

## Original papers

## Comprehensive analysis of hyperspectral features for monitoring canopy maize leaf spot disease



Yali Bai<sup>a,b,c,d</sup>, Chenwei Nie<sup>a,b,c</sup>, Xun Yu<sup>a,b,c</sup>, Mingyue Gou<sup>e</sup>, Shuaibing Liu<sup>a,b,c</sup>, Yanqin Zhu<sup>a,b,c</sup>, Tiantian Jiang<sup>a,b,c</sup>, Xiao Jia<sup>a,b,c</sup>, Yadong Liu<sup>a,b,c</sup>, Fei Nan<sup>a,b,c</sup>, Liming Li<sup>a,b,c</sup>, Bedir Tekinerdogan<sup>d</sup>, Yang Song<sup>a,b,c</sup>, Qingzhi Liu<sup>d,\*</sup>, Xiuliang Jin<sup>a,b,c,\*\*</sup>

<sup>a</sup> Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>b</sup> State Key Laboratory of Crop Gene Resources and Breeding, Beijing 100081, China

<sup>c</sup> National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China

<sup>d</sup> Information Technology Group, Wageningen University & Research, Wageningen, 6708 PB, the Netherlands

<sup>e</sup> State Key Laboratory of Wheat and Maize Crop Science, Collaborative Innovation Center of Henan Grain Crops, Center for Crop Genome Engineering, College of Agronomy, Henan Agricultural University, Zhengzhou, 450002, China

## ARTICLE INFO

## Keywords:

Hyperspectral features

Canopy

Unmanned aerial vehicle (UAV)

Leaf spot

Reinforcement learning

## ABSTRACT

Accurate quantification of hyperspectral features altered by plant disease infection is pivotal for effective disease management. However, the sensitivity of hyperspectral features to plant disease progression remains elusive, primarily because these features are often influenced by plant growth and environmental factors in addition to the specific disease. This study explores the sensitivity of biophysical and spectral features as indicators for maize adaptation to leaf spot disease. Using high-resolution UAV hyperspectral imaging, we captured maize adaptation dynamics over 30 days post-infection. We evaluated the sensitivity and importance of hyperspectral features for disease monitoring, including biophysical parameters retrieved by the PROSAIL model, and spectral features, including spectral reflectance, vegetation indices (VIs), and wavelet features (WFs). Our findings reveal that WFs first indicate disease response as early as 6 days after infection (DAI), followed by VIs at DAI 8, and variations in chlorophyll content ( $C_{ab}$ ) become apparent by DAI 10. The  $C_{ab}$ , plant senescence reflectance index (PSRI), and normalized photosynthetic reflectance index (NPRI) are determined to be important features at the early stage of the disease. Our experimental results show that the different feature sets are complementary at the early and severe stages of the disease. Our classification models integrating  $C_{ab}$ , VIs, and WFs showed higher overall accuracy than models using only spectral features or VIs, with a maximum improvement of 9.36%. However, these feature sets are redundant in the mild and initial severe disease stages, where models using only spectral features achieve the highest overall accuracy of 86.21%. This study underscores the novel insights by offering an understanding of plant responses to disease infection and enhancing early detection strategies.

## 1. Introduction

Maize (corn, *Zea mays* L.) is one of the most widely demanded and rapidly growing grains globally (Martinez and Fernandez, 2019; Yan and Tan, 2019). It is susceptible to numerous plant diseases, with as many as 112 reported worldwide (Kumar et al., 2014). Leaf spot, a prevalent disease across different regions, has led to significant yield losses in vulnerable maize germplasm (Dhami et al., 2015; Gonçalves et al., 2013). Traditional methods for monitoring leaf spot rely on visual

inspection by experts in the field, which is time-consuming, labor-intensive, and challenging to implement on a large scale.

With the advent of remote sensing technology, particularly hyperspectral imaging, a powerful tool has emerged that is capable of non-invasive, large-scale monitoring of plant disease (Chen et al., 2017; Hornero et al., 2020). Hyperspectral imaging's fine spectral resolution holds the promise of monitoring and diagnosing plant diseases (Mahlein et al., 2018; Thenkabail et al., 2004). This technology captures a broad spectrum of information, potentially revealing intricate details about

\* Corresponding author.

\*\* Corresponding author at: Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

E-mail addresses: [Qingzhi.liu@wur.nl](mailto:Qingzhi.liu@wur.nl) (Q. Liu), [jinxuliang@caas.cn](mailto:jinxuliang@caas.cn) (X. Jin).

<https://doi.org/10.1016/j.compag.2024.109350>

Received 7 January 2024; Received in revised form 21 June 2024; Accepted 11 August 2024

Available online 17 August 2024

0168-1699/© 2024 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

plant physiology and health that are invisible to the naked eye or through traditional multispectral imaging (Mahlein et al., 2013; Tian et al., 2021). Recent research in hyperspectral disease detection has predominantly focused on constructing machine learning models using hyperspectral features. For instance, Huang et al. (2022) underscored the necessity of choosing suitable machine learning methods and feature selection algorithms to enhance the accuracy of early and mid-term remote sensing monitoring of wheat stripe rust. Tian et al. (2021) demonstrated that a limited yet carefully chosen set of spectral features, identified through the machine learning based sequential floating forward selection (ML-SFFS) algorithm, effectively distinguishing rice leaves infected with leaf blast. Similarly, Luo et al. (2021) evidenced that machine learning models based on vegetation indices are highly promising for surveilling maize dwarf mosaic virus infections. However, remote sensing captures a multifaceted array of surface information, extending beyond vegetation and diseases to include elements like plant growth, environmental conditions, and so on (Weiss et al., 2020). These additional factors add complexity when interpreting hyperspectral features for disease monitoring and model transferability in canopy scales.

