FILIPE VIEIRA DA SILVA

DISCRIMINATION OF COFFEE DISEASES AND PESTS USING MULTISPECTRAL IMAGING

Dissertation presented to the Graduate Program in Agriculture and Geospatial Information at the Federal University of Uberlândia, *Campus* Monte Carmelo, as a partial requirement for obtaining the title of 'Master in Agriculture and Geospatial Information '.

Advisor
Prof. PhD. George Deroco Martins.

MONTE CARMELO
MINAS GERAIS – BRAZIL 2026

FILIPE VIEIRA DA SILVA

DISCRIMINATION OF COFFEE DISEASES AND PESTS USING MULTISPECTRAL IMAGING

Dissertation presented to the Graduate Program in Agriculture and Geospatial Information at the Federal University of Uberlândia, *Campus* Monte Carmelo, as a partial requirement for obtaining the title of 'Master in Agriculture and Geospatial Information '.

APPROVED on February 23, 2026

Examination Board

Prof. PhD. George Deroco Martins - Institute of Geography, Geosciences and Collective Health (UFU)

Prof. PhD. Rodrigo Bezerra de Araújo Gallis - Institute of Geography, Geosciences, and Collective Health (UFU)

Prof. PhD. Charles Cardoso Santana - Institute of Agricultural Technology of Pitangui (EPAMIG)

Prof. PhD. George Deroco Martins Institute of Geography, Geosciences and Collective Health (UFU) (Advisor)

MONTE CARMELO

MINAS GERAIS – BRAZIL

2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU com dados informados pelo(a) próprio(a) autor(a).

S586
2026

Silva, Filipe Vieira da, 2000-
Discrimination of Coffee Diseases and Pests Using Multispectral Images [recurso eletrônico] / Filipe Vieira da Silva. - 2026.

Orientador: George Deroco Martins.
Coorientadora: Gleice Aparecida de Assis.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Agricultura e Informações Geoespaciais.
Modo de acesso: Internet.
DOI http://doi.org/10.14393/ufu.di.2026.97
Inclui bibliografia.
Inclui ilustrações.

1. Agronomia. I. Martins, George Deroco,1987-, (Orient.). II. Assis, Gleice Aparecida de,1986-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Agricultura e Informações Geoespaciais. IV. Título.

CDU: 631

Bibliotecários responsáveis pela estrutura de acordo com o AACR2: Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Agricultura e Informações Geoespaciais
Rodovia LMG 746, Km 01, s/nº, Bloco 1AMC, Sala 1A202 , Monte Carmelo-MG, CEP 38.500-000
Telefone: (34) 3810-1033 - ppgaig@iciag.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Agricultura e Informações Geoespaciais				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	23/02/2026	Hora de início:	9:00h	Hora de encerramento:	12:00h
Matrícula do Discente:	32412AIG007				
Nome do Discente:	Filipe Vieira da Silva				
Título do Trabalho:	Detecção E Discriminação de Doenças E Pragas do Cafeeiro por Meio de Imagens Multiespectrais				
Área de concentração:	Informações geoespaciais e tecnologias aplicadas à produção agrícola				
Linha de pesquisa:	Desenvolvimento e aplicações de métodos em informações geoespaciais				
Projeto de Pesquisa de vinculação:	Mapeamento de doenças e variáveis agrícolas em culturas agrícolas a partir de dados de sensoriamento remoto				

Reuniu-se em sala virtual a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agricultura e Informações Geoespaciais, composta pelo prof. Dr. George Deroco Martins (Universidade Federal de Uberlândia-UFU), orientador do candidato, prof. Dr. Rodrigo Bezerra de Araújo Gallis (Universidade Federal de Uberlândia-UFU), e prof. Dr. Charles Cardoso Santana (Instituto Tecnológico de Agropecuária de Pitangui - EPAMIG).

Iniciando os trabalhos o presidente da mesa, Dr. George Deroco Martins, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação da discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **George Deroco Martins**, **Professor(a) do Magistério Superior**, em 25/02/2026, às 16:08, conforme horário oficial de Brasília, com fundamento no art. 6º,
§ 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Rodrigo Bezerra de Araujo Gallis**, **Professor(a) do Magistério Superior**, em 25/02/2026, às 20:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Charles Cardoso Santana**, **Usuário Externo**, em 26/02/2026, às 08:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7081672** e o código CRC **F0F0644F**.

Referência: Processo nº 23117.010636/2026-03

SEI nº 7081672

THANKS

First and foremost, I thank God for the strength, wisdom, and light that guided me through every challenge.

To my mother, Sirlene Maria da Silva, my eternal foundation and inexhaustible source of love, sacrifice, and encouragement. Thank you for believing in me even when I doubted myself and for being the safe harbor that always welcomed me.

To my family and friends, for your love, unconditional support, and companionship.

To the professors and staff for the knowledge they shared. Special thanks to my advisor, Prof. Dr. George Deroco Martins, for his patience, availability, and for providing guidance full of opportunities and a sense of ease.

To the Federal University of Uberlândia – Monte Carmelo campus, for the learning and research environment and facilities provided.

To the Coordination for the Improvement of Higher Education Personnel (CAPES), for the crucial financial support through the scholarship.

Finally, to my friends and research colleagues. Thank you for your camaraderie, assistance with every task, for the moments of relaxation, for the encouraging words in times of tension, and for making the journey lighter and more joyful.

I leave here my sincere thanks!

BIOGRAPHY

Filipe Vieira da Silva was born in Monte Carmelo, Minas Gerais, Brazil, 1997. He holds a degree as a Survey and Cartographic Engineer (2023) from the Federal University of Uberlândia (UFU), Minas Gerais, Brazil. During undergraduate studies, he developed scientific initiation projects and conducted work using geospatial data for agricultural mapping, such as yield estimation, plant classification, and plant health analysis. He is a Master's student in the Graduate Program in Agriculture and Geospatial Information, focusing on precision agriculture and remote sensing applied to multispectral imaging for pest and disease detection in coffee plants.

SUMÁRIO

ABSTRACT

Accurate identification of diseases and pests in coffee plants is crucial for targeted crop management and avoiding productivity losses. In this study, multispectral images were used to monitor a drip-irrigated coffee plantation of the cultivar Topázio MG-1190 planted in 2016 at the Federal University of Uberlândia – Monte Carmelo Campus, with the objective of discriminating symptoms caused by Leucoptera coffeella (coffee leaf miner), Cercospora coffeicola (Cercospora leaf spot), and Hemileia vastatrix (coffee leaf rust). The images, captured monthly from February to November 2024 using a DJI Phantom 4 drone, were analyzed using supervised machine learning classification models (Neural Networks and Random Forests) developed in Weka software. Concurrently, 4 leaf pairs per plant (middle third) were assessed to analyze the incidence of each disease/pest. The input data for the classifications consisted of the sensor's original spectral bands, derived vegetation indices, and the designated class per plant. Subsequently, the methodology was tested on a commercial crop in July 2025. The results demonstrated that the classification models achieved up to 85% accuracy and a Kappa index of 0.8, highlighting the potential use of multispectral images for the continuous monitoring of major coffee crop diseases and pests. Furthermore, the maps generated by the models allowed for the analysis of the spatial distribution of the studied phytosanitary problems, proving to be a tool to support localized management practices. Thus, the research showed that this tool can be an alternative for promoting sustainable coffee farming, contributing to more efficient disease and pest control.

Keywords: Coffee; Monitoring; Plant Diseases; Precision Agriculture; Remotely Piloted Aircraft; Remote Sensing.

RESUMO

A identificação precisa de doenças e pragas em cafeeiros é crucial para o manejo dirigido da lavoura e para evitar perdas de produtividade. Neste estudo, utilizaram-se imagens multiespectrais para monitorar um cafezal irrigado por gotejamento da cultivar Topázio MG-1190, implantado em 2016 na Universidade Federal de Uberlândia – Campus Monte Carmelo, com o objetivo de discriminar sintomas causados por Leucoptera coffeella (bicho-mineiro), Cercospora coffeicola (cercosporiose) e Hemileia vastatrix (ferrugem do cafeeiro). As imagens, capturadas mensalmente de fevereiro a novembro de 2024 com um drone DJI Phantom 4, foram analisadas por meio de modelos de classificação supervisionada de aprendizado de máquina (Redes Neurais e Random Forests) desenvolvidos no software Weka. Paralelamente, foram avaliados 4 pares de folhas por planta (terço médio) para analisar a incidência de cada doença/praga. Os dados de entrada para as classificações consistiram nas bandas espectrais originais do sensor, nos índices de vegetação derivados e na classe designada por planta. Posteriormente, a metodologia foi testada em uma lavoura comercial em julho de 2025. Os resultados demonstraram que os modelos de classificação atingiram até 85% de acurácia e índice Kappa de 0,8, evidenciando o potencial do uso de imagens multiespectrais para o monitoramento contínuo das principais doenças e pragas da cafeicultura. Além disso, os mapas gerados pelos modelos permitiram analisar a distribuição espacial dos problemas fitossanitários estudados, mostrando-se uma ferramenta de apoio a práticas de manejo localizado. Desse modo, a pesquisa evidenciou que essa ferramenta pode ser uma alternativa para a promoção da cafeicultura sustentável, contribuindo para o controle mais eficiente de doenças e pragas.

Palavras-chave: Café; Monitoramento; Doenças de Plantas; Agricultura de Precisão; Aeronave Remotamente Pilotada; Sensoriamento Remoto.

1. INTRODUCTION

The incidence of diseases and pests in coffee plants is one of the factors influencing productivity and beverage quality, making correct diagnosis necessary to ensure rational and sustainable management. However, physiological disorders and nutritional deficiencies can cause symptoms like those caused by disease and pest infestations, which complicate accurate diagnosis and the safe recommendation of phytosanitary products (Pozza, Carvalho e Chalfoun., 2010; Wan *et al*., 2022).

In this context, there are phytosanitary products intended for the exclusive control of a single disease or pest, as well as those that can treat multiple diseases simultaneously. Their incorrect use, whether due to dosage error, number of applications, tank mixture, or application technology, can lead to phytotoxicity, which favors the development of other diseases and pests (Losch *et al.*, 2020 Considering the control of Brown Eye Spot (BES) and Coffee Leaf Rust (CLR), the use of systemic and protectant fungicides, such as those based on triazole, strobilurin, among others, is typically recommended. Regarding the Coffee Leaf Miner (CLM), control is achieved using systemic insecticides applied via soil and foliar spraying with active ingredients such as thiamethoxam, abamectin, etc. When applied at high doses, these products can cause toxic effects (mainly anti-gibberellin action), resulting in curled leaves, changes in leaf color, defoliation, small fruits, and delayed maturation (Vieira Junior et al., 2020; Matiello et al., 2015). In this context, accurately diagnosing phytosanitary problems can guide product selection, as well as the appropriate level of control for application.

