

Early detection of Chickpea Ascochyta Blight using Hyperspectral imaging Coupled with Machine learning

Mohamed ARAME

mohamed.arame@um6p.ma

<https://orcid.org/0009-0004-0785-9309>

Issam Meftah Kadmiri

Oussama Elbaraghi

Francois Bourzeix

Yahya Zennayi

Ilyass Maafa

Abdelghani Chehbouni

Research Article

Keywords: Chickpea, Ascochyta blight, Hyperspectral imaging, Early disease detection, Machine learning, Precision agriculture

Posted Date: March 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9186587/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Early detection of Chickpea Ascochyta Blight using Hyperspectral imaging Coupled with Machine learning

Mohamed Arame^{a,b}, Issam Meftah Kadmiri^a, Oussama Elbaraghi^b,
Francois Bourzeix^b, Yahya Zennayi^b, Ilyass Maafa^c, Abdelghani Chehbouni^d

^a Plant and Microbial Biotechnology Center, Moroccan Foundation for Advanced Science, Innovation and Research (MAScIR), Mohammed VI Polytechnic University, Benguerir, Morocco

^b Analytics-Lab (A-LAB), Mohammed VI Polytechnic University (UM6P), Rabat, Morocco

^c Laboratory of Entomology and Phytopathology, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat, Morocco

^d Center for Remote Sensing Applications (CRSA), Mohammed VI Polytechnic University (UM6P), Benguerir, Morocco

* Corresponding author.

E-mail addresses: mohamed.aram@um6p.ma (M. Arame).

ABSTRACT

Fungal diseases such as Ascochyta pose major threats to chickpea production, causing significant losses if not detected early. Conventional diagnostic methods, including visual inspection and molecular assays, are often time-consuming, subjective, and ineffective for early detection of infection. This study investigates the use of hyperspectral imaging (HSI) combined with machine learning for early, non-destructive detection of Ascochyta blight in chickpea leaves, an application that remains underexplored in previous research. Hyperspectral data in the 400–1000 nm range were acquired under controlled laboratory conditions from artificially infected chickpea plants. In this study, we developed a new comprehensive processing pipeline to address critical challenges associated with hyperspectral data, including noise, artifacts, and illumination variations. Subsequently, unsupervised learning approaches, such as K-means clustering, were employed to construct a clean, well-labeled database of mean leaf spectra. Using this refined dataset, we evaluated a classification framework based on supervised learning models, leveraging selected vegetation indices, visible and infrared spectral bands, along with features derived from statistical analyses. The proposed approach achieved an overall classification accuracy exceeding 95% in distinguishing healthy chickpea plants from those infected with Ascochyta blight. Results demonstrate that HSI can capture subtle physiological changes in leaves before visible symptoms appear, offering a reliable and scalable tool for precision agriculture. This study contributes a promising step toward AI-powered early disease detection in chickpea farming, enabling timely interventions, reducing fungicide use, and supporting sustainable crop protection strategies. Future work will focus on real-world deployment and cost-effective integration into existing monitoring systems.

Keywords: Chickpea, Ascochyta blight, Hyperspectral imaging, Early disease detection, Machine learning, Precision agriculture.

Introduction

Chickpea (*Cicer arietinum* L.) is the third most important pulse crop globally, cultivated extensively across arid and semi-arid regions, particularly in India, Australia, Turkey, and North Africa. According to the Food and Agriculture Organization (FAO, 2023), global chickpea production exceeded 18.1 million tons in 2022, with Morocco contributing significantly as a key producer in North Africa (Basiru et al. 2025). However, chickpea production is highly vulnerable to biotic stressors, notably fungal infections. Previous research by Jha et al. (2023) highlighted that the susceptibility of local chickpea varieties directly influences yield loss under pathogen pressure, underscoring the need for cultivar-specific disease monitoring tools.

Fungal pathogens, particularly *Ascochyta rabiei* (causing *Ascochyta* blight), is among the most destructive to chickpea plants. *Ascochyta* blight is spread via infected seeds and airborne spores, rapidly infecting leaves, stems, and pods, especially under wet and humid conditions (Ilyas et al. 2022, Bayraktar et al. 2016). Symptoms typically appear 7–30 days post-infection. Importantly, the asymptomatic period, before visible symptoms manifest, is critical for early intervention and recent studies suggest that biotic stress from fungal diseases weakens plant immunity, increasing susceptibility to secondary pest infestations such as leaf miners (Vasanth-Srinivasan et al. 2025), thus aggravating overall plant health and yield outcomes.

Traditional fungal disease detection relies on visual symptoms or molecular diagnostics, which are often delayed until disease progression is advanced. RGB-based image analysis and convolutional neural networks (CNNs) have been applied for plant disease detection (Mohanty et al. 2016), but these techniques are limited by their reliance on visible symptoms and sensitivity to background noise. Consequently, they are inadequate for early-stage detection, especially during the latent period. Hyperspectral imaging (HSI), with its ability to record spectral reflectance data across hundreds of narrow bands, offers a compelling alternative. It

can detect subtle biochemical and physiological changes in plant tissues that precede visible symptoms, making it ideal for early disease diagnosis (González et al. 2015, Saeidan et al. 2025). Recent advances have shown promising results for HSI in detecting pathogens in crops such as sugar beet, grapevine, and tomato (Mahlein et al. 2019, Rumpf et al. 2010), but its application to chickpea remains underexplored.

Over the past decade, HSI has evolved from a lab-based tool into a powerful agricultural diagnostic technique, capable of detecting plant stress responses with high spatial and spectral resolution. With spectral coverage from the visible to the near-infrared region (typically 400–1000 nm), HSI captures detailed reflectance variations linked to changes in pigment concentration, water content, and cellular structure which is a key indicator of stress (Mahlein 2016, Arame et al. 2025, Pfordt and Paulus 2025). In disease detection, studies have successfully used HSI to detect early infections of downy mildew in grapevines by Lacotte et al. (2022) and powdery mildew in wheat by Khan et al. (2021). These advancements demonstrate the potential of HSI in supporting sustainable agricultural practices by reducing dependence on chemical inputs and enabling targeted interventions.

This work investigates the feasibility of leveraging advanced remote sensing specifically hyperspectral imaging (HSI) in combination with proven machine learning algorithms for the early detection of *Ascochyta* fungal diseases in chickpea crops. Our objective is to establish a rapid, non-destructive, and accurate diagnostic framework that empowers timely disease management decisions. Conventional disease scouting methods, visual leaf inspections, and laboratory assays are slow, laborious, and often miss asymptomatic infections, delaying interventions and risking major yield losses. Such shortcomings underscore the urgent need for more efficient and sensitive detection tools deployable under real field conditions. While HSI has shown promise in detecting stresses in other crops, most of the prior work has targeted late-stage symptoms. In contrast, this study directly evaluates how HSI coupled with algorithms like

Random Forests and Support Vector Machines can discriminate healthy and *Ascochyta*-infected plants under artificial infection regimes. By capturing subtle spectral alterations in the 400–1000 nm range before visual signs emerge, our approach aims to fill a critical gap in chickpea disease monitoring; it lies in the underexplored application of hyperspectral imaging combined with machine learning for the early, non-destructive detection of *Ascochyta* blight in chickpea leaves, together with the development of a comprehensive processing pipeline designed to address key challenges of hyperspectral data, including noise, artifacts, and illumination variability, paving the way for field-deployable, precision agriculture solutions that minimize fungicide use and maximize crop protection.

Materials and Methods

Plant material and experimental design

This study experiment was structured to enable the early detection of *Ascochyta* blight during the vegetative stage of chickpea, targeting a critical phase of the pathogen colonization and infection progression. Among the most widely cultivated chickpea cultivars in Morocco, Arifi (FLIP84-78C), a Moroccan cultivar released in 2007, and Garbanzo, a commercially prominent cultivar known for its superior market quality, were selected for their contrasting resistance profiles to *Ascochyta* blight (Table 1) (Houasli et al. 2021, Bencheqroun et al. 2021).

Table 1 Characteristics of chickpea cultivars are included in the study.

Cultivar	Type	Traits
Arifi (FLIP98-50 C)	Kabuli	Erect plant, large seed size, moderately resistant to <i>Ascochyta</i> blight
Garbanzo	Kabuli	Extra-large seeds, highly susceptible to <i>Ascochyta</i> blight

The experimental design was a split-plot layout, with inoculation status (inoculated vs. non-inoculated [healthy]) as the main plot factor and chickpea cultivars as the subplot factor. A

total of 12 pots were used, with six allocated to the inoculated treatment and six to the non-inoculated pots, which served as the control. Each treatment subgroup included three replicates per cultivar, with each pot containing five seedlings. All pots were maintained under uniform controlled conditions: temperature ($22 \pm 2^{\circ}\text{C}$), photoperiod (16 h light/8 h dark), and watering regime to effectively isolate spectral signals related to pathogen infection. This design facilitated an efficient comparison of cultivar responses under uniform controlled conditions while enabling the assessment of treatment effects (Hartung et al. 2019, Grewal et al. 2009).

Inoculum Production and Inoculation

Five monoconidial, aggressive isolates of *Ascochyta rabiei* obtained from the International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat, Morocco, were cultured on potato dextrose agar (PDA; Sigma-Aldrich, Germany) and incubated at $22 \pm 1^{\circ}\text{C}$ for 8 days. Pycnidiospores were harvested by scraping the culture surface with sterile distilled water (SDW), followed by filtration through two layers of sterile cheesecloth. The resulting suspension was adjusted to a concentration of 1×10^6 conidia/mL using a hemocytometer. To enhance spore adhesion during application, 0.05% Tween 20 (Sigma-Aldrich, Germany) was added to the inoculum. Plants were uniformly sprayed on both the adaxial and abaxial leaf surfaces, as well as on the stems, using a handheld spray.