Hyperspectral features widely used in plant disease tasks primarily include biophysical parameters and spectral features (Hernandez-Clemente et al., 2019). Biophysical parameters such as leaf pigment content, derived from radiative transfer models, have been used as an indicator of vegetation health (Hornero et al., 2021; Tian et al., 2021; Zarco-Tejada et al., 2018). Increasing disease severity over time can ultimately cause variations among leaf photosynthetic pigment pools, known as constitutive changes (Hernandez-Clemente et al., 2019). Additionally, spectral features like spectral reflectance, vegetation indices (VIs), and wavelet features (WFs) have proven effective for plant disease monitoring (Camino et al., 2021; Tian et al., 2021). Nonetheless, their interpretation is challenged by illumination, structural, and influenced by different crop varieties and growth stages (Barton and North, 2001; Damm et al., 2015; Myneni et al., 1995). Furthermore, knowledge of causal relations between dynamics in plant response to leaf spot and required sensitivity of hyperspectral features is not fully exploited yet. Therefore, it is urgent to comprehend the sensitivity of different hyperspectral features for monitoring plant disease in canopy scales.

In this study, we hypothesize that the complexity of observations in plant disease monitoring obscures the inherent sensitivity of these observations for plant-pathogen relationships during leaf spot infection. Therefore, it is crucial to disentangle unwanted sensitivities from the targeted features for remote sensing-based classification in monitoring disease. We designed leaf spot infection experiments in three maize fields. The objectives are to (i) unravel the temporal sensitivity of

various hyperspectral features as disease severity increased using a time series analysis approach, (ii) analyze the importance of hyperspectral features (biophysical parameters and spectral features) in response to leaf spot infection, (iii) build classification models for assessing maize resistance to leaf spot disease based on the selected features.

## 2. Materials and methods

### 2.1. Experimental design

The study area was in Xinxiang City, China, a warm temperate continental monsoon climate characterized by hot and rainy and breed disease in summer. This region belongs to North Henan Plain (Yubei Plain), with maize as the main crop. Three sites were selected in this study: sites A, B, and C (Fig. 1).

Site A and B were in Yuanyang (35.15° N, 113.97° E). Both sites A and B were maize breeding fields with a planting density of 90,000 plants/ha planted on May 30, 2022. Site A consists of 120 rows, each 3 m, and site B consists of 1031 rows, each 3 m. In site A, 20 rows were considered as a genotype plot, with a total of 6 genotype plots. In each plot, 3 rows \* 5 plants were selected and treated with the leaf spot bacterial solution on the evening of August 2, 2022. The inoculation of maize with the bacterial leaf spot pathogen adhered to the technical specification for the identification of maize disease and pest resistance of China (NY/T 1248.10–2016). A water-dispersible granule formulation of *Cercospora zeina* bacterial solution was prepared at a concentration ranging from  $1 \times 10^5$  to  $1 \times 10^6$  pores per milliliter. The bacterial leaf spot solution was dispensed using a sprayer to ensure thorough coverage. The infected areas (6 areas) are shown in Fig. 1 using a red mask. The healthy areas (uninfected areas, yellow masked plots) were selected in equal numbers and sizes to the infected areas. When choosing these areas, we aimed to minimize the impact of pathogen transmission. The two adjacent plots' infection and healthy areas were connected separately, and the infected and healthy areas were distributed at intervals. In site B, each row was considered a plot with a genotype. Totally site B contained 1013 genotypes including 483 susceptibility genotypes and 530 resistant genotypes. This entire site was simultaneously treated with the leaf spot bacterial solution, paralleling the treatment applied to site A.

Site C, located in Xinxiang (35.30° N, 113.93° E), consists of 492 plots, each with a single genotype measuring 1.25 x 1.8 m. Site C contained 492 genotypes, including 311 susceptible and 181 resistant genotypes. This site, with a planting density of 90,000 plants/ha, was planted on May 30, 2021. Each plot of this site was treated with the leaf

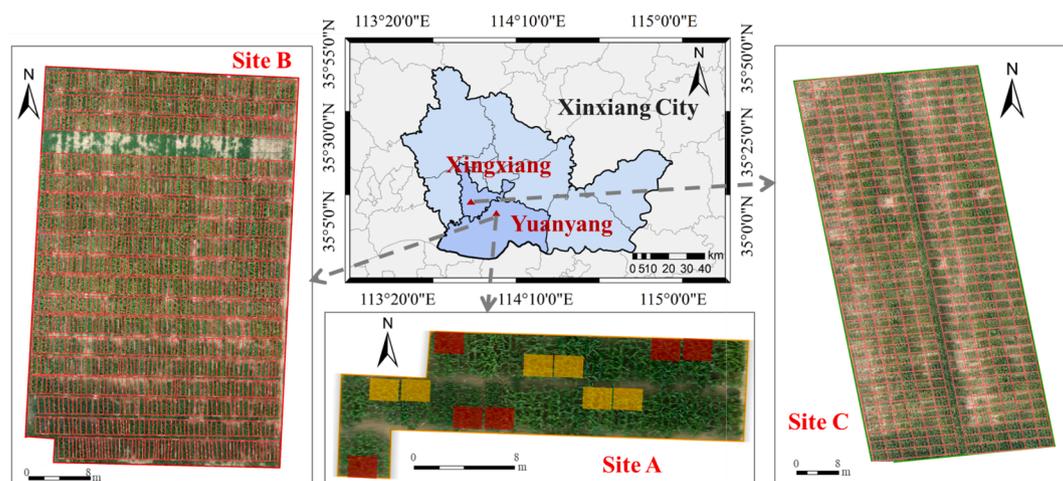


Fig. 1. Overview of field experiments at two locations (Xinxiang and Yuanyang) in 2021 and 2022. The upper center figure shows the geographic location of the experimental area. Site A (lower center) and B (left) in Yuanyang in 2022. The red mask in site A presented infected areas, and the yellow mask showed healthy areas. The site C (right) in Xinxiang in 2021.

spot bacterial solution on the evening of 2 August 2021.