From this perspective, although monitoring coffee crops using aerial or satellite imagery is already a reality, significant limitations still exist for the accurate discrimination of diseases within the same crop. These include climatic conditions that limit data acquisition, the similar response of the plant to different phytosanitary problems, the influence of crop management practices on the spectral response of plants, and even the technology for data acquisition and processing (Ennouri et al., 2025).

These challenges can be observed in recent studies. For instance, Baht, Dominguez e Aljumaili (2025) demonstrated the potential use of imagery for detecting coffee leaf diseases. However, despite the high accuracy, it is important to highlight the large number of samples used (over 30,000 images), which makes the entire process time-consuming, lacks evaluations at the crop canopy level, and lacks practical field applications. In contrast, Mendonça (2022) managed to classify plants using aerial images, but with low accuracy (~47%). His contribution stands out in the spectral characterization of the confusion between classifying healthy plants and plants with Brown Eye Spot (BES), depicting the similar spectral behavior of these plants under different conditions.

Furthermore, recent studies have focused on the identification of diseases, disorders, or severity at the leaf level, as demonstrated in the works of Adelaja e Pranggono (20225), Silva et al. (2025) e Cajas et al. (2025). In addition, studies that considered the classification of entire crops presented the evaluation of a single phytosanitary problem, disregarding symptom overlap and the incidence of other diseases and pests, as observed in the research by Vilela et al. (2024), Da Silva et al. (2024) e Dos Santos et al. (2024). Thus, the overlap of visual and spectral symptoms among different diseases still represents a significant technical challenge for the practical application of remote sensing monitoring.

It is necessary to emphasize that, much like plant health, the incidence and severity of coffee diseases and pests are directly influenced by climate and the plant's phenological stage (Vieira Júnior et al., 2020). Additionally, Chedid, Cortez e Arcoverde (2024) demonstrated that the spectral response of crops changes throughout plant development, highlighting that vegetative vigor can influence image readings. This reinforces the need for research that considers seasonal variations and the phenological stages of coffee plants to identify the ideal time for image capture, aiming to provide more effective monitoring (Longchamps & Philpot, 2023; Diao, 2020). Moreover, Zhang et al. (2025) point out in their research that the use of remote sensing applied to the detection of plant pests and diseases is not widely explored, especially in coffee crops, which limits the advancement of accurate diagnostic capabilities and the practical implementation of targeted management.

Without a clear understanding of the spectral behavior associated with each type of disease and pest under different phenological stages, it becomes challenging to develop algorithms or models capable of distinguishing them automatically and accurately. Therefore, progress in using imagery for monitoring coffee diseases largely depends on expanding knowledge about the spectral signatures of these diseases and a more detailed investigation into the optimal timing for data collection (Mahlein, 2016; Orlando et al., 2021; Zhang et al., 2025).

Given this, the hypothesis is that different diseases and pests affecting coffee plants induce specific biophysical changes, generating distinct spectral signatures detectable through multispectral images, especially when combined with vegetation indices (VIs) and machine learning algorithms. In this context, this study aimed to evaluate the effectiveness of multi-scale remote sensing, combined with spectral indices and supervised classification models, in differentiating healthy coffee plants from those infested by the Coffee Leaf Miner (L. coffeella, CLM), Brown Eye Spot (C. coffeicola, BES), and Coffee Leaf Rust (H. vastatrix, CLR), under different phenological stages, providing a phytosanitary monitoring tool to support localized and sustainable crop management practices.

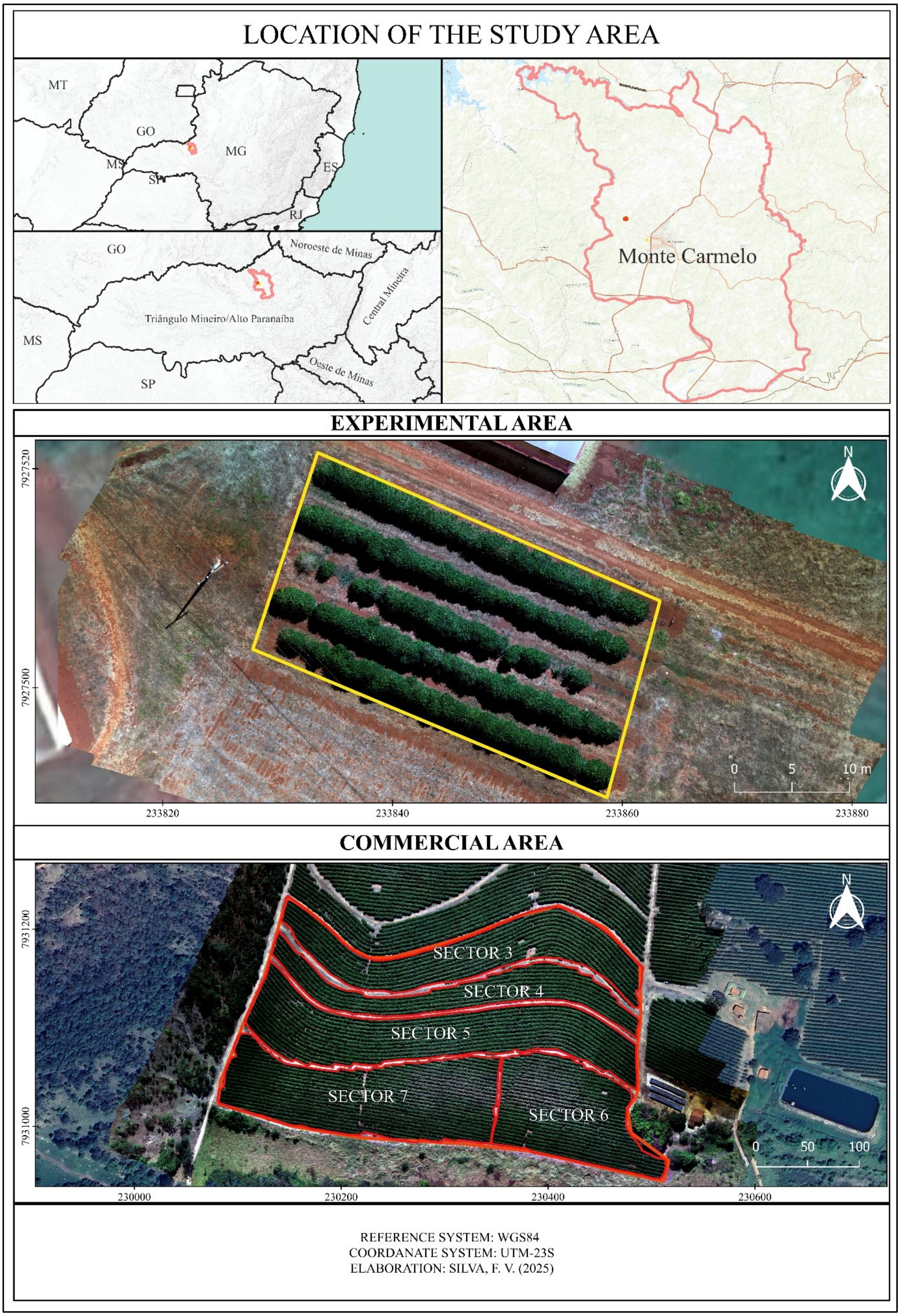
2. STUDY AREA CHARACTERIZATION

In the study conducted at the Federal University of Uberlândia – Araras Unit, in Monte Carmelo, MG, Brazil, the incidence of CLM, BES, and CLR was evaluated in an experimental area containing 185 plants of the Topázio MG-1190 cultivar, planted in 2016 with a spacing of 3.5 m between rows x 0.6 m between plants. The plot is equipped with drip irrigation, with an average flow rate of 1.6 L/h and 60 cm between emitters. Subsequently, a test was conducted in a commercial crop at the Juliana Farm in Monte Carmelo, covering 7.82 ha and divided into 5 sectors. This crop contained the same cultivar with a spacing of 3.8 m between rows x 0.6 m between plants. However, the plants were established in 2009 (sectors 3, 4, and 5) and 2012 (sectors 6 and 7). The aim was to assess monitoring capability on a larger scale and in a crop with management practices closer to field reality. The commercial crop also utilizes drip irrigation, with an average flow rate of 2.2 L/h and 75 cm between emitters.

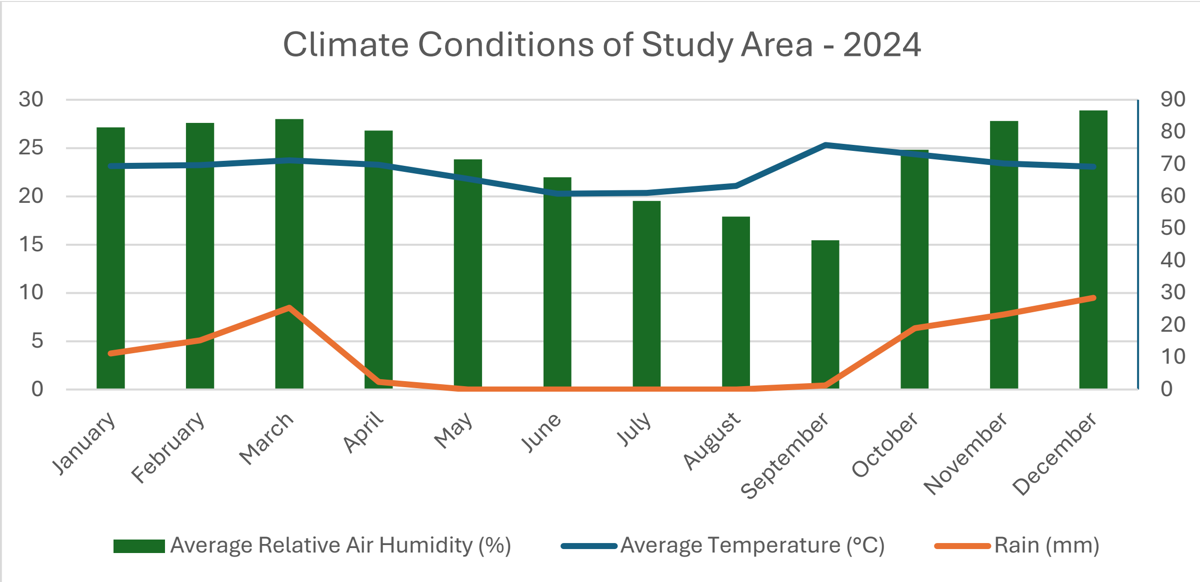
Regarding crop management, in the experimental plot during the research period, it is important to note the following applications: Priori Xtra in March 2024 for the control of BES and CLR, Fastac in April 2024 for CLM control, Auge for BES, and Vertimec for the control of CLM, coffee berry borer, and mites in October 2024. Additionally, herbicide applications, weeding, removal of volunteer plants, fertilization, and liming were carried out. Meanwhile, in the commercial crop, the following phytosanitary applications should be noted: for the control of CLM, Polytrin and Taura were applied in August/September 2024, with an additional application of Taura in November 2024; Joiner was applied in January and March 2025. For BES, Auge was applied in August/September and December 2024; Comet in October 2024 and February 2025; and Cercobin in December 2024 and April 2025. For the control of CLR, Comet was applied in October 2024 and February 2025.