Data collection

Three *Ascochyta* blight-inoculated pots and three healthy-control pots per cultivar were transferred daily to the hyperspectral imaging laboratory throughout the asymptomatic phase, before visible lesions appeared. This allowed continuous monitoring of spectral changes from initial colonization through symptom onset. Because *Ascochyta* lesions typically emerge on leaves and occasionally on stems at latencies influenced by cultivar resistance and environmental factors (Thomas et al. 2018, Pena-García et al. 2023, Pérez-Pizá et al. 2023), this

daily sampling strategy ensured capture of all critical spectral windows. Plant health was assessed daily via expert visual inspection to confirm the exact timing of symptom emergence. This two-step framework, comprising artificial inoculation followed by systematic hyperspectral acquisition, provides a robust foundation for correlating spectral signatures with early phases of *Ascochyta* blight infection in the chickpea vegetative stage (Fig. 1).

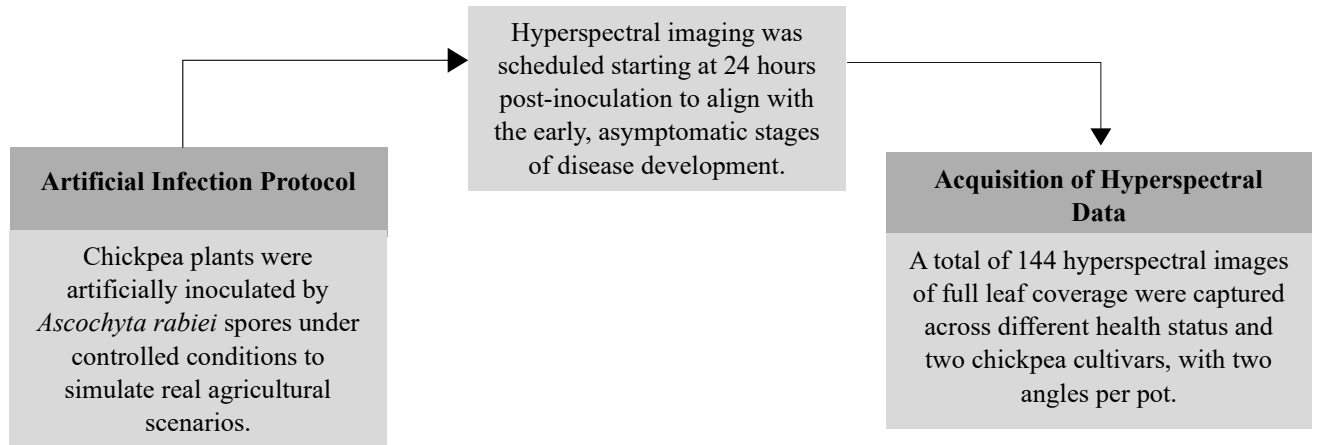


Fig. 1. Data collection diagram.

The imaging system used in our study was the Specim FX10 hyperspectral camera (Specim, Oulu, Finland), depicted in Fig. 2. This line-scan camera operates in the visible and near-infrared (VNIR) region, covering a spectral range from 400 to 1000 nm, and is equipped with a high spatial resolution of 1024 pixels. It captures 448 spectral bands with a spectral resolution of 5.5 nm, enabling the detection of subtle spectral variations. The camera achieves an imaging speed of up to 330 frames per second at full resolution, facilitating rapid data acquisition. To ensure stability during imaging, the camera was securely mounted on a rigid frame. The setup incorporated a laboratory scanner to facilitate precise movement and accurate imaging. Uniform illumination was achieved using halogen lamps, which provided consistent lighting and minimized shadowing effects. Before every imaging session, the system was calibrated by using a white reference panel for obtaining valid reflectance measurements. However, it is important to note that as the spectral range of the Specim FX10

is limited to 400–1000 nm, which excludes the shortwave infrared (SWIR) region beyond 1000 nm. This limitation may impact the detection of certain vegetation-associated spectral features, such as water content and lignin-cellulose absorption bands, which are prominent in the SWIR region. Consequently, while the camera is effective for capturing the visible and near-infrared (VNIR) plant characteristics, its inability to detect SWIR features may limit comprehensive vegetation analysis.



Fig. 2. The experiment site at ICARDA in Rabat, Morocco, using hyperspectral camera measuring a line of 1024 pixels with 448 spectral bands on chickpea plant pot.

Hyperspectral imaging was performed on both Arifi and Garbanzo cultivars across two health statuses, healthy and *Ascochyta*-infected. Initially, data were acquired from 1-day post-infection (dpi) up to 25 dpi, covering the full progression until symptom appearance. For analysis, however, the data were grouped into three representative periods: 2 and 4 dpi, 15 and 17 dpi, and 22 and 24 dpi. This grouping was chosen to reduce redundancy, capture key stages of disease development (early, intermediate, and late), and ensure manageable, statistically robust datasets for model training. In total, 144 images were acquired (72 from infected plants and 72 from healthy controls). For each status and cultivar, three pots were imaged twice per session, with the camera repositioned to capture two distinct spectral viewing angles that encompassed all leaf surfaces. This protocol generated six images per status per cultivar per

day (2 images $\times$ 3 pots), producing consistent temporal coverage of disease progression under identical illumination and environmental conditions (9:00–15:00 GMT). After each imaging session, infected pots were returned to the humid chamber to maintain pathogen development, while healthy controls remained in the non-infested chamber. Table 2 summarizes this distribution: three infected pots and three healthy control pots per cultivar, each imaged twice per day resulting in six Infected and six healthy, yielding a balanced dataset for robust statistical comparison. By synchronizing the imaging of infected and control plants at each time point, our design ensures that any spectral differences reflect true physiological responses to *Ascochyta* infection rather than extraneous variability in plant orientation or environmental factors.

Table 2 Summary of hyperspectral images (HSI) size by chickpea health status for each cultivar by 3 pots during seedling stage.

	Data Measurement Time (GMT Central Time)	Days post infection	Arifi Pots	Garbanzo Pots	Total
Ascochyta blight inoculated pots and healthy control	09 June 2025 9:00–15:00	2 days	6 Infected and 6 healthy	6 Infected and 6 healthy	72 Infected and 72 healthy
	11 June 2025 9:00–15:00	4 days	6 Infected and 6 healthy	6 Infected and 6 healthy	
	22 June 2025 9:00–15:00	15 days	6 Infected and 6 healthy	6 Infected and 6 healthy	
	24 June 2025 9:00–15:00	17 days	6 Infected and 6 healthy	6 Infected and 6 healthy	
	29 June 2025 9:00–15:00	22 days	6 Infected and 6 healthy	6 Infected and 6 healthy	
	31 June 2025 9:00–15:00	24 days	6 Infected and 6 healthy	6 Infected and 6 healthy	
Total			36 Infected and 36 healthy	36 Infected and 36 healthy	72 Infected and 72 healthy

Of the 144 hyperspectral images, 96 were allocated to the training set and 48 to the test set to ensure balanced representation of each infection status and cultivar. For training, we selected four images per day from each of the first two infected pots and same from the healthy control pot of each cultivar, resulting in 8 images per cultivar over the six days (4 images of 2 infected pots + 4 healthy images = 8 images $\times$ 2 cultivars $\times$ 6 days = 96, Fig. 3). The remaining images comprising the third infected pot and one healthy control image per cultivar on each day (2 images $\times$ 2 pots $\times$ 2 cultivars $\times$ 6 days = 48), formed the test set (Table 3). This split maintains identical class and varietal proportions between training and testing, ensuring that each health status and cultivar is equally represented. By holding out an entire pot per status for testing, we guard against overfitting to a single plant's spectral signature and provide a robust evaluation of model generalization to unseen samples. The limited spectral variability inherent to hyperspectral data further amplifies the importance of this strategy in preventing bias.

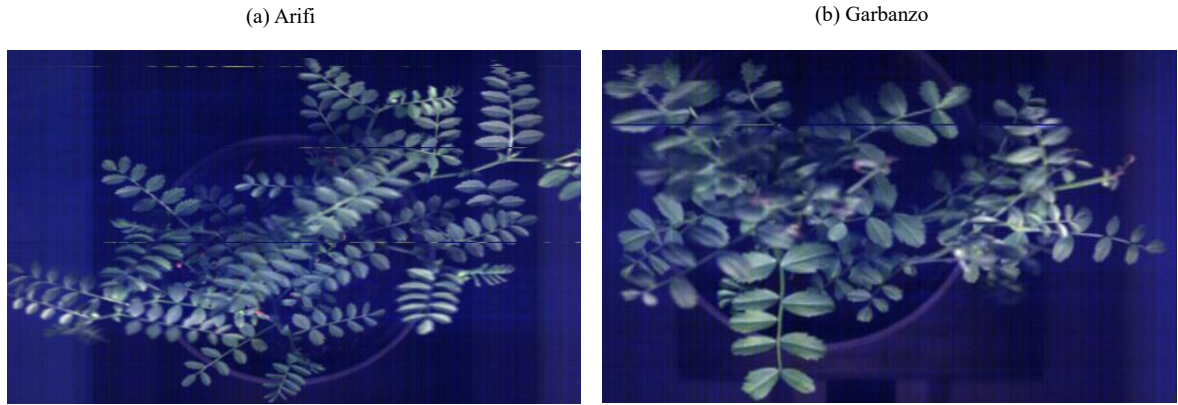


Fig. 3. Red-green-blue (RGB) hyperspectral images example for training set. (a) Arifi sample. (b) Garbanzo sample.

Table 3 Summary of hyperspectral images (HSI) size by chickpea health status for train set and test set during vegetative stage.