Overall, all the infection processes occurred during the trumpet stage of maize growth. In managing the fields, we adhered to fertilizer application and irrigation practices guided by the expertise of local agricultural professionals. This approach aimed to provide a consistent balance of water and nutrients to reduce their potential impact on disease occurrence in the maize plants.

2.2. Field surveys

In field surveys, two different disease severity assessment methods were used. In site A, a temporal ground survey was employed to analyze the response time of remote sensing features to the disease (Fig. 2). In sites B and C, the resistance to leaf spot in genotypes was evaluated (Fig. 3) to verify the performance of hyperspectral features in disease monitoring tasks. In all field surveys, 3 maize plants were randomly selected from each area/plot. From each plant, 3 leaves were collected: the leaf above the ear, the ear leaf, and the leaf below the ear. There was a total of 9 leaves per area/plot.

In site A, field surveys were conducted simultaneously with hyperspectral data acquisition after the onset of the disease infection (Table 1). Following the disease severity grading method from Tian et al. (2021), disease severity was classified into different developmental stages. Though infections varied among leaves, visual inspections helped broadly categorize the infection processes. Disease severity was assessed in three levels based on lesion size and condition (Fig. 2). Specifically, early infection, from initial infection to scattered lesions on leaves, occurred between 2 and 10 days after infection (DAI); mild infection, with two or more connected linear lesions, was at DAI 12 and 15; severe infection, with three or more connected planar lesions, occurred at DAI 22, 24, and 30.

In sites B and C, the field surveys were conducted on August 26, 2022 (DAI 24) and September 2, 2021 (DAI 30), respectively (Table 1). The images of the selected leaves were documented. Subsequently, multiple experts collaboratively interpreted the proportion of leaf spots presented. Based on visual inspections, leaves were classified into one of six categories depending on the percentage of leaf spot coverage. Following this classification, the disease index (DI) for every plot was computed using the methodology of Huang et al. (2007). The resistance of maize was then determined by referencing the DI values. Maize resistance was classified into two groups, resistant (DI: 0–30 %) and susceptible (DI: 31 %–100 %), following the China national standard (NY/T 1248.1–2006), titled technical specification for identification of maize resistance to leaf spot disease.

2.3. Hyperspectral images collection and preprocessing

Hyperspectral images were obtained from the study site using a Pika L hyperspectral camera (Resonon, Bozeman, MT, USA) mounted on an unmanned aerial vehicle (UAV) (Matrice 600 Pro Hexacopter, DJI, Shenzhen, China). The images were taken at 50 m above the ground, flying at a speed of 1.5 m/s. All flights took place between 10:00 and 14:00 to minimize the effects of sun angle. The DJI M600 Pro UAV had a maximum payload capacity of 6 kg and a 30-minute endurance. It was

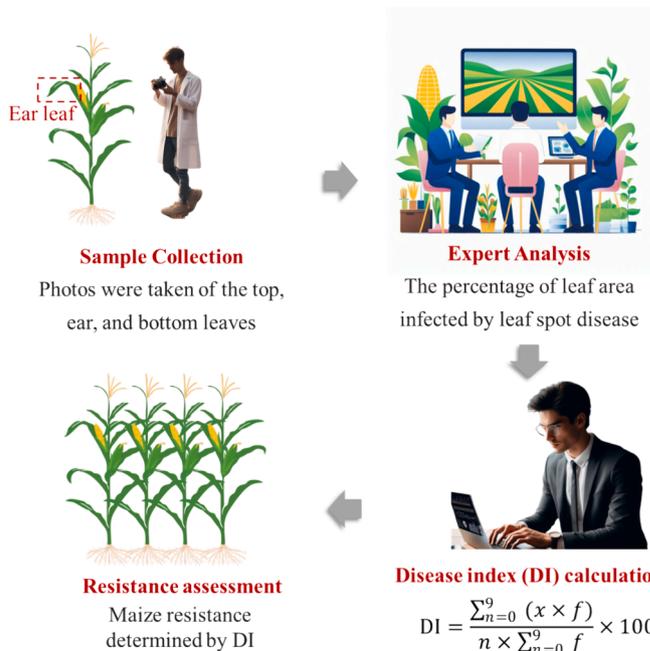


Fig. 3. Flowchart of maize resistance assessment. x is defined as 0, 1, 3, 5, 7, 9, representing respective disease coverage of 0 %, 0 %–5%, 6 %–25 %, 26 %–50 %, 76 %–100 % on the leaf surface; f is the total number of leaves of each class; n is the highest degree of disease severity (in this paper, n = 9). The resistance of maize genotypes was categorized as resistant and susceptible, corresponding to DI value ranges of 0–30 % and 31–100 %, respectively.

Table 1

Field surveys and hyperspectral images acquisition time.

Site	Year	Field surveys time	Hyperspectral images acquisition time
A	2022	Simultaneous with hyperspectral data acquisition	DAI 0, DAI 2, DAI 4, DAI 6, DAI 8, DAI 10, DAI 12, DAI 15, DAI 22, DAI 24, DAI 30
B	2022	DAI 24	DAI 12, DAI 15, DAI 22, DAI 30, DAI 37
C	2021	DAI 30	DAI 18, DAI 25, DAI 31, DAI 37

DAI: days after infection.

equipped with an onboard RTK-GNSS (real-time kinematic global navigation satellite system) unit for centimeter-level navigation. The Pika L camera featured 150 bands ranging from 400 to 1000 nm with a 2.1 nm spectral resolution and a 3.3 nm FWHM. The hyperspectral sensor’s 17 mm optical focal length provided an instantaneous field of view (IFOV) of 1.47 mrad and an angular field of view (FOV) of 30.8°. The image acquisition time is shown in Table 1.