Both study regions have a Red Latosol soil type and a tropical savanna climate (Aw according to the Köppen classification) with a dry season in winter (Reibota et al. (2015); Amaral et al (2004)). The experimental area is located at an average altitude of 889 m, while the commercial area is situated 840 meters above sea level. The geographic location of both areas is shown in Figure 1. The climatic conditions recorded throughout the year 2024, including temperature and precipitation, are detailed in Graphic 1.

FIGURE 1. Location map of the experimental area at the Federal University of Uberlândia, Monte Carmelo Campus, and the commercial area at Juliana Farm.


LOCATION OF THE STUDY AREA
MT
GO
MG
MS
SP
ES
RJ
GO
Noroeste de Minas
Central Mineira
Triângulo Mineiro/Alto Paranaíba
MS
SP
Oeste de Minas
Monte Carmelo
EXPERIMENTAL AREA
7927520
7927500
233820
233840
233860
233880
N
0
5
10 m
COMMERCIAL AREA
7931200
7931000
230000
230200
230400
230600
SECTOR 3
SECTOR 4
SECTOR 5
SECTOR 7
SECTOR 6
N
0
50
100
REFERENCE SYSTEM: WGS84
COORDANATE SYSTEM: UTM-23S
ELABORATION: SILVA, F. V. (2025)

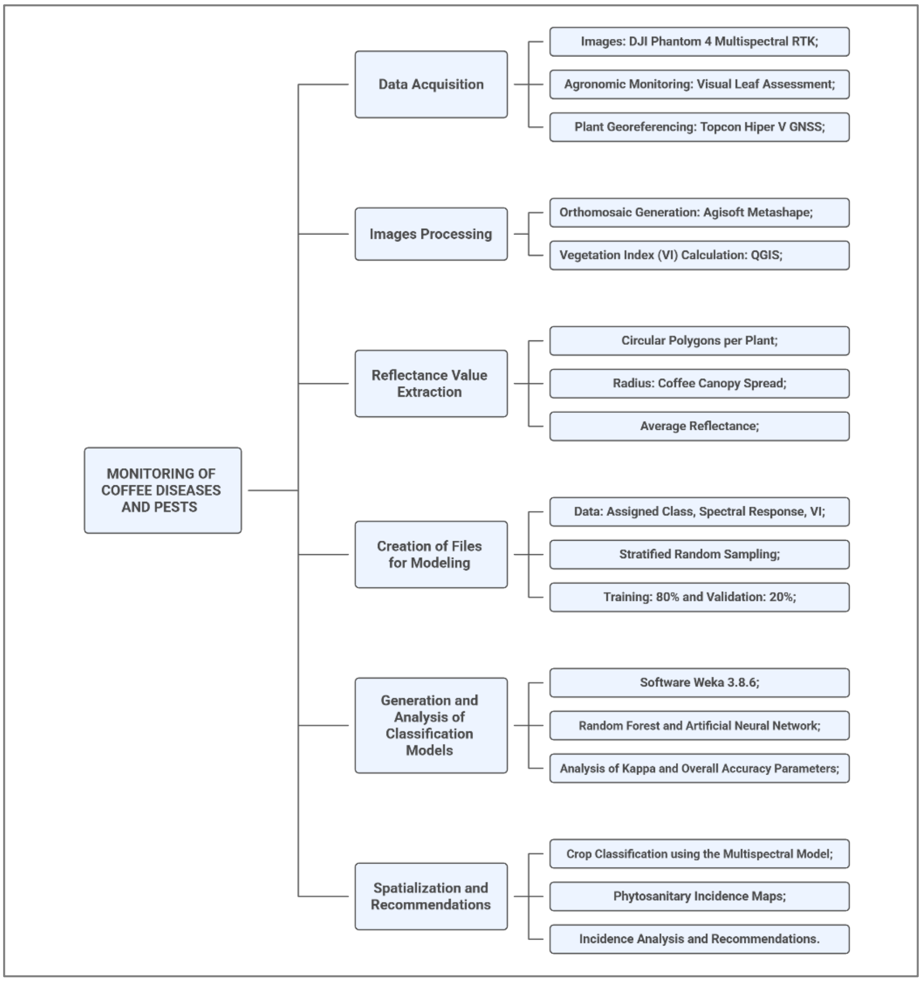
GRAPHIC 1. Climate regime of the study region: monthly variation of temperature, relative humidity, and precipitation in the year 2024.



3. METHODOLY

The methodology developed for this study is briefly presented in the flowchart shown in Figure 2. Subsequently, the materials used and the entire methodological process are described in detail.

FIGURE 2: Climate Methodological workflow integrating remote sensing and machine learning for coffee disease and pest (CLM, BES, CLR) monitoring. Key steps include: (1) multispectral drone imaging and field sampling; (2) image processing and reflectance extraction; (3) dataset preparation for modeling; (4) supervised classification using Random Forest and Neural Networks; and (5) spatial analysis and map generation for phytosanitary management.



3.1 Data Acquisition

For imaging, the DJI Phantom 4 Multispectral drone was used, equipped with a multispectral camera (visible bands 450 nm - 650 nm, red edge 730 nm, and near-infrared 840 nm). According to Figueiredo e Figueiredo (2018), the flight plan should be defined according to the objective of interest, where parameters such as flight altitude, speed, and overlap (photo overlap) can be adjusted. For overlaps, the minimum recommended is 70% front overlap and 60% side overlap, while the flight altitude is related to the level of detail required, with the GSD (Ground Sampling Distance), with a minimum of 30 m being recommended for drone safety. Finally, speed can impact the rolling shutter effect which causes target displacement (dragging/blurring in the image), with low speed being recommended in cases of low flight altitudes.

Considering this, the experimental plot was imaged monthly from February to November 2024, with the following flight parameters: altitude of 30 m, flight trajectory parallel to the planting rows, 75% front overlap and 60% side overlap, speed of 3 m/s, and static photography, providing a GSD of 1.6 cm and high-quality images. For the commercial area, the aerial survey was conducted in July 2025, with an altitude of 90 m, maintaining the other parameters and obtaining a GSD of 4.8 cm, to assess whether cartographic generalization, even if small, would significantly impact the results.

Additionally, to expand the spatial analysis and validate the results on an orbital remote sensing scale, a Sentinel-2 satellite scene with atmospheric correction – Level 2A – was acquired and used. The MSI (Multispectral Instrument) sensor onboard the satellite has 13 spectral bands (visible, near-infrared, red edge, shortwave infrared, and atmospheric correction bands) with spatial resolutions of 10, 20, and 60 meters and a radiometric resolution of 12 bits. In order to ensure direct comparability with the data acquired via drone, the Sentinel-2 bands that best correspond to the spectral ranges of the multispectral camera used on the drone were selected. Thus, the following bands were used for processing: B2 (Blue – 490 nm), B3 (Green – 560 nm), B4 (Red – 665 nm), B6 (Red Edge – 740 nm), and B8 (Near InfraRed – 842 nm).

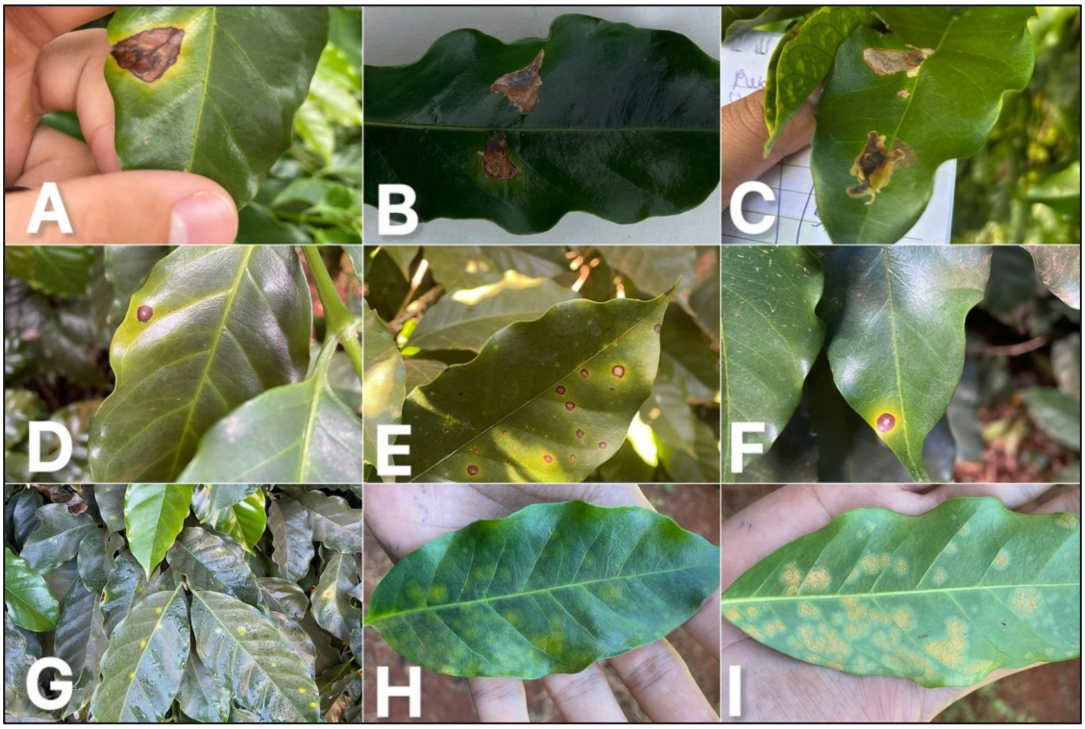
The period chosen for the test in the commercial area is justified by the coffee plant's reproductive phase, which confers a high fruit load to the plants, making the incidence of diseases more severe (CARVALHO; MATOS; PEREIRA, 2019). Furthermore, the month of July is important for monitoring the main coffee pest (CLM), as it precedes the peak of infestation and allows for preventive actions to minimize the incidence curve (Silva et al., 2019). Additionally, even though it is not a favourable moment for the development of fungal diseases, there were records of mild temperatures and rain, which allowed the development of all diseases and the pest under study (Vieira Junior et al., 2020).