Status	Arifi Pots		Garbanzo Pots		Total
	Train Set (x 6 days post infection)	Test Set (x 6 days post infection)	Train Set (x 6 days post infection)	Test Set (x 6 days post infection)	
Inoculated	4	2	4	2	12
Non-Inoculated (Healthy)	4	2	4	2	12
Total	8	4	8	4	144

Methods for Image Preprocessing, Feature Extraction, and Model Development

Image preprocessing

This study employed a new developed pipeline of structured sequence of hyperspectral data preprocessing steps, as illustrated in Fig. 4, to ensure the extraction of reliable and high-quality spectral features for machine learning-based *Ascochyta* blight detection in chickpea plants. Preprocessing is a critical phase in hyperspectral imaging workflows, as raw hyperspectral images are often affected by various sources of noise, illumination inconsistencies, and sensor artifacts that can compromise analytical accuracy (Gao et al. 2020). Techniques such as dark current and white reference correction, spectral smoothing, background segmentation, and normalization were systematically applied to transform the raw spectral data into a clean, standardized format. These steps not only reduce irrelevant variability but also enhance the discriminative power of the spectral information associated with plant physiological changes due to *Ascochyta* infections. Previous studies have highlighted that robust preprocessing pipelines significantly improve classification accuracy and model generalization in agricultural hyperspectral applications (Behmann et al. 2015). By adopting these proven methodologies, this study ensures the reliability and reproducibility of downstream disease classification processes.

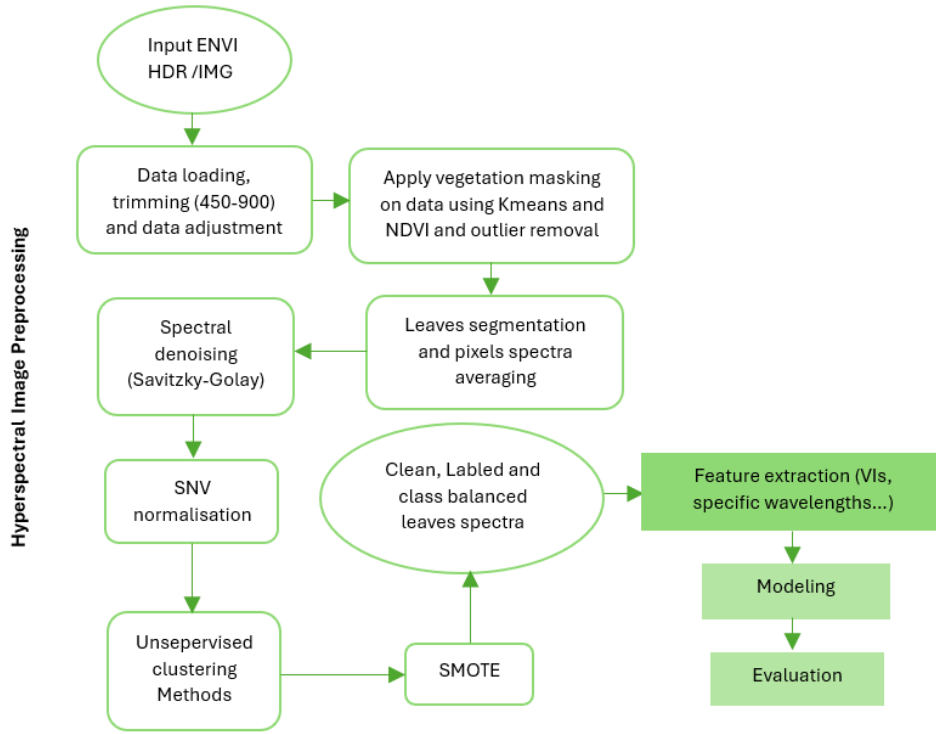


Fig. 4. The overall workflow diagram of data processing and modeling pipeline.

Hyperspectral images were acquired in ENVI format (. HDR/IMG) and imported into the analysis environment. The initial preprocessing steps involved data trimming, where the spectral range was reduced from 400–1000 nm to 450–900 nm to eliminate noisy extremes. Subsequently, data adjustment was performed by replacing negative and infinite sample values with the mean reflectance of the corresponding spectral band. A binary vegetation mask was generated using the Normalized Difference Vegetation Index (NDVI) as illustrated in Fig. 5, calculated as:

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}) \quad (1)$$

where NIR and Red are surface reflectance in the near-infrared and red bands, respectively (Tucker 1979). NDVI thresholds were empirically selected to retain green, photosynthetically active plant tissue while excluding soil, pot, and background. Kmeans clustering method also is used in addition to NDVI, especially for the first two stages of chickpea plants post infections, to guarantee more precision of background removal. This masking is critical for ensuring that

the subsequent spectral analysis focuses solely on leaf tissue, which reflects the physiological status of plant (Zarco-Tejada et al. 2005). Then, spectral outliers were identified and removed. This step involved analyzing the reflectance distribution across all leaf pixels and excluding spectra that deviated beyond ± 2 standard deviations from the mean, calculated as:

$$|R_i - \mu| > 2\sigma \quad (2)$$

where R_i is the reflectance of pixel i , μ is the mean reflectance, and σ is the standard deviation. Outliers may result from sensor noise, specular reflections, or the inadvertent inclusion of soil spectra instead of vegetation, all of which can distort average spectral profiles if not addressed (Kohavi and John 1997, Mahlein et al. 2012).

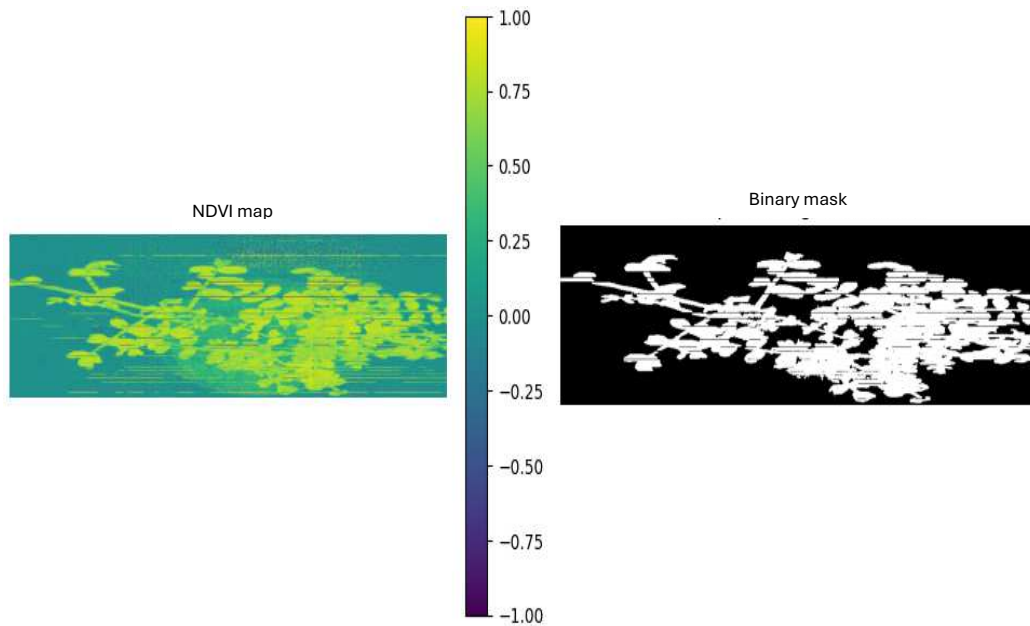


Fig. 5. Applied NDVI on a HSI sample

A key methodological choice in this study was the leaf-level spectral analysis approach. Following vegetation masking, A watershed segmentation algorithm was employed to isolate individual leaves within each hyperspectral image (Fig. 6). Compared to other traditional segmentation approaches, watershed was selected because it provided the most accurate separation of adjacent leaves and minimized misclassification of background regions (Gehan et

al. 2017). However, when leaf boundaries were weak or overlapping, the algorithm sometimes produced region-level segments rather than perfectly defined leaf objects. The watershed method treats the NDVI image as a topographic surface, where bright areas represent ridges and dark areas represent valleys. Segmentation is achieved by identifying catchment basins based on changes in image gradients. Mathematically, the watershed regions W are obtained as:

$$W = \text{Watershed}(\|\nabla I\|) \quad (3)$$

where I is the NDVI image and $\|\nabla I\|$ is its gradient magnitude. Once segmented, the average reflectance spectrum for each leaf was calculated by averaging the reflectance across all pixels within the segmented area, across the entire spectral range of 450–900 nm. This strategy yields one spectral signature per leaf, which greatly reduces the impact of pixel-level noise and enables robust, interpretable comparisons between leaves of different infection statuses. Additionally, this aligns with recent findings that disease-related changes manifest at the organ level (Mahlein et al. 2013). By averaging spectra per leaf, inter-pixel variability is minimized, and more stable features are provided to the classifiers downstream.

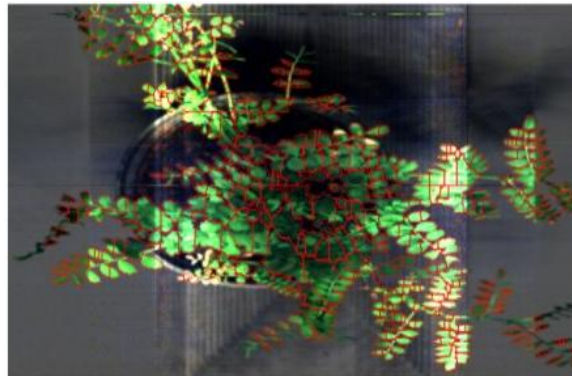


Fig. 6. Applied watershed leaves segmentation on a HSI sample (442 segments)

Additionally, to reduce random noise while preserving the integrity of plant features, spectral denoising was applied, using the Savitzky-Golay filter, which smooths the leaf reflectance curves by fitting successive subsets of adjacent data points with a low-degree polynomial via linear least squares (Savitzky and Golay 1964). This method retains peak shapes

and fine details of the spectral signature, which are crucial for detecting infection-related stress. Finally, to reduce multiplicative scatter effects, improve comparability between spectrum and create a clean leave spectrum, Standard Normal Variate (SNV) normalization was applied (Barnes et al. 1989). In this method, each spectrum is standardized by subtracting its mean and dividing by its standard deviation, expressed as:

$$\text{SNV}(R_i) = \frac{R_i - \mu}{\sigma} \quad (4)$$

where R_i is the reflectance at wavelength i , μ is the mean reflectance of the spectrum, and σ is its standard deviation. SNV is particularly effective for leaf-level hyperspectral data, where variations in leaf thickness, surface orientation, or illumination can introduce variability unrelated to plant health (Rinnan et al. 2009).