The raw hyperspectral cubes were converted to radiance using the instrument’s radiometric calibration file. Cubes on the same flight path were combined into a linear cube. The linear cubes were georeferenced, based on high-resolution RGB images collected simultaneously and the

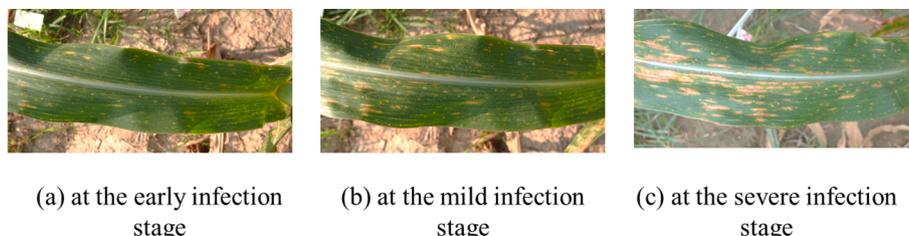


Fig. 2. RGB images of maize leaves under different leaf spot disease infection stages.

ground control points, using ArcGIS 10.2 (Esri, Redlands, CA, USA). Radiometric calibration was performed with a 60 % reflectance gray cloth with known reflectance collected at the same time as hyperspectral data in ENVI 5.3 (ITT Visual Information Solutions, Boulder, United States). Then, a seamless mosaic for the research site was created using all the linear cubes. Finally, zonal statistics were performed on the vector boundaries of each plot. This step was completed using the zonal statistics as a table plugin in ENVI 5.3.

#### 2.4. Feature extraction

This study investigated hyperspectral features comprising biophysical parameters and spectral features. Biophysical parameters derived from radiative transfer models (RTM) are considered, leveraging the comprehensive data captured by hyperspectral systems across an array of continuous narrow bands (Kokaly and Clark, 1999). Simultaneously, spectral features based on one or more spectral reflectance exhibited empirical relationships with plant information.

##### 2.4.1. Spectral features

Mean reflectance values at the canopy scale for each plot were utilized to calculate various narrow-band spectral features. These features include spectral reflectance, VIs, and WFs. Spectral reflectance within the wavelength range of 400 nm to 900 nm, comprising 117 bands, was chosen to avoid the lower signal-to-noise ratio typically observed at the edges of the sensor’s spectral range. VIs (Table S1) gathered from the literature were computed for spectral feature analysis. WFs were derived from the spectral data using the continuous wavelet transform (CWT), leveraging the PyWavelets library’s pywt (Python 3.7). The CWT was executed on the reflectance spectra to obtain wavelet coefficients, which encapsulate the spectral information at various scales. Specifically, for each reflectance spectrum within the dataset, the transformation was conducted over scales 3 to 8 and applied the Mexican Hat wavelet as the mother wavelet basis, following the methodology outlined by Cheng et al. (2010).

##### 2.4.2. Biophysical parameters

Biophysical parameters, including chlorophyll content ( $C_{ab}$ ), carotenoid content ( $C_{ar}$ ), equivalent water thickness ( $C_w$ ), dry matter content ( $C_m$ ), and anthocyanin content ( $A_{nth}$ ), were quantified by inverting the PROSAIL model, using the mean reflectance values of each plot (Fig. 4). The PROSAIL model couples the PROSPECT-D leaf reflectance model with the SAIL canopy model. PROSPECT-D accounts for the biophysical parameters in the leaf, while SAIL considers the properties of the canopy structure. We chose this model because, at the time of this study, the study plots had reached canopy closure and could be considered a uniform chaotic medium.

Initially, a simulation dataset containing 79.872 million measures was generated with the PROSAIL model. The input variables for the model were estimated based on leaf optical property datasets and prior studies (Nie et al., 2023), aiming to encompass the range of variability in the practical maize field (Table 2). The output spectra were weighted average according to the spectral response function of the PikaL sensor, ensuring uniform and comparable spectral resolution across different spectra.

A multilayer perceptron (MLP) model was trained on the simulated dataset to establish the relationship between the simulated spectra and biophysical parameters. The leaf area index (LAI) was obtained using actual values measured in the field with a SUNSCAN instrument (Delta-T Devices, Cambridge, UK) to address the ill inversion issue. The MLP network was designed with an input layer with the same number of nodes as the number of the PikaL bands plus one (LAI). The inner layers successively applied a compression ratio of two, reducing the number of nodes by half. The output layer had the same number of nodes as the biophysical parameters. The network employed dropout regularization to prevent overfitting with a probability of 0.2. The ReLU activation

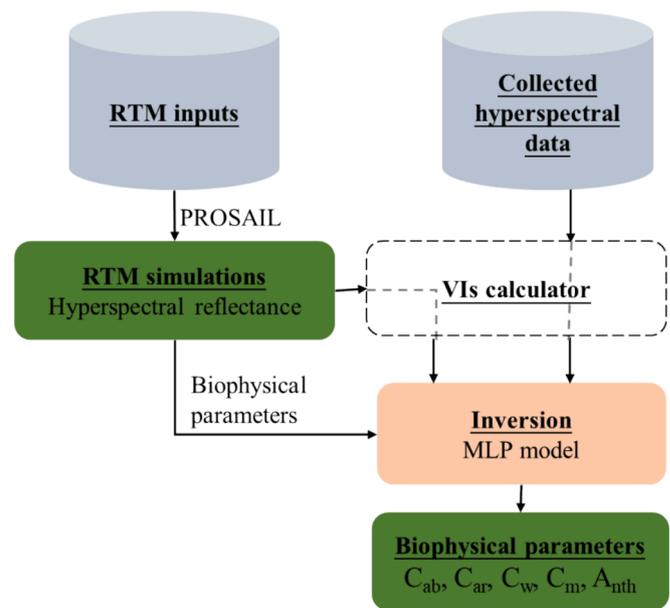


Fig. 4. Flowchart of PROSAIL simulation and retrieval.  $C_{ab}$ : chlorophyll content,  $C_{ar}$ : carotenoid content,  $C_w$ : equivalent water thickness,  $C_m$ : dry matter content,  $A_{nth}$ : anthocyanin content, VIs: vegetation indices, MLP: multi-layer perceptron.