According to data from the meteorological station of the Federal University of Uberlândia – Monte Carmelo Campus (18° 43' 40.2" S - 47° 31' 22.6" W), in July 2025, the average relative humidity was 63.48% and the temperature varied on average from 13.13°C to 26.53°C, which allowed the development of CLR, CLM, and BES, since such characteristics made the environment favourable for the development of all the phytosanitary problems under study (Silva et al., 2019; Vieira Junior et al., 2020). It is worth noting that the incidence of CLR normally occurs 20 to 30 days after the occurrence of light to moderate rain (Vieira Junior et al., 2020). Thus, despite the lack of precipitation in July, in the month prior to monitoring there were small precipitations totaling 6.6 mm, thus allowing the emergence of the disease and, consequently, the evaluation of all diseases and the pest under study.

For each flight, plants were visually assessed for the occurrence of phytosanitary problems by analyzing 4 leaves per plant (the third or fourth pair of leaves in the middle third of the plant, with 2 on the north face and 2 on the south face of the plant). Given that CLM shows the formation of mines, rupturing of leaf tissues, and more brownish spots, while BES is characterized by circular pinpoint yellow spots with a white center and CLR presents a circular chlorosis (most often) on the adaxial leaf surface and fungal spores on the abaxial leaf surface, the visual assessment was performed by observing the presence of these characteristics in each sampled leaf. It is important to highlight that a diagrammatic scale was not used to classify the severity of the diseases and pest, but rather only the incidence (presence or absence) of these phytosanitary problems was assessed, since the objective was to classify the plants according to their plant health status and not the level of infestation.

Figure 3 presented below illustrates the variability of lesions caused by CLM (frame A, B, C), BES (frame D, E, F), and CLR (frame G, H, I) that were assessed in the field, where distinct patterns for each lesion can be observed.

FIGURE 3: Standard pattern of leaf lesions caused by CLM, BES, and CLR sampled in the study region.



The monitored plants were georeferenced using a Topcon Hiper V GNSS receiver. The collection of plant coordinates in the experimental area was performed using the RTK (Real Time Kinematic) method, with an antenna height of 2 meters, obtaining coordinates with an average accuracy of 9 mm. It is noteworthy that the SAT 99657-point (located at the Federal University of Uberlândia – Araras Unit) is a forced-centering benchmark that served as the base for the survey, with the base data post-processed using the PPP (Precise Point Positioning) method. Following the same workflow, in the commercial area, the antenna height was changed to 3 meters (due to the tall stature of the plants) and a point was established to serve as the base for the survey (with 20 minutes of tracking). This involved static positioning post-processed for the base, and the monitored points were obtained using the RTK method. However, the coordinate accuracy varied, averaging 7 mm and 14 mm in the Easting and Northing coordinates, respectively.

3.2 Image Processing, Spectral Data Extraction, and Modeling File Creation

Data processing began with the generation of orthomosaics using Agisoft Metashape software, with photo alignment quality and point cloud construction parameters set to maximum quality. From this image, 16 vegetation indices (VIs - see Appendix A) were calculated using the QGIS raster calculator to highlight plant biophysical characteristics, such as chlorophyll content, vegetative vigor, biomass density, and leaf structure.

The selection of these VIs was based on their recognized ability to identify physiological changes in plants under biotic stress. For example, the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), and Soil Adjusted Vegetation Index (SAVI) are widely used to estimate vegetative vigor and leaf cover density (Liu, 2006; Tayade et al., 2022). The Modified Chlorophyll Absorption in Reflectance Index (MCARI) and Average Rate Index (ARI) are sensitive to chlorophyll content and the presence of secondary compounds associated with plant defence responses (Tayade et al., 2022). Meanwhile, VIs such as the Chlorophyll Index - Green (CIG) and Normalized Difference Red Edge Index (NDRE) utilize bands near the red edge, proving useful for detecting subtle leaf changes before the manifestation of visual symptoms (Mahlein, 2016; Wan et al., 2022). The inclusion of VIs like the Modified Triangular Vegetation Index 2 (MTVI2) and Triangular Vegetation Index (TVI) also helps capture variations in vegetation structure and density, aiding in the distinction of areas affected by pests and diseases with similar spectral symptoms.

Subsequently, circular polygons were created around each point, considering the lower third of the coffee plant as the radius, to represent individual plants. These features were used to extract the average reflectance values for each band of the orthomosaic and the corresponding VI responses. Finally, files for generating plant classification models were created, containing point coordinates, spectral responses from the bands and VIs, and the designated class (healthy, with CLM, with BES, or with CLR). Plants showing more than one disease simultaneously on the assessed leaves were discarded.

Since the selection of spectral bands from the Sentinel-2 satellite image ensured spectral resolution equivalence between the two datasets, the main remaining differences were spatial resolution (~5 cm from the drone vs. 10 m from the satellite) and radiometric resolution (12-bit from the satellite vs. 8-bit from the drone). To isolate and quantify the individual impact of these resolutions on the results, a controlled comparative analysis was conducted. To test the effect of radiometric resolution in isolation, the high-resolution drone image was resampled to 10 meters, simulating the spatial resolution of the satellite image band, and then the spectral signature extraction process was repeated using this resampled image. This approach allowed for the separate assessment of: (A) the influence of the satellite's superior radiometric resolution and (B) the effect of spatial resolution loss (from 4.8 cm to 10 m) on the classification model's accuracy.

Finally, the files for generating the plant classification models were created, containing the point coordinates, the spectral responses of the bands and VIs, and the designated class (healthy, with CLM, with BES, or with CLR), discarding plants that concurrently presented more than one disease on the evaluated leaves.

This dataset was divided into two independent subsets, one for model training and another for validation. The division was performed through stratified random sampling, maintaining class proportions and ensuring balanced representativeness in both subsets. This strategy aims to avoid learning bias caused by class imbalance, which is particularly crucial for supervised models applied to multi-class classification problems multiclasse (Kotsiantis, 2007).

To determine sample sizes, the number of samples per class was standardized at 15, representing the smallest number of samples identified in one class during the months of May and July in the experimental area. Thus, each evaluated month contained 10 points per class in the training file and 5 in the validation file. However, the months of August and October were discarded for not meeting this minimum criterion, as 175 out of 185 plants were infested by CLM, resulting in an insufficient number of samples for the other classes. This same sampling protocol was applied to the commercial area, where 60 samples (15 points per class) were randomly collected to compose the models.

3.3 Generation of Classification Models

For plant classification, the Random Forest (RF) and Artificial Neural Networks (ANN) algorithms were employed, implemented in Weka 3.8.6 software. The default software configurations were used: for ANN, learning rate = 0.3, momentum = 0.2, training time = 500, batch size = 100, and validation threshold = 20; for RF, number of iterations = 100, number of features = $\sqrt{n}$ (value 0 in the software) and maximum depth = 0 (unlimited).

The selection of these methods is justified by their robustness and wide application in multivariate classification problems in agronomic contexts, particularly in complex scenarios such as distinguishing between biotic stresses with overlapping spectral symptoms. RF, being an ensemble of decision trees, performs well even with correlated variables and provides clear interpretations of attribute importance. On the other hand, ANN can model nonlinear relationships and capture complex patterns in the data, which is particularly useful when working with variables that respond differently to physiological changes in plants (Kotsiants, 2007). Model performance was evaluated based on Overall Accuracy and the Kappa Index, metrics suitable for measuring performance in multi-class classifications with potential imbalances, as commonly occurs in field experiments.

Additionally, attributes were ranked according to their correlation with the 'class' variable. Subsequently, variables with lower correlation were progressively removed to simplify the model without loss of accuracy. This approach is particularly relevant for precision agriculture, as it optimizes computational costs and, more importantly, facilitates practical field application by selecting a minimal dataset that still guarantees good results.

3.4 Spatialization and Recommendations

Evaluating the performance of the models for each monitored month, the parameters of the best models, their component attributes, and the highest-performing algorithm were consolidated into a table. This made it possible to identify the optimal timing for image acquisition to achieve more accurate classification, assess the methodology's ability to detect diseases and the pest before the infestation peak, and determine whether the selected attributes followed a consistent pattern or varied each month.

Finally, for the spatialization of phytosanitary problem distribution, the classification generated for the commercial area was chosen. This choice is due to the fact that it is a larger plot with management practices closer to real-world commercial production, which gives the results greater practical applicability. For this purpose, the image was resampled so that each pixel had a 1-meter resolution. The centroid of each pixel was generated, and those corresponding to the crop canopy were selected, discarding exposed soil, shadows, and planting gaps. Based on these points, a Weka file was assembled to generate the class for each plant.

In the software, the training points from the commercial area were added to generate the classification model, and the remaining points (with no defined class) were inserted to compute their class. With the class defined for each point, in QGIS, an estimated surface of disease distribution across the plot was generated by rasterizing the classified points. As a result, phytosanitary incidence maps for the coffee crop were produced, enabling the analysis of the spatial distribution of the studied stresses and serving as a support tool for sustainable crop management.

To conclude, an analysis was conducted on the behavior of spatial distribution and the incidence level of the studied phytosanitary problems, aiming to support agronomic recommendations for the efficient control of CLM, BES, and CLR in the crop.

4. RESULTS

Supervised classification models demonstrate the feasibility of using multispectral images combined with machine learning algorithms to distinguish between healthy coffee plants and those infected with CLM, BES, and CLR. The results from the experimental plot, shown in Table 1, highlight that the models achieved an overall accuracy ranging from 66.67% to 80%, with Kappa indices between 0.5 and 0.7. The models corresponding to the months of February and November showed the best performance, both with 80% accuracy.

TABLE 1: Best Classification Models – Experimental area.