After completing the preprocessing and cleaning procedures, some hyperspectral images of the infected pots, where leaf spectra were labeled as infected still contained regions of non-symptomatic and healthy tissue. This outcome is expected during the symptomatic stage, as symptoms often emerge only on a subset of leaves within an infected pot. This mislabeling could introduce noise into the dataset and reduce the reliability of the classification task. To address this issue, unsupervised clustering methods, including DBSCAN, HDBSCAN, and K-means, were employed to extract distinct clusters within the infected images. Each method was evaluated in terms of its ability to discriminate between subtle spectral variations. Among them, K-means clustering was identified as the most appropriate technique (Fig. 7) due to its robustness, simplicity, and compatibility with spectral feature distributions. Based on this clustering analysis and on literature-supported interpretations of chickpea spectral signatures (Singh et al. 2021), the datasets were relabeled to improve class assignment accuracy. Fig. 7 illustrates two distinct groups ($k = 2$) of leaf spectra, which differ primarily in the near-infrared region (700 to 950 nm): the supposed “infected” cluster (blue curve) shows a marked decline

in reflectance, whereas the supposed “healthy” cluster (orange curve) maintains a relatively stable spectral pattern. In line with established disease-progression dynamics, we hypothesized that the proportion of infected leaves within each infected plant increases from the early to the final assessment stage for both cultivars, with Grbz exhibiting earlier and more pronounced sensitivity to *Ascochyta* infection compared to Arifi. This hypothesis is further examined and confirmed in the Results section.

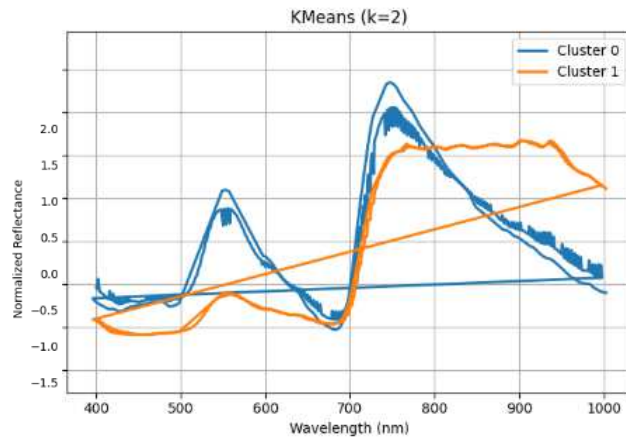


Fig. 7. Best clustering results using KMeans

Thus, a major challenge was created; the dataset was characterized by a strong class imbalance, as the number of infected samples was significantly lower than that of healthy ones. To overcome this limitation, the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE is a widely used resampling strategy that generates new synthetic samples for the minority class by interpolating between existing neighbors in the feature space, thereby improving the class distribution while preserving the intrinsic variability of the spectral data (Chawla et al. 2002, Douzas and Bacao 2018). The final hyperspectral dataset in this study consisted of approximately 17,000 leaf spectra for training and 7,500 spectra for testing, equally distributed between “healthy” and *Ascochyta*-infected or “stressed” classes.

These refined and balanced spectra then served as the foundation for the next stage: feature extraction, where key spectral indicators such as vegetation indices and stress-related

wavelength regions were identified. The extracted features were subsequently used to train machine learning models for classifying plant health status, and the models were evaluated to assess their accuracy, robustness, and applicability for the early detection of *Ascochyta* blight in chickpea plants.

Feature extraction

In this study, a set of spectral and statistical features was selected based on literature evidence demonstrating their effectiveness in discriminating against healthy and diseased plants. Several vegetation indices (VIs) known to be sensitive to pigment content, structural changes, and stress responses were considered, including the Normalized Difference Vegetation Index (NDVI), Red-Edge Position (REP), Modified Red-Edge NDVI (mRENDVI), Photochemical Reflectance Index (PRI), and Anthocyanin Reflectance Index (ARI). Their definitions, equations, and references are summarized in Table 4.

Table 4 Vegetation indices (VIs) used for classification of chickpea plant health.

Vegetation Index (VI)	Acronym	Equation	References
Normalized Difference Vegetation Index	NDVI	$NDVI = (NIR - Red) / (NIR + Red)$	(Beisel et al. 2018, Mohd Johari et al. 2025)
Red-Edge Position	REP	$\lambda_{red-edge}$, calculated using linear interpolation between reflectance values at 670, 700, 740 nm	(Goswami et al. 2021)
Modified Red-Edge NDVI	mRENDVI	$(R750 - R705) / (R750 + R705 - 2 \times R445)$	(Sims and Gamon 2002)
Photochemical Reflectance Index	PRI	$(R531 - R570) / (R531 + R570)$	(Zarco-Tejada et al. 2013)
Anthocyanin Reflectance Index	ARI	$(1/R550) - (1/R700)$	(Roberts et al. 2018)

In addition to VIs, specific narrow spectral bands (700 nm, 763 nm, 770 nm, and 884 nm) were selected, as previous studies have consistently reported their importance for disease detection, particularly due to chlorophyll absorption at 540–750 nm (Carter 1994), red-edge shifts around 763–770 nm (Moshou et al. 2005, Li et al. 2021), and structural or water-related changes near 884 nm (Li et al. 2021). Furthermore, higher-order statistical and derivative-based

features were extracted from leaf-level spectra to capture distributional characteristics. These include mean reflectance, which summarizes the overall spectral profile of each leaf and has been reported as a strong indicator of disease-related changes (Bravo et al. 2003); skewness and kurtosis, which describe the asymmetry and peakedness of spectral distributions and are sensitive to infection-induced reflectance variations (Baranowski et al. 2015); and the red-edge slope, derived from reflectance between 680–750 nm, which is widely recognized as a reliable proxy for chlorophyll degradation under stress (Leucker et al. 2016, Horler et al. 1983). Additionally, first-derivative spectra were used to compute the derivative mean and derivative standard deviation, which enhance subtle inflection points and variability not captured in raw spectra, and have been shown to significantly improve disease detection accuracy (Liang et al. 2015). Collectively, these features provide a comprehensive description of both global and localized spectral variations, thereby strengthening the discrimination between healthy and *Ascochyta*-infected chickpea leaves.

Model Selection, Training, and Evaluation

In this study, unsupervised clustering and supervised classification were applied consecutively to meet distinct but complementary objectives arising from the nature of the experimental data and the final application target. First, unsupervised clustering (k-means) was used as a preprocessing and relabeling step to mitigate label noise in the initial dataset. However, clustering alone is not sufficient to achieve the objectives of this study. Unsupervised methods such as k-means do not provide an explicit decision function, probabilistic confidence, or mechanism for evaluating generalization on unobserved data. Furthermore, cluster assignments are specific to the dataset and cannot be directly transferred to new acquisitions, especially under different environmental conditions or sensor configurations. For this reason, supervised classification models were then trained on the relabeled dataset to learn stable decision boundaries between healthy and infected spectral responses. Supervised models offer

several advantages that are essential to the goal of this work: (i) quantitative evaluation using standard performance metrics (precision, recall, F1 score), (ii) explicit evaluation of generalization on independent test data, and (iii) applicability to new and unobserved data, which is essential for future field deployment using multispectral sensors.

For the classification stage, several machine learning algorithms were selected to evaluate their ability to discriminate between healthy and *Ascochyta*-infected (stressed) chickpea leaves. The chosen models included Gradient Boosting (GB), Support Vector Machine with Radial Basis Function kernel (SVM-RBF), Random Forest (RF), k-Nearest Neighbors (k-NN) and Logistic Regression (LR). These models were selected based on their proven effectiveness in hyperspectral plant disease detection tasks: ensemble models such as GB and RF are well-known for handling high-dimensional and nonlinear spectral data, SVM-RBF has demonstrated strong performance in disease detection due to its capacity to separate complex boundaries in hyperspectral space (Rumpf et al. 2010), while SVM-RBF, k-NN, and LR provide interpretable baselines and allow comparisons with more advanced classifiers (Cover and Hart 1967, Wijesinghe et al. 2025). Hyperparameter optimization was conducted using a grid search strategy combined with 5-fold cross-validation to avoid overfitting and ensure the selection of optimal parameters across all models (Bergstra and Bengio 2012). Following training, the models were tested on an independent dataset, and their performance was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score, which provide a comprehensive understanding of both class balance and error distribution. Among the tested algorithms, Gradient Boosting emerged as the most effective model, achieving the highest F1-score and recall, thereby indicating its robustness in detecting infected samples at an early stage. To further assess its generalization capacity, the best-performing model was validated on newly acquired chickpea datasets. The predictions were mapped back onto the hyperspectral images at the leaf level, enabling spatial visualization of infection patterns across

individual leaves and confirming the model's capacity to not only classify but also localize disease symptoms. This step demonstrated the practical applicability of hyperspectral imaging combined with machine learning for real-time monitoring of chickpea health status.

Results

Data analysis

A data analysis has been carried out focusing on the interpretation of hyper-spectral signatures to detect early physiological changes in chickpea plants following infection by *Ascochyta*, in the asymptomatic stage (Fig. 8). We will focus specifically on the near-infrared (NIR) spectral region, as this region is well known for its sensitivity to alterations in internal leaf structure, water content and overall plant vigor, which are key indicators of physiological stress (Li et al. 2023). The spectral range of interest for this detailed examination extends roughly from spectral wavelength 700 nm to 950 nm, where subtle changes related to plant health are often prominently expressed in reflectance data captured by hyper-spectral imaging.