Table 2  
Inputs for PROSAIL simulation.

Model	Variable	Acronym	Range	Steps	Units
PROSPECT-D (leaf level)	Chlorophyll content	$C_{ab}$	1–100	5	$\mu\text{g}/\text{cm}^2$
	Carotenoid content	$C_{ar}$	0–26	1	$\mu\text{g}/\text{cm}^2$
	Anthocyanin content	$A_{nth}$	0–20	1	$\mu\text{g}/\text{cm}^2$
	Brown pigment content	$C_{brown}$	0.2	N/A	N/A
	Equivalent water thickness	$C_w$	0–0.08	0.01	cm
	Dry matter content	$C_m$	0–0.04	0.01	$\text{g}/\text{cm}^2$
	Leaf structure parameter	N	1–1.5	0.1	N/A
SAIL (canopy level)	Leaf area index	LAI	1–5	0.3	N/A
	Hot-spot size parameter	Hspot	0.25	N/A	$\text{mm}^{-1}$
	Solar zenith angle	SZA	65	N/A	Degree
	Observed zenith angle	OZA	0	N/A	Degree
	Relative azimuth angle	RAA	0	N/A	Degree
	Average leaf inclination angle	ALA	30–50	10	Degree

Note: The ‘Range’ column indicates the variation interval for each parameter, and ‘Steps’ describes the quantification intervals within this range. For example,  $C_{ab}$  quantified within a range of 1 to 100  $\mu\text{g}/\text{cm}^2$ , with 5  $\mu\text{g}/\text{cm}^2$  increments. N/A: not applicable.

function was applied throughout the network. Finally, the inverted  $C_{ab}$  and  $A_{nth}$  were evaluated using the coefficient of determination ( $R^2$ ) and root mean square error (RMSE) based on the field measured  $C_{ab}$  and  $A_{nth}$  at site A (Fig. S1). The  $C_{ab}$  and  $A_{nth}$  were measured with Dualex Scientific+ (Force-A, Orsay, France) following the measurements described by Zhou et al. (2023). These measurements were conducted simultaneously with field surveys.

## 2.5. Feature analysis

We proposed a comprehensive feature analysis method to identify sensitive features, as illustrated in Fig. 5. This process begins by utilizing all hyperspectral features retrieved from infected and healthy plots at site A. Given that each plot consists of numerous observations (pixels), we subdivided each plot into twelve equal parts, enriching the hyperspectral data for analysis. In total, both infected and healthy areas had 72 samples. We applied time series analysis to unravel the temporal sensitivity of various hyperspectral features to increase leaf spot severity (Section 2.5.1), considering the effect size using Cohen’s d. Subsequently, we trained an intelligent agent capable of automatically learning policies to select an optimal feature subset to address multicollinearity and select-rich information features. These features are named as ‘parsimony features’ (Section 2.5.2).

### 2.5.1. Response time analysis

The response time analysis is crucial for understanding and quantifying the temporal dynamics of hyperspectral features in response to the vegetation stress process (Behmann et al., 2014; Langenkämper et al., 2022). To accurately focus on the target disease, a time-based normalization approach was employed to account for canopy structural influences that frequently overlay dynamics in hyperspectral features. This normalization strategy is essential to avoid misinterpretation of data and has been proven to compensate for most of the remote sensing parameter variation caused by illumination effects and canopy structure (Damm et al., 2022; Zarco-Tejada et al., 2012). This approach calculated the difference between the hyperspectral feature values observed at a given time (denoted as  $F_t$ ) and the average of these features recorded before the infection event on August 2nd (denoted as  $F_r$ ). We use  $\Delta F$  to represent the time series of change in each hyperspectral feature, which is calculated by

$$\Delta F = F_t - F_r \quad (1)$$

The resulting time series represents the increment of individual hyperspectral features considering the first observations (DAI 0) in

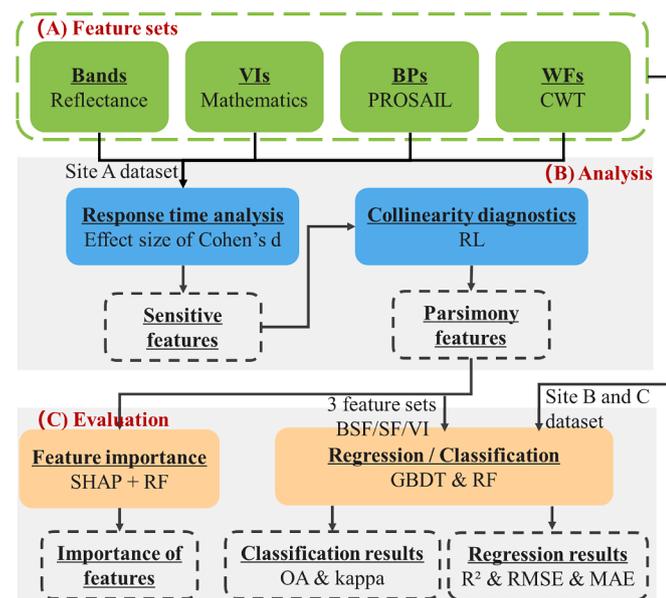


Fig. 5. Overview of the methodology used for feature analysis and classification. Note: VIs: vegetation indices, BPs: biophysical parameters, WFs: wavelet features, CWT: continuous wavelet transform, RL: reinforcement learning, SHAP: Shapley additive explanations, GBDT: gradient-boosting decision tree, RF: random forest, OA: overall accuracy; 3 feature sets include biophysical parameters in combination with spectral features (BSF), spectral features alone (SF), and vegetation indices alone (VI).

