpretive near-infrared spectroscopy*. Boca Raton: CRC Press, 2008.

BURNS, D. A.; CIURCZAK, E. W. (eds.). *Handbook of near-infrared analysis*. 3. ed. Boca Raton: CRC Press, 2007.

THENKABAIL, P. S.; LYON, J. G.; HUETE, A. (eds.). *Hyperspectral remote sensing of vegetation*. Boca Raton: CRC Press, 2012.

CURRAN, P. J. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, v. 30, n. 3, p. 271–278, 1989.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, v. 374, n. 2065, 2016.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. *A user-friendly guide to multivariate calibration and classification*. Chichester: NIR Publications, 2002.

ESBENSEN, K. H.; GUYOT, D.; WESTAD, F.; HOUMØLLER, L. P. *Multivariate data analysis: in practice*. 5. ed. Oslo: CAMO Process AS, 2002.

BELLMAN, R. E. *Adaptive control processes: a guided tour*. Princeton: Princeton University Press, 1961.

MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. 7. ed. Hoboken: Wiley, 2018.

DEVORE, J. L. *Probability and statistics for engineering and the sciences*. 9. ed. Boston: Cengage Learning, 2016.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3–4, p. 591–611, 1965.

RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, v. 2, n. 1, p. 21–33, 2011.

STUDENT. The probable error of a mean. *Biometrika*, v. 6, n. 1, p. 1–25, 1908.

RUXTON, G. D. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, v. 17, n. 4, p. 688–690, 2006.

WELCH, B. L. The generalization of Student's problem when several different population variances are involved. *Biometrika*, v. 34, n. 1–2, p. 28–35, 1947.

MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, v. 18, n. 1, p. 50–60, 1947.

FISHER, R. A. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.

MONTGOMERY, D. C. *Design and analysis of experiments*. 9. ed. Hoboken: Wiley, 2017.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. Upper Saddle River: Pearson, 2007.

COHEN, J. *Statistical power analysis for the behavioral sciences*. 2. ed. Hillsdale: Lawrence Erlbaum Associates, 1988.

PEARSON, K. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, v. 58, p. 240–242, 1895.

SPEARMAN, C. The proof and measurement of association between two things. *American Journal of Psychology*, v. 15, n. 1, p. 72–101, 1904.

MAHALANOBIS, P. C. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, v. 2, n. 1, p. 49–55, 1936.

HOTELLING, H. The generalization of Student's ratio. *Annals of Mathematical Statistics*, v. 2, n. 3, p. 360–378, 1931.

EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993.

KOLMOGOROV, A. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, v. 4, p. 83–91, 1933.

SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, v. 19, n. 2, p. 279–281, 1948.

ZAR, J. H. *Biostatistical analysis*. 5. ed. Upper Saddle River: Pearson, 2010.

GOMEZ, K. A.; GOMEZ, A. A. *Statistical procedures for agricultural research*. 2. ed. New York: Wiley, 1984.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. 3. ed. Hoboken: Wiley, 2013.

MENARD, S. *Applied logistic regression analysis*. 2. ed. Thousand Oaks: Sage Publications, 2002.

AGRESTI, A. *An introduction to categorical data analysis*. 3. ed. Hoboken: Wiley, 2018.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

MOLNAR, C. *Interpretable machine learning*. 2. ed. [S.l.]: Lulu.com, 2022.

LEGENDRE, P. Spatial autocorrelation: trouble or new paradigm? *Ecology*, v. 74, n. 6, p. 1659–1673, 1993.

CLIFF, A. D.; ORD, J. K. *Spatial autocorrelation*. London: Pion, 1973.

ROBERTS, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, v. 40, n. 8, p. 913–929, 2017.

VALAVI, R. et al. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, v. 10, n. 2, p. 225–232, 2019.

WENGER, S. J.; OLDEN, J. D. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, v. 3, n. 2, p. 260–267, 2012.

TUIA, D. et al. Domain adaptation in remote sensing data analysis. *IEEE Geoscience and Remote Sensing Magazine*, v. 4, n. 1, p. 41–57, 2016.

MAXWELL, A. E.; WARNER, T. A.; FANG, F. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, v. 39, n. 9, p. 2784–2817, 2018.

VAPNIK, V. N. *Statistical learning theory*. New York: Wiley, 1998.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

KAUFMAN, S.; ROSSET, S.; PERLICH, C.; STROBL, E. Leakage in data mining: formulation, detection, and avoidance. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2011. p. 556–563.

MAXWELL, A. E.; STRONG, A.; WARNER, T. A. Training and evaluating machine learning classifiers for remote sensing applications. *Remote Sensing*, v. 13, n. 3, p. 526, 2021.