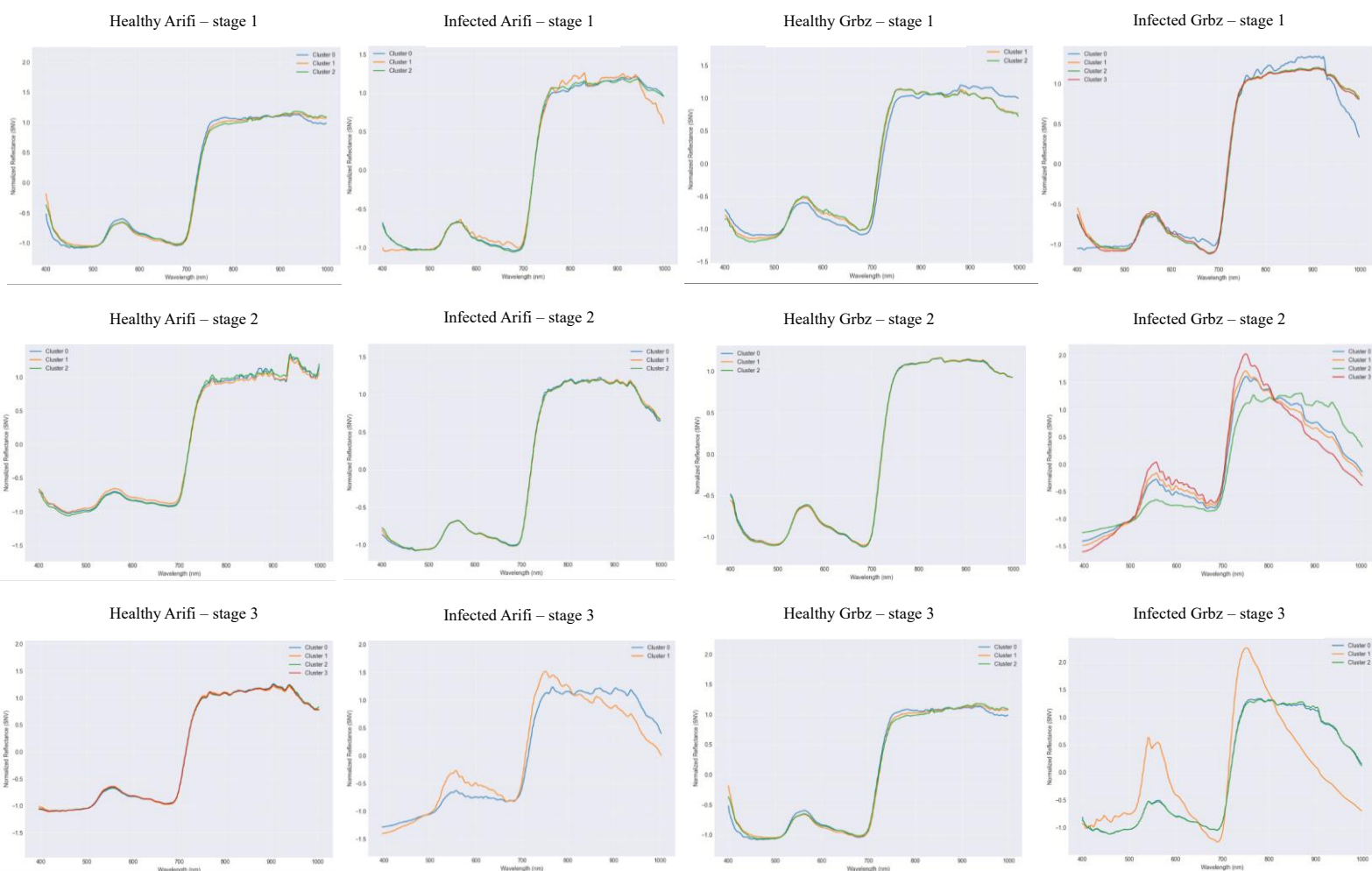


Fig. 8. Distribution of hyperspectral data clusters using K-Means for healthy and *Ascochyta*-infected chickpea leaves by stage and cultivar

The spectral profile of chickpea leaves across different varieties (Arifi and Garbanzo) and developmental stages reveal clear differences between healthy and infected samples, particularly in the near-infrared (NIR) region. In agreement with previous findings (Li et al. 2023, Grishina et al. 2024), healthy leaves exhibited stable reflectance in the NIR domain (750–900 nm), whereas infected leaves consistently showed a decrease in NIR reflectance, reflecting structural and physiological degradation associated with *Ascochyta* blight infection. This confirms the first hypothesis that healthy spectra remain stable in the NIR, while infection induces a decline in reflectance.

When comparing the two cultivars, the Garbanzo (Grbz) displayed more pronounced spectral changes in the presence of infection compared to Arifi. At stage 1, both varieties exhibited relatively similar spectra between healthy and infected leaves, with only subtle differences in NIR reflectance, suggesting that disease onset begins but is not yet strongly manifested. By stage 2, infected Grbz leaves demonstrated a clearer separation from their healthy counterparts, with a marked decline in NIR reflectance, while Arifi leaves showed only moderate changes, indicating partial resistance. At stage 3, the divergence became most evident: infected Grbz spectra showed substantial NIR reflectance decline and greater variability across clusters, suggesting extensive leaf tissue damage. In contrast, Arifi spectra, while affected, remained comparatively more stable, supporting the second hypothesis that Arifi is more resistant to *Ascochyta* blight than Grbz.

In terms of disease progression, the results suggest that *Ascochyta* infection begins to significantly alter spectral signatures from stage 2 onward. Table 5 summarizes the approximate proportion of healthy and infected leaves observed in the two chickpea cultivars across the disease-assessment stages. These proportions were derived from the two clusters identified by the k-means algorithm, which grouped all spectral observations at each stage and for each cultivar into two dominant spectral classes. Based on a comparison of the mean spectral signatures of these clusters with reference spectral patterns reported in the literature for healthy and stressed chickpea leaves, each cluster was subsequently interpreted as representing either a healthy or an infected physiological status. Owing to the complexity and limited accuracy of tracking individual diseased leaf areas under laboratory conditions, direct visual ground-truthing at the leaf-zone level was not feasible. Consequently, this clustering-based approach provided a consistent and data-driven alternative for estimating infection proportions across cultivars and stages. As expected, the proportion of infected leaves within each infected plant increases from the early to the final assessment stage for both cultivars, with Grbz consistently

exhibiting a higher proportion of infection compared to Arifi at each stage, reflecting its earlier onset of symptoms and greater sensitivity to *Ascochyta* blight. The clear and progressive shift in cluster membership across stages and cultivars, as reported in Table 5, supports our initial hypothesis on disease-progression dynamics and further aligns with the spectral differences reported in the clustering analysis.

Table 5 Approximate Proportion of Healthy and Infected Leaves per Cultivar Across Stages.

Stage	Arifi Leaves (%)		Garbanzo Leaves (%)	
	Healthy	Infected	Healthy	Infected
Stage 1 (Early)	0	0	0	5
Stage 2 (Intermediate)	0	22	0	41
Stage 3 (Late)	0	55	0	70

Overall, these findings confirm that *Ascochyta* blight significantly impacts spectral behavior in the NIR region, with cultivar-dependent differences in resistance. Grbz appears more sensitive and rapidly affected, while Arifi demonstrates relative resilience, delaying the spectral degradation associated with infection. These observations provide a crucial foundation for the subsequent machine learning classification analysis, where the capacity of different algorithms to discriminate between healthy and infected leaves across varieties and stages will be explored.

Classification of chickpea plant status

To evaluate the capacity of hyperspectral data to discriminate between healthy and *Ascochyta*-infected chickpea leaves, several machine learning models were trained and tested, including Gradient Boosting, SVM with RBF kernel, Random Forest and Logistic Regression. Models were trained separately for each cultivar (Arifi and Grbz) and at each developmental stage (1–3), as well as on combined datasets across all stages (Fig. 9). Hyperparameter optimization was performed through grid search with 5-fold cross-validation. To account for the critical need to minimize false negatives in disease detection (i.e., misclassifying infected leaves as healthy), model selection prioritized recall and F1-score as suggested by recent works

in hyperspectral plant pathology (Lin et al. 2024). For the evaluation across all stages, the best-performing model in terms of recall was selected and applied to test dataset to generate confusion matrices.

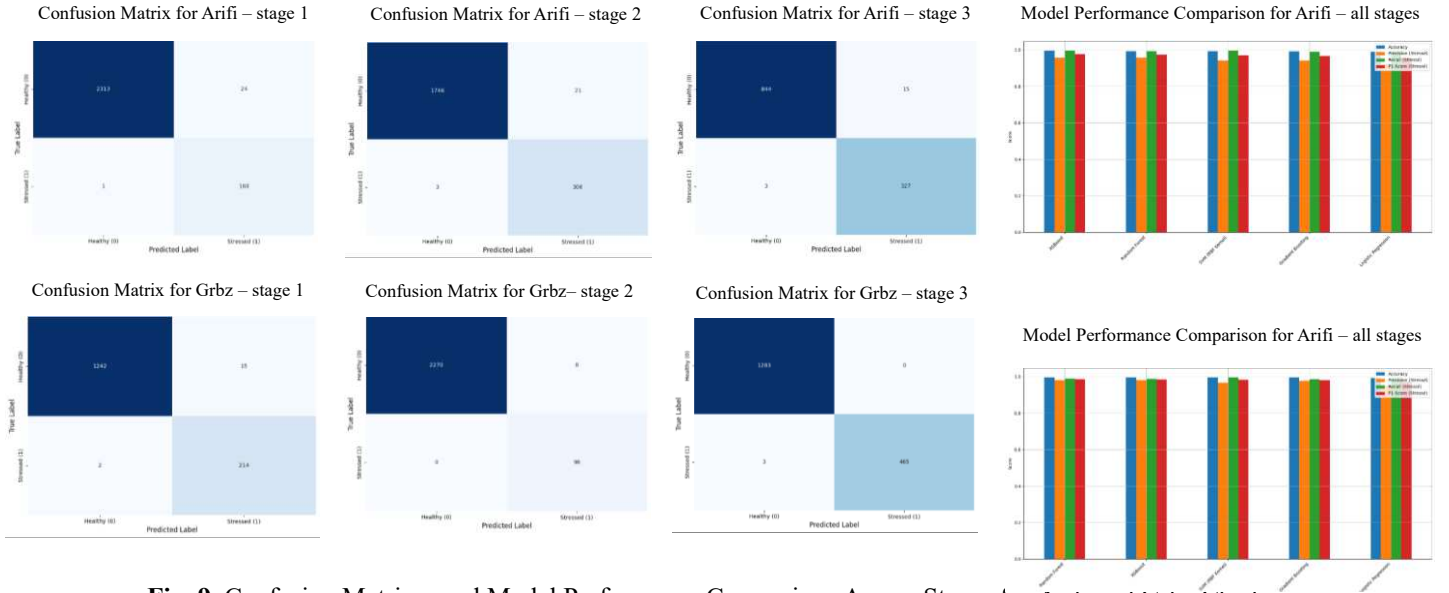


Fig. 9. Confusion Matrices and Model Performance Comparison Across Stages for Arifi and Grbz Chickpea Varieties in Healthy vs. Infected Classification

For the Arifi cultivar, the confusion matrices across stages 1 to 3 revealed that classification performance was consistently high, with most healthy and infected samples correctly identified. At stage 1, misclassifications were minimal, indicating that early infection symptoms in Arifi are moderately detectable. At stage 2, classification improved further with a higher proportion of infected leaves correctly identified, reflecting the increasing visibility of spectral changes. By stage 3, the model maintained robust performance, though some misclassifications remained consistent with the higher variability in infected spectra.