Experimental Area					
Training: 10 samples – Validation: 5 samples					
Month	Algorithm	Kappa	Accuracy (%)	Class estimated	Components
Feb	**MP**	**0.7**	**80**	**CLM, BES, Healthy**	**G, NDRE, SRI, RE**
Apr	MP	0.6	73.33	CLM, BES, Healthy	CIG
Mai	MP	0.6	70	CLM, BES, CLR, Healthy	ARI, RVI, ARVI, G, R
Jun	RF	0.5	66.67	CLM, CLR, Healthy	R, G, RE, NIR
Jul	RF	0.6	73.33	CLM, CLR, Healthy	SRI, R, MSAVI, NIR, G, MTVI2, B, CVI, GLI
Nov	**RF**	**0.6**	**80**	**CLM, Healthy**	**SRI, NDRE, MSAVI, NIR, GNDVI, NDWI, MTVI2, TVI, SAVI, NDVI, CIG, ARVI, RE, MCARI, CVI, GLI, R, RVI, G, ARI**

Considering the six months evaluated in the experimental area, regarding the tested algorithms, it was not possible to establish the exclusive use of a single algorithm, as the first three months performed best with the Artificial Neural Network (ANN) and the following three with Random Forest (RF). However, it is noteworthy that the models generated using the ANN demonstrated greater adaptability to the data, as they used fewer components and achieved a higher average accuracy.

Regarding the components of the best models from the experimental area, a considerable variation in the number of attributes was observed, ranging from 1 (in April) to 20 (in November). Consequently, it was not possible to establish a standard set of VIs or spectral bands that repeated across all models. However, it is important to note that the spectral band of the green wavelength was present in 5 out of the 6 evaluated models and was also the only VI component in April (CIG).

The variation in model performance throughout the year indicates the presence of more favorable temporal windows for more accurate monitoring. It is observed that the models in February and November showed the highest accuracy (80%), demonstrating this to be the ideal time for image acquisition and model generation. Furthermore, the monthly variation in each model's performance can be understood through the confusion matrices shown in Figure 4, where it is possible to identify the classification error patterns for each evaluated month, as the matrices illustrate the confusions each algorithm had in formulating the classification model.

FIGURE 4: Confusion matrix of the best models – Experimental area.

February 2024					
		Estimated Class			Omission Error
		Healthy	CLM	BES	
Ground Truth	Healthy	5	0	0	0
	CLM	0	4	1	20
	BES	2	0	3	40
Inclusion Error		28.57	0	25	

June 2024					
		Estimated Class			Omission Error
		Healthy	CLM	BES	
Ground Truth	Healthy	2	3	0	60
	CLM	0	4	1	20
	BES	1	0	4	20
Inclusion Error		33.33	42.86	20	

April 2024					
		Estimated Class			Omission Error
		Healthy	CLM	CLR	
Ground Truth	Healthy	5	0	0	0
	CLM	1	2	2	60
	CLR	1	0	4	20
Omission Error		28.57	0	33.33	

Juy 2024					
		Estimated Class			Omission Error
		Healthy	CLM	CLR	
Ground Truth	Healthy	3	2	0	40
	CLM	1	4	0	20
	CLR	1	0	4	20
Inclusion Error		40	33.33	0	

May 2024						
		Estimated Class				Omission Error
		Healthy	CLM	BES	CLR	
Ground Truth	Healthy	4	0	0	1	20
	CLM	1	3	0	1	40
	BES	2	0	2	1	60
	CLR	0	0	0	5	0
Inclusion Error		42.86	0	0	37.5	

November 2024				
		Estimated Class		Omission Error
		Healthy	CLM	
Ground Truth	Healthy	4	1	20
	CLM	1	4	20
Inclusion Error		20	20	

Monthly analysis can be performed by evaluating each matrix individually, but comparing periods that share the same classes reveals that the confusion between them varies. For example, in February, 67% of the error (2 out of the 3 misclassified plants) is attributed to the classification of plants with BES, while in April, 75% of the error (3 out of the 4 misclassified plants) is due to plants with CLM. This demonstrates that the spectral signature of plants with the same phytosanitary problems can vary according to the phenological stage and incidence level.

Therefore, table 2 was generated to enable an overall analysis of the matrices. In it, the F1-Score indicates which classes the algorithms had the most difficulty classifying, which affects the final accuracy of each model. In this regard, plants with BES rank as the most difficult to classify, while those with CLR are the easiest. This pattern suggests that BES symptoms cause less pronounced spectral changes, making them more difficult for the algorithm to detect.

TABLE 2: Statistical analysis – Experimental area.

Class	Total	True Positive	False Negative	False Positive	Precision	Recall	**F1-Score**
Healthy	30	23	7	11	0.676	0.767	0.719
CLM	30	21	9	6	0.778	0.700	0.737
BES	15	9	6	3	0.750	0.600	**0.667**
CLR	15	13	2	4	0.765	0.867	**0.813**

Regarding the results of the classification models for the commercial area, Table 3 shows that the best model generated for the commercial area using the high-resolution aerial image (~5 cm) was obtained through MP, achieving an accuracy of 70% and a Kappa index of 0.6. As the spatial resolution decreased (10 m), accuracy dropped to 65%. However, considering the use of the orbital image, the best plant classification model achieved 85% accuracy and a Kappa index of 0.8. Meanwhile, Tables 4, 5, and 6 present the confusion matrices for the models generated using the higher-resolution aerial image (4.8 cm), the resampled aerial image (10 m), and the orbital image, respectively.

TABLE 3: Best Classification Models for the Commercial Area.

<table>
<tr><th colspan="6">Commercial Area</th></tr>
<tr><th colspan="6">Aereal Imagery - GSD = 4.8 cm</th></tr>
<tr><th>Month</th><th>Algorithm</th><th>Kappa</th><th>Accuracy (%)</th><th>Class estimated</th><th>Components</th></tr>
<tr><td>Jul</td><td>ANN</td><td>0.6</td><td>70</td><td>CLM, BES, CLR and Health</td><td>R, G, B, RE, NIR</td></tr>
<tr><td colspan="6">Aereal Imagery - GSD = 10 m</td></tr>
<tr><td>Jul</td><td>RF</td><td>0.53</td><td>65</td><td>CLM, BES, CLR and Health</td><td>NDVI, ARVI, ARI, SAVI, GNDVI, RVI, MCARI, CVI, CIG, NDWI</td></tr>
<tr><td colspan="6">Orbital Imagery – Sentinel 2</td></tr>
<tr><td>Jul</td><td>ANN</td><td>0.8</td><td>85</td><td>CLM, BES, CLR and Health</td><td>R, G, B, RE, NIR</td></tr>
</table>

TABLE 4: Confusion Matrix of the Best Aerial Image Classification Model – Commercial Area.

Commercial Area - July 2025						
		Estimated Class				Omission Error
		Healthy	CLM	BES	CLR	
Ground Truth	Healthy	3	1	0	1	40
	CLM	1	3	1	0	40
	BES	1	0	3	1	40
	CLR	0	0	0	5	0
Inclusion Error		40	25	25	28.57	

TABLE 5: Confusion Matrix of the Best Resampled Aerial Image Classification Model (10m) – Commercial Area.

Commercial Area - July 2025						
		Estimated Class				Omission Error
		Healthy	CLM	BES	CLR	
Ground Truth	Healthy	3	0	1	1	40
	CLM	1	4	0	0	20
	BES	2	0	3	0	40
	CLR	1	1	0	3	40
Inclusion Error		57.14	20	25	25	

TABLE 6: Confusion Matrix of the Best Orbital Image Classification Model – Commercial Area.

Commercial Area - July 2025						
		Estimated Class				Omission Error
		Healthy	CLM	BES	CLR	
Ground Truth	Healthy	4	1	0	0	20
	CLM	1	4	0	0	20
	BES	2	0	3	0	40
	CLR	1	1	0	3	40
Inclusion Error		50	33.33	0	0	

To turn this result into a practical monitoring tool, the maps presented in Figures 5 and 6 spatialized the classification of plants in the commercial area based on the multispectral model, allowing for the visualization of critical infestation areas. Figure 5 displays the overall distribution of all diseases, pests, and healthy plants in a single map, while Figure 6 shows in detail the incidence of each class.

FIGURE 5: Spatial distribution map of phytosanitary problems in the commercial area.