physical units. To assess the impact and reliability of leaf spot infection on plant responses as estimated by various hyperspectral features, we employed two statistical measures: (1) 95 % confidence intervals based on a z-distribution; (2) effect size using Cohen’s d, which accounts for the standard deviation of the healthy canopy.

We define a threshold effect size Z to represent the substantial difference between the healthy and infected observations. Based on our comprehensive testing and previous study (Charach et al., 2011),  $Z=0.8$  is the most effective value and, therefore, chosen for the implementations. In this step, we established a sensitive feature selection criterion: the start time of consecutive substantial difference with the same sign is determined as when the hyperspectral feature begins to respond to the leaf spot disease. In addition, only features that showed a response before DAI 15 were selected for further analysis in the following stages. These were observed at more than three-time points, demonstrating the substantial differences between the healthy and infected observations.

### 2.5.2. Collinearity diagnostics

Following sensitivity analysis in Section 2.5.1, we identified several sensitive features, some exhibiting multicollinearity and redundancy. These sensitive features will now be subjected to an in-depth collinearity assessment to refine our feature selection for the classification model. We modeled the task of collinearity diagnostics as a Markov decision process (MDP) from Mou et al. (2022). Consequently, RL was adopted as the strategy of choice for formulating a sophisticated collinearity diagnostics methodology (Fig. 6). Compared with traditional approaches, RL allows for continuous learning and adjustment, a critical advantage in environments where the features change over time (Mou et al., 2022). In addition, RL can efficiently manage the complex, high-dimensional spaces typical of hyperspectral images (Feng et al., 2022).

In this MDP, we define the action, state, and reward. Let  $F$  denote the set of all the sensitive features, representing the total count of these features, and let  $f_i$  represent an individual feature within  $F$ . The RL model’s state, action, and reward are defined as follows.

**State:** The state ( $s$ ) is a subset of  $F$ , with the state space being the power set of  $F$ . Each state is uniquely depicted by a feature encoding—a vector of length  $n$ . If a feature is selected, its corresponding element is set to 1; otherwise, it remains 0.

**Action:** The agent’s primary objective is to select a feature from  $F$  at each iteration. Action refers to the position of a feature in a list, where the chosen position’s value changes from 0 to 1. This selection process continues until the predefined number of features is reached;  $x$  features were optimal for ensuring classification models.

The transition function defines how the system moves from one state to another based on an action taken. Specifically, if an action is taken to select a feature already chosen within the current episode, the state  $s_t$  remains unaltered in the succeeding step. If the chosen feature  $f_i$  has not been selected in the current episode, then the state  $s_t$  evolves to a new state  $s_{t+1}$ , where the corresponding element of  $f_i$  is set to 1 in the feature encoding. When the count of chosen features matches the predefined threshold, the state transitions to a “Terminal” state, signifying the end of the episode.

Mathematically, the transition function can be described as:

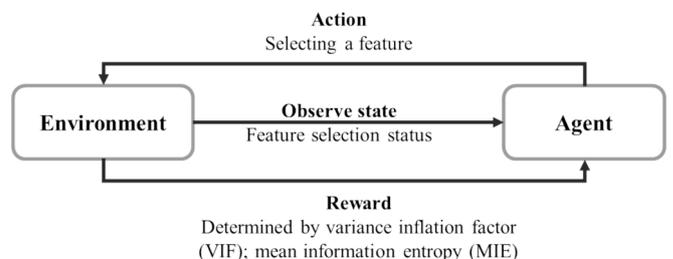


Fig. 6. Overview of reinforcement learning strategy.

$$s_{t+1} = \begin{cases} \text{Terminal} & \text{if reaching the preset number} \\ s_t & \text{if } a_t = \text{select a feature } f_i \text{ already in } s_t \\ s_t + f_i & \text{if } a_t = \text{select a new feature } f_i \end{cases} \quad (2)$$

in which  $s_t$  is the current state,  $s_{t+1}$  is the next state,  $f_i$  is the selected feature, and  $a_t$  is the current action.

**Reward:** The agent is rewarded based on the efficacy of its feature selection. A penalty of  $-1$  is applied if a repeated feature is chosen. Conversely, a reward of  $1$  is granted when a non-collinear feature is selected. To determine collinearity, we employ the variance inflation factor (VIF), a metric introduced by Gareth et al. (2013), to measure the variance increase of a regression coefficient due to collinearity. A VIF below 10 suggests that the chosen feature is not collinear with previously selected ones. However, if collinearity is selected (VIF>10), the information entropy, gauging the richness of spectral information, is used. The reward difference between states and  $s_{t+1}$  is governed by the disparity in information entropy. Specifically, the average information entropy of the selected features was measured as Equation (3), in which  $A$  represents a set of previously executed actions,  $N$  designates the total count of chosen features, and  $P$  stands for the probability mass function.

$$\text{MIE}(s) = -\frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} \sum_{n=1}^N P(f_i^n) \log_2 P(f_i^n) \quad (3)$$

Subsequently, the reward function is mathematically structured as:

$$r_t = \begin{cases} -1 & \text{if } a_t = \text{select a feature } f_i \text{ already in } s_t \\ 1 & \text{if } a_t = \text{select a feature } f_i \text{ noncollinearity} \\ & \text{with features already in } s_t \\ \text{MIE}(s_{t+1}) - \text{MIE}(s_t) & \text{if } a_t = \text{select a feature } f_i \text{ collinearity} \\ & \text{with features already in } s_t \end{cases} \quad (4)$$

in which  $s_t$  is the current state,  $s_{t+1}$  is the next state,  $f_i$  is the selected feature,  $a_t$  is the current action, and  $r_t$  is the current reward.