For the Grbz cultivar, the confusion matrices demonstrated slightly stronger performance than Arifi across all stages. At stage 1, classification was already highly accurate, exhibits earlier detectable spectral differences between healthy and infected leaves. At stages 2 and 3, the accuracy and recall remained consistently high, with very few false negatives, confirming the hypothesis that Grbz is more sensitive to the disease and therefore more easily classified.

When combining all stages per cultivar, the most effective model (Gradient Boosting) maintained high performance across precision, recall, and F1-score for both Arifi and Grbz. The confusion matrices for the combined datasets closely mirrored those of the individual stages, suggesting that the model generalized well across phenological stages. Importantly, the priority criterion of minimizing false negatives was achieved, with recall values remaining consistently high. Grbz again exhibited slightly better classification performance, with sharper separation between healthy and infected classes, consistent with its greater sensitivity to *Ascochyta* infection. Arifi, while still classified with high accuracy, showed a slightly higher rate of false negatives, which aligns with its partial resistance and more stable spectral response under infection stress.

To further strengthen the robustness of our approach, all data from both chickpea varieties (Arifi and Grbz) and across the three disease development stages were combined into a single dataset. This integration allowed the training and evaluation of models on a more diverse and representative sample set, simulating real-world scenarios where multiple varieties and phenological stages co-exist. The best-performing classifier was employed to distinguish between healthy and infected leaves (Fig. 10). Prioritizing recall is consistent with plant disease detection literature, where minimizing missed infections is critical for early intervention and crop protection.

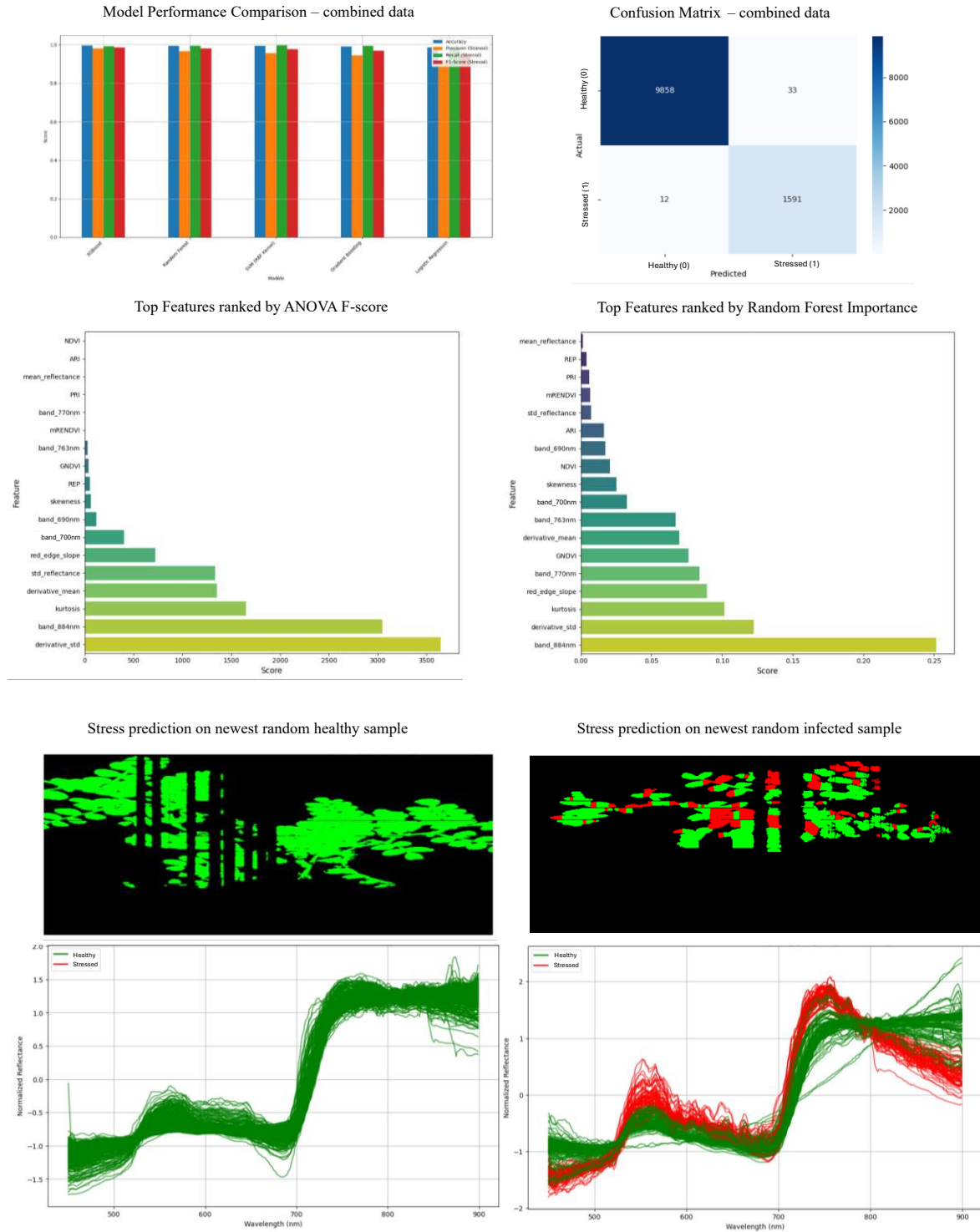


Fig. 10. Model performance, feature importance, and stress mapping results for combined chickpea varieties and stages.

The performance of the selected model on the combined dataset is illustrated in Fig. 10 (top row). Across all evaluation metrics, the model achieved consistently high scores, with precision, recall, and F1-score values exceeding 95% for both healthy and infected classes. The confusion matrix further reinforces these outcomes, showing 9803 true negatives and 1993 true

positives, with only 33 false positives and 12 false negatives. These results indicate not only strong overall performance but also, more importantly, a very low false-negative rate, which aligns with the primary objective of early disease detection where missing infected samples could have severe agronomic consequences. The balanced performance across classes demonstrates the robustness of the model in distinguishing between healthy and infected leaves even under combined cultivar and stage conditions.

Feature selection and ranking analyses were conducted to identify the most discriminative spectral and derived features distinguishing between healthy and infected classes. As presented in Fig. 10 (middle row), two complementary approaches were used: ANOVA F-score ranking and Random Forest feature importance. Both methods converged on a consistent set of highly informative features, including vegetation indices such as NDVI, REP, and PRI, as well as specific spectral bands (e.g., 700 nm, 770 nm, and 884 nm) and statistical descriptors (e.g., skewness, derivative mean, and kurtosis). Both methods highlighted `derivative_mean`, `red_edge_slope`, `NDVI`, `mean_reflectance`, and `skewness` as the most informative indicators of plant status, confirming their robustness across univariate and multivariate perspectives. ANOVA tended to emphasize features with strong individual discriminatory power, such as `mean_reflectance`, `NDVI`, and `red-edge slope`, while Random Forest additionally assigned high importance to specific narrow bands (e.g., 700 nm, 770–763 nm, and 884 nm) and higher-order features (e.g., `derivative_std` and `kurtosis`), reflecting their contribution through nonlinear interactions within the model.

Finally, the generalization capacity of the model was assessed on newly acquired, independent chickpea samples not included in the training process. As depicted in Fig. 10 (bottom row), the stress mapping correctly detected infection zones (highlighted in red) within infected leaves while maintaining a clean, stable mapping for healthy leaves. Corresponding spectral plots confirm the expected disease-related deviations in the near-infrared region,

consistent with known physiological stress signatures. These results demonstrate not only the accuracy of the classification model on controlled datasets but also its ability to extend to novel, unseen data, a critical step toward real-world field deployment.

Discussion

The results of this study demonstrate the effectiveness of hyperspectral imaging (HSI) combined with machine learning for the early detection of *Ascochyta* blight in chickpea leaves. The analysis of spectral signatures revealed that the near-infrared (NIR) domain, particularly between 750–900 nm, plays a decisive role in distinguishing between healthy and infected plants. This agrees with prior studies that have consistently identified NIR as a sensitive region for structural and physiological changes linked to plant stress (Mahlein 2016, Carter and Miller 1994). In line with our first hypothesis, healthy leaves exhibited stable NIR reflectance, whereas infected leaves showed a progressive decline, confirming that spectral degradation is a reliable proxy for infection-related tissue damage.

A key contribution of this work lies in the cultivar-specific responses observed at the early and asymptomatic stages. The Grbz cultivar consistently showed stronger spectral alterations under infection compared to Arifi, especially from stage 2 onward. These findings support the second hypothesis that Grbz is more sensitive, while Arifi displays partial resistance to *Ascochyta* blight. Such genotype-dependent differences are consistent with reports from crop phenotyping studies, where resistant cultivars maintain more stable spectral profiles under pathogen stress (Bauriegel and Herppich 2014, Thomas et al. 2018). The observed infection prevalence, reaching over 70% in Grbz leaves at stage 3 compared to ~55% in Arifi, further corroborates this differential susceptibility. This underscores the potential of hyperspectral imaging not only for accurate disease detection, but also for cultivar screening, resistance evaluation, and monitoring disease progression across developmental stages.