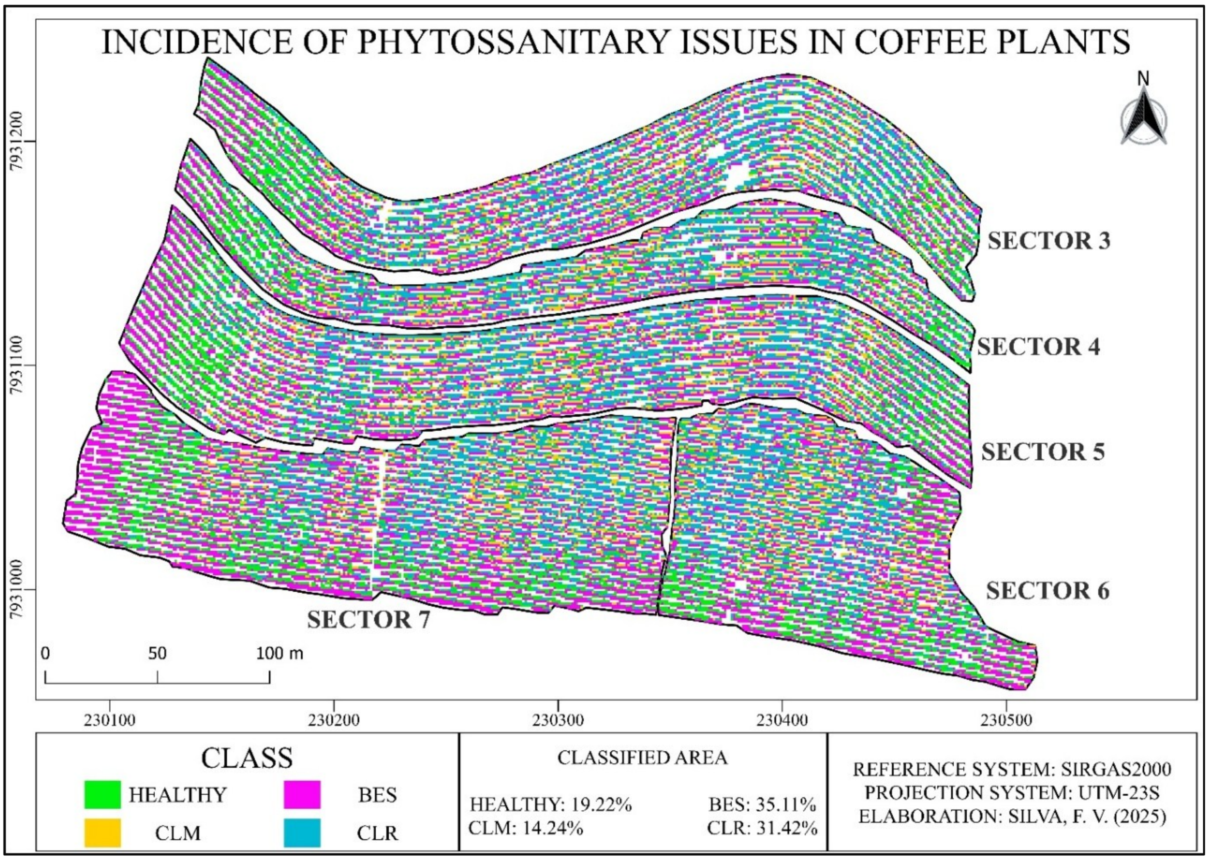
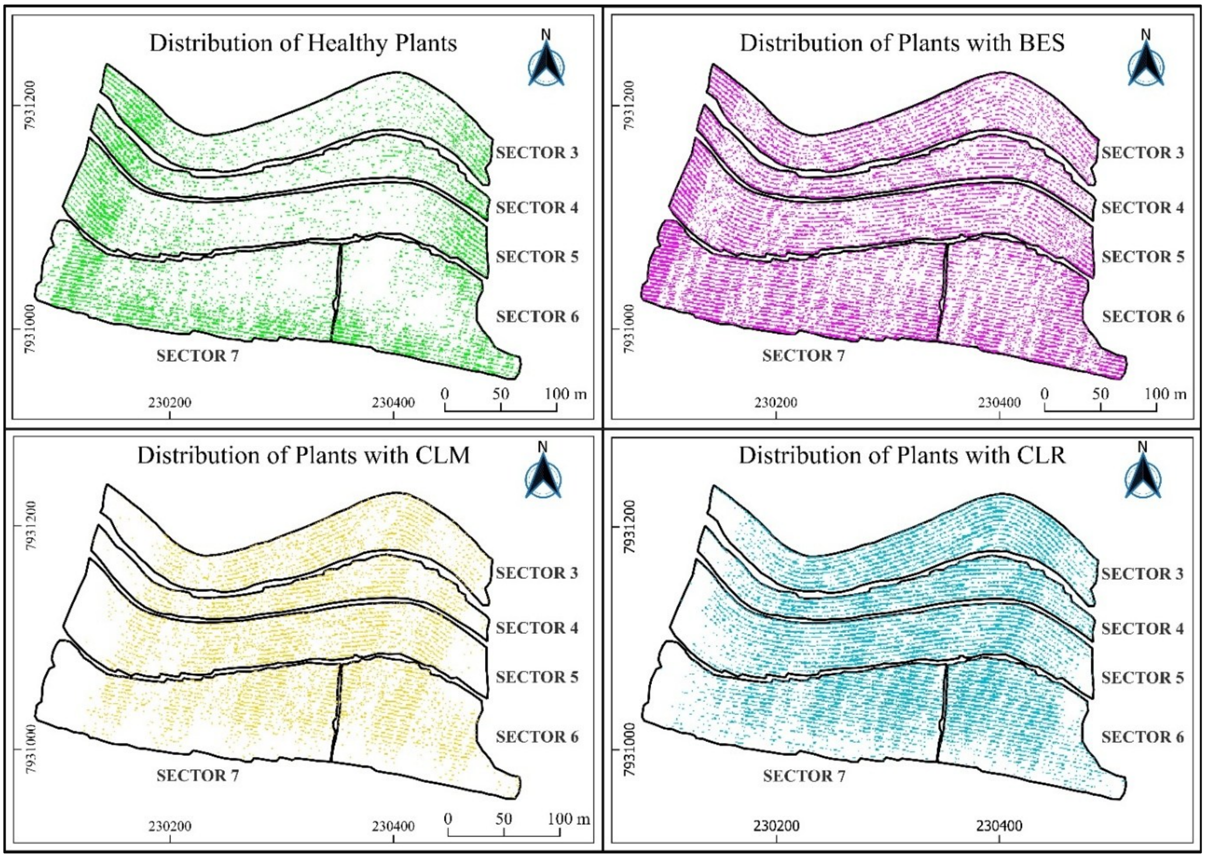


Figure 6: Spatial distribution map of plants with CLM, BES, CLR, and healthy plants.



The maps revealed the distribution patterns of the phytosanitary problems, with incidence occurring in straight-line strips at distinct levels. It is worth highlighting the incidence of BES spread throughout the area, mainly at the edges, and CLR more present in the center of the plot.

5. DISCUSSION

Figure 3 presents visual examples of leaf lesions caused by CLM, BES, and CLR, highlighting relevant differences among the symptoms. It can be observed that CLM lesions are more extensive and have a distinct coloration compared to BES and CLR lesions, which tend to be smaller, more punctual, and visually less evident. This difference may partially explain the greater difficulty of the model in correctly identifying plants with BES, since subtle lesions generally cause smaller changes in the reflectance captured by the sensor's spectral bands, making it difficult to distinguish them from healthy plants.

This limitation is reflected in the confusion matrices of the experimental area, which showed higher error rates in the classification of plants with BES, while healthy plants, those with CLR, and those affected by CLM were classified with greater accuracy. In the commercial area, with the decrease in spatial resolution of the aerial image (from ~5 cm to 10 m), the confusion matrix of the model generated with the resampled image did not follow the confusion pattern of the experimental area models, nor that of the model generated with the high-resolution image. In this model, plants with CLR, which were previously classified more easily, showed a higher confusion rate (Table 5). This also holds true for the confusion matrix of the orbital image classification model, where healthy plants and CLR caused the greatest errors. This demonstrates that the overlap of symptoms, which occurs with the generalization of spectral information through resampling, can be decisive for generating more accurate models. This behavior reinforces the importance of symptom intensity and extent for the success of remote sensing-based classifications, particularly in supervised methods.

The literature corroborates these findings. According to Mahlein (2016), healthy plants exhibit high absorption in the visible region and higher reflectance in the near-infrared, while stressed plants tend to reverse this pattern. The VIs used specifically explore these spectral variations to detect physiological changes associated with biotic stress. CIG, highlighted in the April model for example, is sensitive to chlorophyll content and therefore useful for identifying plants with compromised photosynthesis. Furthermore, the G component present in all models is responsive to leaf pigmentation and highlights chlorophyll concentration, factors that change under biotic stress (Tayade et al., 2022; Liu, 2006).

The use of only one variable to generate the best model in April demonstrates that the infestation level and phenological stage of coffee plants can influence the spectral response of diseased plants, making them more distinguishable. In this case, it becomes evident that different stages may require a more complex or simpler model. However, to develop a generalizable model, a more in-depth study of the spectral behavior of each disease and pest across all phenological stages is necessary, along with consideration of the biennial production cycle of coffee.

In this sense, considering the seasonality of the incidence of the problems under study, the months with the highest accuracy (February and November) demonstrate the ideal window for monitoring. Since both precede the peak infestation of the diseases and pest under study (Nascente et al., 2021; Silva et al, 2019; CARVALHO; MATOS; PEREIRA, 2019), these monitoring efforts can aid decisions regarding the implementation of preventive control in the crop, reducing the incidence curve and severity of phytosanitary problems.

The high accuracy value in these classifications can be understood by analyzing the plant's condition and the cultural practices applied. During these months, the plant has greater leaf coverage and coincides with fertilization periods, thus presenting more vegetative and vigorous plants (Camargo e Camargo, 2001; Sakiyama et al., 2015; Damatta, Rena e Carvalho, 2008). Such characteristics suggest that healthy plants would stand out due to their high vegetative vigor, while those with some phytosanitary problem would have fewer leaves or chlorotic leaves, making the spectral signatures of these plants more distinct and facilitating classification via remote sensing.

The classification may also have been influenced by two other factors: the number of classes and the incidence level of the diseases. As demonstrated by Troung et al. (2019), in supervised classification, overall accuracy tends to decrease as the number of classes increases, which partly explains the performance difference of the models when compared to May, for example. Additionally, the fact that the monitoring preceded the peak infestation resulted in lower and more localized incidence in the crop. This sparse distribution of diseased plants may have made them spectrally more distinct amidst the predominantly healthy plants, facilitating their classification.

Considering the classification of the commercial area, where the image spatial resolution was lower due to the higher flight altitude required to cover the area's extent, the generated model maintained the plant classification pattern, albeit with a 3.33% reduction compared to the experimental area model from the same period. Furthermore, the accuracy of the classification model based on the resampled image showed an even greater reduction (8.33%), reaffirming that decreased spatial resolution tends to worsen classification. This reduction in accuracy may be attributed not only to the lower spatial resolution of the images but may also be related to the distinct management conditions and consequent canopy variability present at the commercial scale.

In the experimental area, the rigorous environmental control and individually applied cultural practices (plant-by-plant fertilization and phytosanitary management) result in a more homogeneous plot. In contrast, in the commercial crop, differences in microclimate, soil fertility, management, and even production load generate a structurally and physiologically more heterogeneous canopy. This heterogeneity increases the complexity of the captured spectral response, as the same pixel may integrate the reflectance of healthy and diseased leaves. Therefore, the fact that the model maintained a classification pattern with high accuracy even under these more complex conditions reinforces the robustness of the method and demonstrates potential for operational scalability.

A comparison with the research by Mendonça (2022), which has a similar methodology, demonstrates that the method proposed in this work increased classification accuracy by up to 53%, due to the increased spectral resolution and VIs tested. Considering the results from the commercial area using the orbital image, classification accuracy is even higher, due to the increased radiometric resolution of the image. Meanwhile, the work by Vilela et al. (2024), which also uses Sentinel images but considers only CLM incidence, presented better results (86% accuracy). This demonstrates that the scalability of the proposed method is possible and that the number of estimated classes, spectral resolution, spatial resolution, and radiometric resolution can influence classification accuracy.

From a practical perspective, the disease distribution map of the plot showed that BES exhibited high dispersion, present in almost the entire area. Furthermore, it is noted that approximately 80% of the canopy was classified with some of the phytosanitary problems under study, highlighting the momentary vulnerability of the crop. According to Vieira Junior et al. (2020), this high rate of disease and pest incidence may be related to the plant's pending load, which may also explain the disease distribution, as central plants tend to be more productive.

Another issue highlighted by the map is the distribution pattern of the diseases. It is observed that they have a rectangular behavior pattern in longitudinal strips interspersed along each sector. This type of well-defined format is generally associated with human interventions, representing possible problems in irrigation, fertilization, and application of agricultural inputs and defensives, but may also represent the presence of soil patches in the region. Regardless of the attenuating factor, this spatial information can be used for localized crop management, guiding the application of phytosanitary products only where necessary, reducing costs, minimizing environmental impacts, and increasing phytosanitary control efficiency.