To implement this strategy, we adapted our approach to incorporate the proximal policy optimization (PPO) algorithm. This training process was facilitated by the Ray framework. The tune library was used to search for the optimal model hyperparameters. Detailed settings of each parameter are provided in Table 3.

## 2.6. Feature evaluation

To validate the selected features, we assessed the importance of the parsimony features (Section 2.6.1) and demonstrated the effectiveness of these features for machine learning models across different sites and times (Section 2.6.2).

### 2.6.1. Feature importance

We utilized the Shapley additive explanations (SHAP) method (Lundberg and Lee, 2017) for model feature interpretation in this evaluation. We generated SHAP summary plots based on random forest (RF) models to visualize the significance of each feature. The

performances of the models are shown in Table S2.

The feature importance analysis process consisted of two main parts. Firstly, we conducted a SHAP summary using the infected areas. We utilized hyperspectral features that were calculated throughout the entire experimental period and compared them to DAI, which increases with leaf spot severity. This phase aimed to gain insights into the relationship between hyperspectral features and disease severity. Secondly, we compared the hyperspectral features obtained from the infected and healthy areas at early, mild, and severe infection stages. This comparison sought to identify the contribution of features in distinguishing between infected and healthy plots at different disease stages.

### 2.6.2. Machine learning models

We employed two regression machine learning algorithms to estimate the DI for each plot on the survey dates: RF and gradient-boosting decision tree (GBDT). Similarly, we applied the same machine learning algorithms in a classification context to categorize maize genotypes as disease-resistant or susceptible.

In this study, we conducted 100 iterations of data division into two samples, namely the training and testing samples, by randomly selecting 80% ( $n = 810$  for 2022;  $n = 394$  for 2021) and 20% ( $n = 203$  for 2022;  $n = 98$  for 2021) of the dataset, respectively. The validation process involved k-fold cross-validation, wherein the original sample was partitioned randomly into 10 subsamples of equal size and repeated five times. Optimal hyperparameters for the models were identified by grid search and are presented in Table 4. Model performance was quantified by calculating R2, RMSE, and mean absolute error (MAE) for regression accuracy. We analyzed the confusion matrix, overall accuracy (OA), and Cohen's kappa coefficient (kappa) for classification accuracy. In the development of the regression and classification models, we utilized Python 3.7 and the scikit-learn library. For GBDT models, we employed the GradientBoostingRegressor and GradientBoostingClassifier from the ensemble module. Similarly, we used the RandomForestRegressor and RandomForestClassifier from the same module for RF models.

To compare the effectiveness of different feature sets in classification models, we evaluated the classification accuracies of three distinct sets of hyperspectral features: (1) biophysical parameters in combination with spectral features (BSF), (2) spectral features alone (SF), and (3) vegetation indices alone (VI).

## 3. Results

### 3.1. Temporal response of features after leaf spot infection

This section explores the temporal dynamics of hyperspectral features following maize leaf spot infection. The effect sizes quantify the analysis by comparing these features between infected and healthy areas to pinpoint the onset and progression of the disease's impact (detailed in Supplementary Table S3–S5).

#### 3.1.1. Biophysical parameters

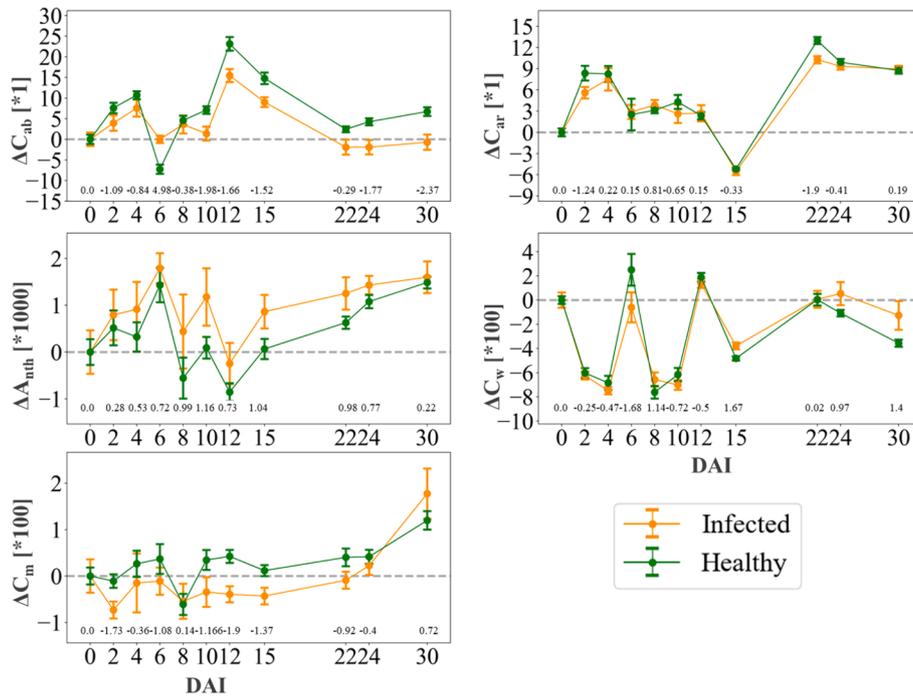
Compared to the DAI 0 measurements (Fig. 7), the change in  $C_{ab}$  ( $\Delta C_{ab}$ ) exhibited an initial increase followed by a gradual decrease, with the accumulation of  $C_{ab}$  in infected areas consistently lagging that in the