From a modeling perspective, the comparative evaluation of classifiers showed that ensemble-based approaches (e.g., Gradient Boosting and Random Forest) consistently achieved high accuracy and recall across varieties and stages above 90%. The prioritization of recall as the main selection criterion is critical, as false negatives (infected leaves misclassified as healthy) pose a serious risk in crop disease management (Liakos et al. 2018, Kamilaris and Prenafeta-Boldú 2018). In our case, the selected model achieved very low false-negative rates ($<0.5\%$), which is particularly encouraging for early detection applications where timely intervention is essential. The robustness of performance across both stages, specific and combined datasets indicates strong generalization ability, aligning with findings from recent hyperspectral disease detection studies in cereals and legumes (García-Vera et al. 2024, Urva 2021).

Feature importance analysis further supports the biological relevance of the extracted features. Both ANOVA and Random Forest approaches converged on `derivative_mean`, `red_edge_slope`, `NDVI`, `mean_reflectance`, and `skewness` as key indicators. These features are physiologically linked to chlorophyll content, red-edge shifts, and internal leaf structure, all of which are affected by fungal infections (Carter and Knapp 2001, Mahlein et al. 2019). The complementary perspectives of ANOVA (emphasizing univariate separability) and Random Forest (capturing nonlinear interactions) provided a more holistic understanding of discriminative traits. This methodological combination aligns with best practices in spectral feature selection, reducing the risk of discarding features that may appear weak in isolation but are informative in multivariate contexts (Belgiu and Drăguț 2016).

Finally, the evaluation of the model on new, independent chickpea samples confirmed its ability to generalize to unseen data. The accurate mapping of infection zones and stable classification of healthy leaves underscore the potential of this framework for real-world deployment. These outcomes highlight that beyond academic validation, the developed

approach can be scaled for field-based disease monitoring, integrating into precision agriculture platforms. The capacity to detect infections early and reliably across varieties provides tangible value for disease management, yield protection, and breeding programs.

Overall, this study advances the understanding of hyperspectral-based disease detection in chickpea leaves by demonstrating (i) the diagnostic power of NIR spectral changes, (ii) Cultivar differences in disease response and infection progression. (iii), the robustness of ensemble learning models under recall-prioritized evaluation, and (iv) the physiological relevance of selected spectral features. Future work should focus on scaling this approach to field conditions, integrating multisensory data (e.g., thermal, fluorescence), and expanding to other stress factors to further enhance resilience-oriented agricultural monitoring systems.

Conclusion

This study demonstrates that hyperspectral imaging coupled with machine learning can detect *Ascochyta* infection in chickpea leaves at a pre-symptomatic stage, before visible symptoms develop. After preprocessing, class balance was achieved via clustering-based relabeling plus SMOTE, yielding approximately 17,000 training and 7,500 testing spectra split evenly by health class, reducing bias and improving reliability.

Across Gradient Boosting, SVM (RBF), Random Forest, Logistic Regression, and k-NN, accuracy, precision, recall, and F1 all exceeded 90%, with near-identical confusion matrices. Recall was prioritized to limit false negatives, critical for early detection.

Feature importance from ANOVA and Random Forest converged on discriminative indices (NDVI, REP, mRENDVI, PRI, ARI), key wavelengths (700, 763–770, 884 nm), and spectral statistics (mean reflectance, skewness, kurtosis, red-edge slope). These relate to pigment degradation, canopy structure, and water-status shifts caused by infection, strengthening interpretability. Consistently across both methods, `derivative_mean`, `red_edge_slope`, NDVI,

mean_reflectance, and skewness emerged as the most informative descriptors of plant status, confirming their robustness in univariate and multivariate contexts.

Performance was held on new random samples, confirming generalization. Varietal analysis showed Grbz was more sensitive (sharper NIR declines), while Arifi appeared more resistant (more stable spectra), aligning with physiological expectations and supporting resistance screening.

The pipeline is adaptable to field-scale monitoring using UAV- or tractor-mounted systems, can inform breeding via targeted spectral features, and is transferable to other crop–pathogen systems. This adaptability is underpinned by the novelty of combining hyperspectral imaging with machine learning for the early, non-destructive detection of *Ascochyta* blight in chickpea leaves, together with the development of a comprehensive processing pipeline that addresses key hyperspectral data challenges. Future work should validate the approach across environments and seasons to ensure robustness and scalability. In sum, integrating hyperspectral imaging with machine learning produced >90% accuracy, stable performance across varieties and stages, and biologically meaningful markers, supporting a reliable, scalable, and non-invasive approach for crop protection and sustainable agriculture.

Declarations

Ethical statement

Not applicable: This manuscript does not include human or animal research.

Consent for publication

All authors have read and agreed to the published version of the manuscript.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This project was financially supported by the SpectraVOCs project (Université Mohammed VI Polytechnique (UM6P)). The lead author financial support from UM6P.

CRediT authorship contribution statement

Conceptualization: Mohamed Arame, Issam Meftah Kadmiri and Oussama Elbaraghi; **Methodology:** Mohamed Arame, Oussama Elbaraghi, Francois Bourzeix and Yahya Zennayi.; **Software:** Mohamed Arame and Ilyass Mafaa; **Validation:** Mohamed Arame and Abdelghani Chehbouni, Francois Bourzeix and Yahya Zennayi; **Formal analysis:** Mohamed Arame and Oussama Elbaraghi; **Investigation:** Mohamed Arame, Oussama Elbaraghi and Ilyass Mafaa; **Resources:** Mohamed Arame, Francois Bourzeix and Yahya Zennayi; **Data curation:** Mohamed Arame; **Writing—original draft preparation:** Mohamed Arame; **Writing—review and editing:** Issam Meftah Kadmiri, Francois Bourzeix, Yahya Zennayi and Ilyass Mafaa; **Visualization:** Mohamed Arame and Oussama Elbaraghi; **Supervision:** Issam Meftah Kadmiri, Francois Bourzeix, Yahya Zennayi, Ilyass Mafaa and Abdelghani Chehbouni; **Project administration:** Issam Meftah Kadmiri; **Funding acquisition:** Issam Meftah Kadmiri.

Acknowledgments

The authors express their gratitude to the members of the International Center for Agricultural Research in the Dry Areas (ICARDA) for establishing and managing the onsite experiments and fieldwork, and to the Analytics Lab team for their valuable support and engagement.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) only to improve the clarity, grammar, and academic style of the text. The tool was not used to generate scientific content, analyze data, or draw conclusions. All content was critically reviewed and edited by the authors, who bear full responsibility for the integrity, originality, and accuracy of the work.

References

- Basiru, S., Ait Si Mhand, K., Elfermi, R., Khatour, I., Errafii, K., Legeay, J., & Hijri, M. (2025). Enhancing chickpea growth through arbuscular mycorrhizal fungus inoculation: Facilitating nutrient uptake and shifting potential pathogenic fungal communities. *Mycorrhiza*, 35(1). <https://doi.org/10.1007/s00572-024-01174-4>
- Jha, U. C., Nayyar, H., Sharma, K. D., Jha, R., Thudi, M., Bakir, M., ... & Siddique, K. H. (2023). Chickpea Diseases: Breeding and “Omics” Approaches for Designing Next-Generation Disease-Resistant Chickpea Cultivar. In *Diseases in Legume Crops: Next Generation Breeding Approaches for Resistant Legume Crops* (pp. 1-30). Singapore: Springer Nature Singapore.
- Foresto, E., Carezzano, M. E., Giordano, W., & Bogino, P. (2023). Ascochyta blight in chickpea: An update. *Journal of Fungi*, 9(2), 203. <https://doi.org/10.3390/jof9020203>
- Ilyas A, Mirza SA, Hussain K, Ilyas A, Shahzad K. A comprehensive review on genetic resistance of chickpea to ascochyta blight. *Journal of Plant Pathology*. 2022 Nov;104(4):1337-54.
- Bayraktar, H., Özer, G., Aydoğan, A., & Palacioğlu, G. (2016). Determination of Ascochyta blight disease in chickpea using real-time PCR. *Journal of Plant Diseases and Protection*, 123(3), 109-117.
- Vasanth-Srinivasan, P., Noh, M. Y., Park, K. B., Kim, T. Y., Jung, W.-J., Senthil-Nathan, S., & Han, Y. S. (2025). Plant immunity to insect herbivores: Mechanisms, interactions, and innovations for sustainable pest management. *Frontiers in Plant Science*, 16, 1599450. <https://doi.org/10.3389/fpls.2025.1599450>
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>
- González, L. A., Bishop-Hurley, G. J., Handcock, R. N., & Crossman, C. (2015). Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and Electronics in Agriculture*, 110, 91–102. <https://doi.org/10.1016/j.compag.2014.10.018>
- Saeidan, A., Caulfield, J., Vuts, J., Yang, N., & Fisk, I. (2025). Detection of aphid infestation on faba bean (*Vicia faba* L.) by hyperspectral imaging and spectral information divergence methods. *Journal of Plant Diseases and Protection*, 132(3), 109.
- Mahlein, A.-K., Kuska, M. T., Behmann, J., Polder, G., & Walter, A. (2019). Hyperspectral sensors and imaging technologies in phytopathology: State of the art. *Annual Review of Phytopathology*, 57, 181–203. <https://doi.org/10.1146/annurev-phyto-080417-050100>
- Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74(1), 91–99. <https://doi.org/10.1016/j.compag.2010.06.009>
- Mahlein, A.-K. (2016). Plant disease detection by imaging sensors—Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2), 241–251. <https://doi.org/10.1094/PDIS-03-15-0340-FE>
- Arame, M., Kadmiri, I. M., Bourzeix, F., Zennayi, Y., Boulamtat, R., & Chehbouni, A. (2025). Detection of Leaf Miner Infestation in Chickpea Plants Using Hyperspectral Imaging in Morocco. *Agronomy*, 15(5), 1106.
- Pfordt, A., & Paulus, S. (2025). A review on detection and differentiation of maize diseases and pests by imaging sensors. *Journal of Plant Diseases and Protection*, 132(1), 40.
- Lacotte, B., Maury, C., Hémon, A., Habas, K., Lucas, M., Cotilloux, S., ... Albetis, J. (2022). Spatial-spectral analysis of hyperspectral images reveals early detection of downy mildew on grapevine leaves. *International Journal of Molecular Sciences*, 23(15), 8490.
- Khan, I. H., Liu, H., Li, W., Cao, A., Wang, X., Liu, H., ... & Yao, X. (2021). Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat. *Remote Sensing*, 13(18), 3612.
- Houasli, C., Sahri, A., Nsarellah, N., & Idrissi, O. (2021). Chickpea (*Cicer arietinum* L.) breeding in Morocco: Genetic gain and stability of grain yield and seed size under winter planting conditions. *Euphytica*, 217, 159. <https://doi.org/10.1007/s10681-021-02885-x>
- Bencheqroun, S. K., Lamarti, A., Ezzahiri, B., & Belabess, Z. (2021). Pathogen diversity and mating types of *Didymella rabiei* isolates collected from Morocco. *Plant Pathology: Quarantine*, 11, 15–29.
- Hartung, J., Wagener, J., Ruser, R., & Piepho, H. P. (2019). Blocking and re-arrangement of pots in greenhouse experiments: which approach is more effective?. *Plant Methods*, 15(1), 143.
- Grewal, R. K., Lulsdorf, M., Croser, J., Ochatt, S., Vandenberg, A., & Warkentin, T. D. (2009). Doubled-haploid production in chickpea (*Cicer arietinum* L.): role of stress treatments. *Plant cell reports*, 28(8), 1289-1299.
- Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., & Mahlein, A.-K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, 14, 45.
- Pena-García, Y., Durán-Patrón, R., & Macías-Sancho, J. (2023). Ascochyta blight in chickpea: An update. *Journal of Fungi*, 9(2), 203.
- Pérez-Pizá, M. C., Striker, G. G., & Stenglein, S. A. (2023). Seed-borne diseases in pasture grasses and legumes: State of the art and gaps in knowledge. *Journal of Plant Diseases and Protection*, 130(2), 225-244.
- Gao, X., Lu, Y., Zheng, C., & Zhang, C. (2020). Hyperspectral image classification based on superpixel and