The occurrence of BES at the edge of the plot is directly related to solar incidence in these regions of the crop, since high sunlight incidence is necessary for cercosporin activation (Daub and Chung, 2007), in addition to being frequently associated with nutritional deficiencies (Vieira Junior et al., 2020). Considering this, the georeferenced identification of affected plants may indicate areas with possible nutritional imbalances. Thus, the model can also serve as a basis for targeted fertilization recommendations and guidance for soil sampling, contributing to more integrated crop management. In this sense, based on the BES distribution map, intensified soil sampling could be recommended in the southern region of the plot and on the left edge (west) of each sector.

The plant classification demonstrated that, for the CLM class, chemical control would not be recommended, since only 14% of the area showed CLM symptoms and the adoption of control measures is recommended when the number of infected plants exceeds the threshold of 15% (Silva et al., 2019). However, continuous monitoring would be recommended to ascertain the evolution of the pest infestation curve or even application for preventive control, since this is a period preceding the peak phase (occurring in Sep./Oct. in the Minas Gerais cerrado) and the infestation rate was already close to the recommended control threshold (Silva et al., 2019). It should be noted that certain products, such as Premio Star, recommend pesticide application when monitoring indicates 3% CLM infestation. Thus, the agronomic recommendation based on plant classification should consider, in addition to the current infestation level, the product, location, and climatic conditions of the crop.

In contrast, for CLR and BES, the application of phytosanitary products would be recommended, as the disease incidence levels exceeded the limits of 5% and 10%, respectively (Vieira Junior et al., 2020; Amorim et al. 2016). Thus, for simultaneous disease control, the application of a fungicide formulated with Epoxiconazole + Fluxapyroxad + Pyraclostrobin, belonging to the chemical groups Triazole, Carboxamide, and Strobilurin, at an average dose of 1-1.5 L ha^{-1} in a spray volume of 400 to 500 L/ha for adequate foliage coverage could be recommended to the producer (Vieira Junior et al., 2020). This finding demonstrates how the use of images and classification models can support agronomic decisions based on technical and quantitative criteria.

Although the incidence map represents a significant practical gain in assessing disease and pest distribution and supporting agronomic recommendations, its generation would still require field monitoring for each analysis, which does not eliminate or reduce the cost of the traditional monitoring model. However, with the adoption of continuous plot monitoring, it would be possible to capture the crop's spectral signature and mitigate the need for field data collection for algorithm training, performing only sporadic monitoring for model recalibration. Furthermore, the spatialization of phytosanitary problem incidence allows for variable rate application, effectively providing precision agriculture and promoting sustainable coffee farming, which would indeed represent a gain from the proposed method.

6. ADVANTAGES, LIMITATIONS, AND FUTURE DIRECTIONS

The methodology of this study was developed to provide a tool as an alternative for phytosanitary monitoring of coffee plots that is practical, fast, expert, and accurate in detecting these problems. Its real contribution stems from the practical demonstration of multiscale integration, identification of strategic temporal windows, validation of operational scalability in commercial areas, multi-class classification, in addition to enabling the mapping of coffee disease and pest incidence.

The monthly monitoring of the experimental area throughout the crop cycle, which allowed the identification of optimal imaging windows in February and November, transforms the classification model into an instrument for early warning and preventive decision support, with clear environmental and economic impact. In turn, the maintenance of accuracy levels (70-85%) in the commercial area, which is more heterogeneous and has more complex data than the experimental area, highlights the scalability power of remote sensing monitoring. In this process of replicating the methodological protocol in the commercial area, the interpretation of the individual impact of spatial and radiometric resolution on the accuracy of classification models represents a major contribution of the study, as the impact of these resolutions is little explored in applied literature.

Furthermore, unlike most remote sensing studies applied to coffee farming, especially in disease and pest detection which often consider the detection of a single phytosanitary problem, this work advances by simultaneously considering the incidence of two fungal pathogens and one pest of great productive and economic impact. This multifocal approach is particularly relevant as it reflects the real complexity of the field, where coffee plantations are commonly affected by multiple phytosanitary problems that interact and coexist. The ability to spectrally distinguish these diseases, which sometimes produce similar visual symptoms, represents a significant contribution of the study, as it highlights the potential of the technique not only to identify stress in a generic way, but to provide a categorized specific diagnosis.

In contrast, the work presents limitations regarding the sampling protocol, operational cost and complexity of execution, research period, and even data scalability. According to the first law of geostatistics, near objects tend to be similar while more distant objects differ (Oliveira, Grego and Brandão, 2015). Therefore, spatial variability tends to make plants different, whether in size, vegetative vigor, coloration, etc., which directly impacts the spectral response of each plant, since higher leaf area indices decrease reflectance in the visible and near-infrared region (Ponzoni; Shimabukuro; Kuplich, 2012). Therefore, sampling can be a determining factor in the success of creating multispectral classification models.

Considering this, traditional agronomic monitoring follows the zigzag sampling pattern, visually evaluating about 60 to 200 leaves for analyzing the incidence of diseases and pests under study, which corresponds to 20 to 40 plants sampled per plot (Silva et al, 2019; CARVALHO; MATOS; PEREIRA, 2019). Thus, traditional monitoring can mitigate these nuances and make sampling more representative. However, for the experimental area this method could not be applied, as it would limit the number of samples precisely because it was such a small area (700 m²), which would compromise the generation of classification models.

In this case, the sampling design presents a fundamental limitation, since instead of performing statistically robust random sampling (such as zigzag), it was necessary to actively search for diseased plants to capture their spectral signatures. This can introduce bias, as the sampled plants may not fully represent the spectral variability of all diseased plants in the crop, especially those with initial or atypical symptoms.

Furthermore, the binary evaluation (presence or absence of phytosanitary problems) of only 4 leaves, without considering a diagrammatic scale, may not represent the actual state of the plant. A plant with minimal symptoms on one sampled leaf is classified as "diseased," while a plant with severe symptoms on unsampled leaves may be classified as "healthy." This introduces noise and uncertainty into the training and validation dataset, which propagates directly to the models, impacting classification accuracy. A more rigorous and detailed protocol analyzing more leaves per plant, with severity assessment, sampling from different parts of the canopy, for example, could improve the reliability of ground truth, but would make the entire process more costly.

The operational cost and complexity of the complete workflow, involving drone flights with specialized sensor, image processing in specific software, high-precision georeferencing, and modeling in specialized software, requires expensive equipment and a certain multidisciplinary expertise. The promise of reducing field monitoring costs will only fully materialize in a future stage of automation, where pre-calibrated models can be applied with minimal field and modeling assistance.

Regarding the research development period, the fact that experimental area monitoring occurred in only one year (2024) may represent a vulnerability of this work. The ideal scenario should consider at least one complete biennial cycle of the crop (2 years), to verify whether the models and temporal windows remain or whether it is necessary to develop specific models for high and low productivity years.

Another weakness is related to actual data scalability. The leap from a small and homogeneous experimental area (700 m^2) to a commercial one (7.8 ha) is significant, but the study did not test the application of models calibrated in the experimental area directly in the commercial area. Instead, new data were collected to train the commercial area model. This reveals that the model is not directly transferable, limiting its operational scalability. True scalability would require a generalizable model trained with data from multiple crops, varieties, and management conditions.

Furthermore, it should be noted that the machine learning models were used statically, with month-by-month analysis, capturing the specific patterns of the tested period. Thus, it was not possible to learn the temporal dynamics of disease spectral signatures. To reduce this variability of input data, more sophisticated machine learning techniques are needed, such as the use of time series algorithms (e.g., long short-term memory, transformers...), data augmentation and domain adaptation techniques (e.g., generative adversarial networks), and extraction of time-invariant features.

Given the limitations and weaknesses presented, we recommend that future work explore variation in sample size, use of time series, consider more than one disease/pest per plant, evaluate different infestation levels, promote integration of LiDAR, hyperspectral, agronomic, and climatic data, in addition to using different algorithms, with the expectation of developing a single efficient model for coffee monitoring.

The development of generalizable models using time series would allow identifying specific dynamic spectral patterns of each disease, regardless of phenological stage, generating more robust models with less need for recalibration. Furthermore, the inclusion of multi-source data is a crucial part for model generalization. Therefore, future work can capture more specific biochemical signatures of diseases through hyperspectral readings with spectroradiometer, can correlate severity with leaf volume loss through LiDAR data integration, and extend the impact of agronomic and climatic information as an additional data layer in models.

Finally, it should be noted that although the inclusion of agronomic and climatic data can increase the robustness of predictive models, this strategy requires cost-benefit analysis, as obtaining such data can increase operational costs. Nevertheless, this represents a promising perspective for future studies seeking to integrate multiple data sources to improve remote diagnosis of coffee phytosanitary conditions.

7. CONCLUSION

The results of this study demonstrate the potential of using multispectral images combined with vegetation indices and machine learning algorithms to distinguish between healthy coffee plants and those affected by CLM, BES, and CLR. The experimental area models, with 73%–80% accuracy and a Kappa index of 0.6–0.7, showed satisfactory classification performance, despite the challenges associated with spectral similarity among the symptoms of the evaluated diseases. The results from the commercial area, where plant classification accuracy ranged from 65% to 85%, demonstrated the scalability capacity of the proposed methodology for coffee monitoring.

The possibility of spatially mapping disease incidence can enable targeted management strategies, such as variable-rate application of phytosanitary products and adjustments to nutritional practices, promoting more efficient and sustainable coffee production. Despite these advances, there is still a need to improve model accuracy, particularly in detecting BES, which showed greater confusion with healthy plants. To address this challenge, it is recommended to expand the dataset, include agronomic and environmental variables, and evaluate other algorithms and approaches for fusing spectral and phenological data.

It is concluded that multispectral remote sensing is a promising tool for phytosanitary monitoring of coffee crops, with the potential to support agronomic decision-making and optimize input use in precision agriculture.

REFERENCE

AMARAL, F. C. S. et al. **Mapeamento de Solos e Aptidão Agrícola das Terras do Estado de Minas Gerais.** Embrapa Solos. Boletim de pesquisa e desenvolvimento, 63, 2004. Disponível em: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/965988 Acesso em: jan. de 2026.

AMORIM, L. et al. **Manual de Fitopatologia**: Doenças de Plantas cultivadas. 5 ed. Ouro Fino: Agronômica Ceres. 2016 v.2. 772 p.

ADELAJA, O.; PRANGGONO, B. **Leveraging Deep Learning for Real-Time Coffee Leaf Disease Identification**. AgriEngineering 2025, *7*, 13. DOI: https://doi.org/10.3390/agriengineering7010013

BAHT N, DOMINGUEZ E, ALJUMAILI S. **Detection of coffee leaf disease and identification using deep learning**. PeerJ Ciência da Computação 11:e3172, 2025. DOI: https://doi.org/10.7717/peerj-cs.3172

CAJAS, J. J. Z. et al. **Estimating the severity of coffee leaf rust using deep learning and image processing.** Inteligência Artificial, *[S. l.],* v. 28, n. 76, p. 200–222, 2025. DOI: https://doi.org/10.4114/intartif.vol28iss76pp200-222.