- structure-preserving graph convolutional networks. *Remote Sensing*, 12(3), 397.
<https://doi.org/10.3390/rs12030397>
- Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, 16(3), 239–260. <https://doi.org/10.1007/s11119-014-9372-7>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H., & Sampson, P. H. (2005). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 928–938. <https://doi.org/10.1109/TGRS.2005.844430>
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Mahlein, A. K., Steiner, U., Dehne, H. W., & Oerke, E. C. (2012). Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture*, 13(6), 671–691. <https://doi.org/10.1007/s11119-012-9267-6>
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., ... & Baxter, I. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, 5, e4088. <https://doi.org/10.7717/peerj.4088>
- Mahlein, A. K., Rumpf, T., Welke, P., Dehne, H. W., Plümer, L., Steiner, U., & Oerke, E. C. (2013). Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, 128, 21–30. <https://doi.org/10.1016/j.rse.2012.09.019>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777. <https://doi.org/10.1366/0003702894202201>
- Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2021). Machine learning for plant stress phenotyping: Trends and future perspectives. *Trends in Plant Science*, 26(9), 882–898. <https://doi.org/10.1016/j.tplants.2021.05.010>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Applications*, 91, 464–471. <https://doi.org/10.1016/j.eswa.2017.09.030>
- Beisel, N. S., Callahan, J. B., Sng, N. J., Taylor, D. J., Paul, A. L., & Ferl, R. J. (2018). Utilization of single-image normalized difference vegetation index (SI-NDVI) for early plant stress detection. *Applications in Plant Sciences*, 6(10), e01186. <https://doi.org/10.1002/aps.3.1186>
- Mohd Johari, S. N. A., Khairunniza-Bejo, S., Mohamed Shariff, A. R., Husin, N. A., Mohd Masri, M. M., & Kamarudin, N. (2025). Effect of datasets size on the machine learning performance of the bagworm, *Metisa plana* (Walker) infestation using UAV remote sensing. *Journal of Plant Diseases and Protection*, 132(1), 52.
- Goswami, J., Das, R., Sarma, K. K., & Raju, P. L. N. (2021). Red edge position (REP), an indicator for crop stress detection: Implication on rice (*Oryza sativa* L.). *International Journal of Environment and Climate Change*, 11(4), 88–96.
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2–3), 337–354.
- Zarco-Tejada, P. J., González-Dugo, V., Williams, L. E., Suárez, L., Berni, J. A. J., Goldhamer, D., & Fereres, E. (2013). A PRI-based water stress index combining structural and chlorophyll effects: Assessment using diurnal narrow-band airborne imagery and the CWSI thermal index. *Remote Sensing of Environment*, 138, 38–50. <https://doi.org/10.1016/j.rse.2013.07.024>
- Roberts, D. A., Roth, K. L., Wetherley, E. B., Meerdink, S. K., & Perroy, R. L. (2018). Hyperspectral vegetation indices. In M. S. Moran & S. J. Goetz (Eds.), *Hyperspectral Indices and Image Classifications for Agriculture and Vegetation* (pp. 3–26).
- Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, 15(3), 697–703. <https://doi.org/10.1080/01431169408954109>
- Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., & Ramon, H. (2005). Automatic detection of yellow rust in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, 44(3), 173–188. <https://doi.org/10.1016/j.compag.2004.04.003>
- Li, S., Zhang, L., Pan, Y., Samui, G., & Song, L. (2021). Identifying optimal wavelengths as disease signatures

- using hyperspectral data and machine learning. *Remote Sensing*, 13(14), 2833. <https://doi.org/10.3390/rs13142833>
- Bravo, C., Moshou, D., West, J., McCartney, A., & Ramon, H. (2003). Early disease detection in wheat fields using spectral reflectance. *Biosystems Engineering*, 84(2), 137–145. [https://doi.org/10.1016/S1537-5110\(02\)00269-6](https://doi.org/10.1016/S1537-5110(02)00269-6)
- Baranowski, P., Jedryczka, M., Mazurek, W., Babula-Skowrońska, D., Siedliska, A., & Kaczmarek, J. (2015). Hyperspectral and thermal imaging of oilseed rape (*Brassica napus*) response to fungal species of the genus *Alternaria*. *PLoS ONE*, 10(3), e0122913. <https://doi.org/10.1371/journal.pone.0122913>
- Leucker, M., Mahlein, A. K., Steiner, U., Oerke, E. C., & Dehne, H. W. (2016). Hyperspectral imaging for monitoring symptom development of *Fusarium* head blight on wheat ears. *Plant Pathology*, 65(6), 1259–1271. <https://doi.org/10.1111/ppa.12491>
- Horler, D. N. H., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2), 273–288. <https://doi.org/10.1080/01431168308948546>
- Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., & Lin, H. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sensing of Environment*, 165, 123–134. <https://doi.org/10.1016/j.rse.2015.05.019>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Wijesinghe, H. D. M. U., Tharupath, K. M. C., & Jayasinghe, G. Y. (2025). Detecting Weligama coconut leaf wilt disease in coconut using UAV-based multispectral imaging and object-based classification. *Journal of Plant Diseases and Protection*, 132(4), 122.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Li, C., Czyż, E. A., Halitschke, R., Baldwin, I. T., Schaepman, M. E., & Schuman, M. C. (2023). Evaluating potential of leaf reflectance spectra to monitor plant genetic variation. *Plant Methods*, 19(1), 108. <https://doi.org/10.1186/s13007-023-01089-9>
- Grishina, A., Lysov, M., Ageyeva, M., Diakova, V., Sherstneva, O., Brilkina, A., & Vodenev, V. (2024). Bacterial and viral-induced changes in the reflectance spectra of *Nicotiana benthamiana* plants. *Horticulturae*, 10(12), 1363. <https://doi.org/10.3390/horticulturae10121363>
- Lin, F., Li, B., Zhou, R., Chen, H., & Zhang, J. (2024). Early detection of rice sheath blight using hyperspectral remote sensing. *Remote Sensing*, 16(12), 2047. <https://doi.org/10.3390/rs16122047>
- Mahlein, A. K. (2016). Plant disease detection by imaging sensors—Parallels and specific demands for precision agriculture. *Sensors*, 16(5), 641. <https://doi.org/10.3390/s16050641>
- Carter, G. A., & Miller, R. L. (1994). Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote sensing of environment*, 50(3), 295–302.
- Bauriegel, E., & Herppich, W. B. (2014). Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to *Fusarium* spec. infections on wheat. *Agriculture*, 4(1), 32–57. <https://doi.org/10.3390/agriculture4010032>
- Thomas, S., Behmann, J., Steier, A., Kraska, T., Müller, O., Rascher, U., & Mahlein, A. K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, 14, 45. <https://doi.org/10.1186/s13007-018-0301-9>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- García-Vera, Y. E., Polochè-Arango, A., Mendivelso-Fajardo, C. A., & Gutiérrez-Bernal, F. J. (2024). Hyperspectral image analysis and machine learning techniques for crop disease detection and identification: A review. *Sustainability*, 16(14), 6064.
- Urva, P. (2021). Detection and classification of grain crops and legumes disease: a survey. *Sparklinglight Transact Artif Intell Quantum Comp*, 1, 41–55.
- Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677–684. <https://doi.org/10.2307/2657068>
- Mahlein, A. K., Kuska, M. T., Behmann, J., Polder, G., & Walter, A. (2019). Hyperspectral sensors and imaging technologies in phytopathology: State of the art. *Annual Review of Phytopathology*, 57, 181–203. <https://doi.org/10.1146/annurev-phyto-082718-100108>
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>