CAMARGO, A. P.; CAMARGO, M. B. P. **Definition and outline for the phenological phases of arabic coffee under Brazilian tropical conditions**. Bragantia, v. 60, n. 1, p. 65–68, 2001. DOI: https://doi.org/10.1590/S0006-87052001000100008

CARVALHO, V. L.; MATOS, C. S. M.; PEREIRA, A. B. Ferrugem-do-cafeeiro. Lavras – MG, EPAMIG, 2019. Disponível em: https://www.livrariaepamig.com.br/wp-content/uploads/2023/02/Ferrugem-do-cafeeiro.pdf . Acesso em: 8 out. 2025.

CHEDID, V.; CORTEZ, J. W.; ARCOVERDE, S. N. S. **Monitoring The Vegetative State Of Coffee Using Vegetation Indices**. Engenharia Agrícola, v. 44, p. e20220212, 2024. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v44e20220212/2024

DAMATTA, F. M.; RENA, A. B.; CARVALHO, C. H. S, 2008. **Aspectos Fisiológicos do Crescimento e da Produção Do Cafeeiro**. In: Carvalho, CHS de. (Ed.). Cultivares de café: origem, característica e recomenda ções. Brasília, EMBRAPA CAFÉ. p.33-55.

DA SILVA, P.C. et al. **Multispectral Images for Drought Stress Evaluation of Arabica Coffee Genotypes Under Different Irrigation Regimes**. Sensors 2024, *24*, 7271. DOI: https://doi.org/10.3390/s24227271

DAUB, M. E.; CHUNG, K. **Cercosporin: A Photoactivated Toxin in Plant Disease.** Disponível em: https://www.researchgate.net/publication/251453457_Cercosporin_A_Photoactivated_Toxin_in_Plant_Disease Acesso em: jan. 2025

DIAO, C.. **Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages**. Remote Sensing of Environment, 248, 111960, 2020. DOI: https://doi.org/10.1016/j.rse.2020.111960

DOS SANTOS, L. M. et al. **Use of Images Obtained by Remotely Piloted Aircraft and Random Forest for the Detection of Leaf Miner (*Leucoptera coffeella*) in Newly Planted Coffee Trees**. Remote Sens. 2024, 16, 728. DOI: **https://doi.org/10.3390/rs16040728**

ENNOURI et al. **Application of remote sensing technology for plant disease detection.** Journal of Arid Arboriculture and Olive Growing. v4(1):13-18, 2025. Disponível em:

https://www.researchgate.net/publication/396371363_Application_of_remote_sensing_technology_for_plant_disease_detection Acesso em: dez. 2025.

FIGUEIREDO, E. O.; FIGUEIREDO, S. M. M. **Planos de Voo Semiautônomos para Fotogrametria com Aeronaves Remotamente Pilotadas de Classe 3**. Embrapa Acre. Circular técnica, 75, 2018 Disponível em: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1100860. Acessoe m: dez. 2025.

KOTSIANTIS, S. B. **Supervised machine learning: A review of classification techniques**. Informatica, 31, 249–268, 2007.

LIU, W. T., **Aplicações de Sensoriamento Remoto**. Campo Grande: UNIDERP, 908p., 2006.

LONGCHAMPS, L., PHILPOT, W. **Full-Season Crop Phenology Monitoring Using Two-Dimensional Normalized Difference Pairs**. Remote Sensing 15(23), 5565, 2023. DOI: https://doi.org/10.3390/rs15235565.

LOSCH, E. L. et al. **Os Agrotóxicos no Contexto da Saúde Única**. Saúde em Debate 46 (2), 438-454, 2022. DOI: https://doi.org/10.1590/0103-11042022E229.

MAHLEIN, A. K. **Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping**. Plant Disease, 100(2), 241-251, 2016. DOI: https://doi.org/10.1094/PDIS-03-15-0340-FE.

MATIELLO, J. B. et al. **Cultura de café no Brasil**: manual de recomendações. Varginha: Fundação Procafé, 2015.

MENDONÇA, R. C. P. Detecção De Pragas E Patógenos Do Cafeeiro Através De Imagens Multiespectrais De Baixo Custo E Algoritmos Baseados Em Aprendizado De Máquina . 2022. Trabalho de Conclusão de Curso. Universidade Federal de Uberlândia, Monte Carmelo, 2022. Disponível em: https://repositorio.ufu.br/bitstream/123456789/34474/1/Detec%c3%a7%c3%a3oPragasPat%c3%b3genos.pdf

NASCENTE, T. F. et al. **Climatic conditions in the cercosporiosis (Cercospora coffeicola) incidence and leaf-miner (Leucoptera coffeella) in coffee cultivars in Monte Carmelo, Minas Gerais, Brazil**. Research, Society and Development, v. 10, n.3, e29210313304, 2021. DOI: http://dx.doi.org/10.33448/rsd-v10i3.13304.

OLIVEIRA, R. P.; Grego, C. R.; Brandão, Z. N. **Geoestatística Aplicada na Agricultura de Precisão Utilizando o Vesper**. 2015. Embrapa, Brasília – DF. Disponível em: https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1051869/1/GeoVespercap3.pdf Acesso em: dez. 2025.

ORLANDO, V. S. W. et al. **SENSORIAMENTO REMOTO: UMA PERSPECTIVA PARA DETECÇÃO DE DOENÇAS NA CAFEICULTURA.** II Simpósio Regional de Agrimensura e Cartografia, 2021. Disponível em: https://www.researchgate.net/publication/371991147_SENSORIAMENTO_REMOTO_UMA_PERSPECTIVA_PARA_DETECCAO_DE_DOENCAS_NA_CAFEICULTURA

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento remoto da vegetação.** 2ª edição – atualizada e ampliada. São Paulo: Oficina de Textos, 2012.

POZZA E. A.; CARVALHO. V. L.; CHALFOUN, S. M. Sintomas de Injúrias Causadas por Doenças em Cafeeiro. In: Guimarães, R. J., Mendes, A. N. G., Baliza,D. P. Semiologia do Cafeeiro: sintomas de desordens nutricionais. Lavras: UFLA, 69-106, 2010.

REIBOTA, M. S. et al. **Aspectos Climáticos Do Estado De Minas Gerais (Climate Aspects In Minas Gerais State).** Revista Brasileira De Climatologia, [*S. l.*], v. 17, 2015. Doi: https://doi.org/10.5380/abclima.v17i0.41493.

SAKIYAMA, N. S. et al. **Café arábica**: do plantio à colheita. Viçosa, MG, 2015. Ed: UFV. 316 p.
SILVA, R. A. et al. **Bicho-mineiro-do-cafeeiro**. Lavras – MG, EPAMIG, 2019. Disponível em: https://livrariaepamig.com.br/wp-content/uploads/2023/02/Bicho-mineiro-do-cafeeiro.pdf Acesso em: 21 out. 2025.

SILVA, C. E. S. E. et al. **Low Computational Cost Deep Learning Approach for Localization and Classification of Diseases and Pests in Coffee Leaves.** IEEE Access, vol. 13, pp. 71943-71964, 2025, doi: 10.1109/ACCESS.2025.3562832.

TAYADE, R. et al. **Utilization of Spectral Indices for High-Throughput Phenotyping**. Plants 11(13), 1712, 2022. DOI: https://doi.org/10.3390/plants11131712.

TRUONG et al. **How Does Land Use/Land Cover Map's Accuracy Depend on Number of Classification Classes?**. SOLA, 2019, Vol. 15, 28−31, DOI:10.2151/sola.2019-006

VIEIRA JUNIOR et al**. Identificação e manejo de doenças do cafeeiro (Coffea canephora)**. In: Vieira Junior, J. R.; Costa, J. N. M. (Ed.). Guia de bolso: diagnose e manejo de doenças e pragas do cafeeiro na Amazônia. Embrapa Rondônia, 1, 11-43, 2020. Disponível em: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1126495 Accessed on 13/11/2024.

VILELA, E. F. et al. **Remote Monitoring of Coffee Leaf Miner Infestation Using Machine Learning**. AgriEngineering 2024, *6*, 1697-1711. DOI: https://doi.org/10.3390/agriengineering6020098

WAN, L. et al. **Hyperspectral Sensing of Plant Diseases**: Principle and Methods. Agronomy, 12 (6), 1451, 2022. DOI: https://doi.org/10.3390/agronomy12061451.

ZHANG, S. et al. **A review of the application of UAV multispectral remote sensing technology in precision agriculture.** Smart Agricultural Technology, v12, 101406, 2025. DOI: https://doi-org.ez34.periodicos.capes.gov.br/10.1016/j.atech.2025.101406

APPENDIX A

Table A.1: Vegetation Index Formulas.

Vegetation Index	
$ARI = \frac{1}{G} - \frac{1}{RE}$	$MTVI2 = \frac{1.5 * \left(2.5 * (NIR - B) - \left(2.5 * (R + G)\right)\right)}{\sqrt{(2 * NIR + 1)^2 - \left(6 * NIR - 5 * \sqrt{R}\right) - 0.5}}$
$ARVI = \frac{(NIR - (2 * R) + B)}{(NIR + (2 * R) + B)}$	$NDVI = \frac{(NIR - R)}{(NIR + R)}$
$CIG = \frac{NIR}{G} - 1$	$NDRE = \frac{(NIR - RE)}{(NIR + RE)}$
$CVI = NIR * \frac{(R)}{(G^2)}$	$NDWI = \frac{(G - NIR)}{(G + NIR)}$
$GLI = \frac{(2 * G - R - B)}{(2 * G + R + B)}$	$RVI = \frac{(NIR)}{(R)}$
$GNDVI = \frac{(NIR - G)}{(NIR + G)}$	$SAVI = \frac{(1.5 * (NIR - R))}{(NIR + R + 0.5)}$
$MCARI = ((RE - R) - (0.2 * (RE - G))) * \left({}^{RE}/_{R}\right)$	$SRI = \frac{(NIR)}{(RE)}$
$MSAVI = \left((2 * NIR + 1) - \sqrt{(2 * NIR + 1) - (NIR - R)}\right)/2$	$TVI = \sqrt{NDVI + 0.5}$

Where: B = Blue, G = Green, R = Red, RE = Red Edge and NIR = Near InfraRed.