**Table 3**  
Parameters for PPO in Ray.

Parameter	Value	Parameter	Value
lr	5e-3	num_workers	6
train_batch_size	256	num_envs_per_worker	1
sgd_minibatch_size	128	rollout_fragment_length	10
num_sgd_iter	10	entropy_coeff	0.01
gamma	0.8	vf_loss_coeff	1.0
lambda	0.95	vf_clip_param	10.0
clip_param	0.3	use_gae	True
kl_coeff	0.2		

**Table 4**  
Hyperparameters for each machine learning models.

Hyperparameter	Search space	Regression	Classification
<b>Gradient-boosting decision tree (GBDT)</b>			
n_estimators	10–100	80	50
learning_rate	0.1–0.5	0.1	0.3
max_depth	2–15	5	5
min_samples_leaf	2–15	5	5
<b>Random forest (RF)</b>			
Number of trees	10–50	40	30
Maximum depth	20–100	50	50



**Fig. 7.** Changes of biophysical parameters in infected areas ( $n = 72$  samples) and healthy areas ( $n = 72$  samples) on different DAIs. The error bars represent the 95 % confidence interval, and the values included represent the effect size, as determined by Cohen's  $d$ .  $\Delta$ : normalized considering the change of the first infection day (2 August);  $C_{ab}$ : chlorophyll content;  $C_{ar}$ : carotenoid content;  $C_w$ : equivalent water thickness;  $C_m$ : dry matter content;  $A_{nth}$ : anthocyanin content.

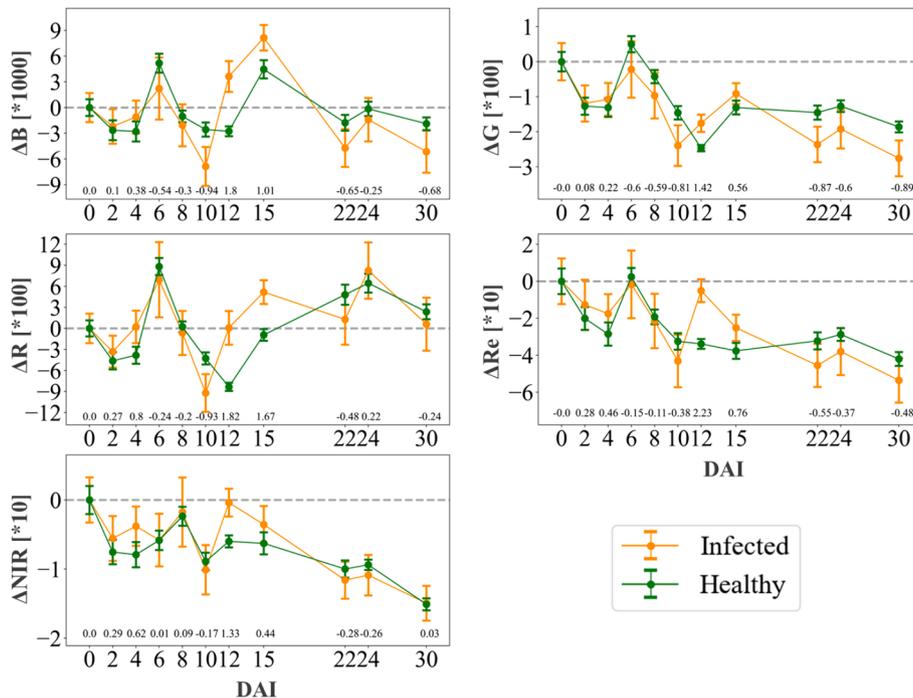
healthy areas. Starting from DAI 10, a substantial difference in the  $C_{ab}$  change rate between infected and healthy areas was observed, substantiating the effect of leaf spot disease on chlorophyll dynamics.

Variations in  $\Delta A_{nth}$ ,  $\Delta C_w$ ,  $\Delta C_m$ , and  $\Delta C_{ar}$  did not show consistently substantial differences between infected and healthy areas. These measurements revealed a noisy nature, making it challenging to find a

clear trend or identify consistently substantial differences between the infected and healthy areas.

### 3.1.2. Spectral reflectance

Spectral reflectance across the 400–900 nm wavelength range did not meet the sensitivity criteria established for feature analysis as



**Fig. 8.** Changes of spectral reflectance in infected ( $n = 72$  samples) and healthy areas ( $n = 72$  samples) on different DAIs. The red (R), green (G), blue (B), red edge (Re), and near-infrared (NIR) bands were presented by wavelengths of 473 nm, 560 nm, 666 nm, 717 nm, and 843 nm corresponding to the central wavelength positions of the RedEdge-MX multispectral camera (Micasense company, USA), respectively. The error bars represent the 95 % confidence interval, and the values included represent the effect size, as determined by Cohen's  $d$ .

outlined in Section 2.5.1 (Fig. 8). The lack of consistency and continuity in the data suggests that spectral reflectance is prone to noise, which could complicate disease detection and risk false positives in practical disease monitoring. These findings underscore the limited transferability of models based on spectral reflectance for robust disease monitoring.

### 3.1.3. VIs

The earliest response of VIs to the disease occurred at DAI 8, with the photochemical reflectance index ( $PRI_n$ ) related to xanthophyll and photosynthetic efficiency continuing to respond to leaf spot disease until DAI 30 (Table 5). Subsequently, starting at DAI 12, the responding VIs included five indices associated with carotenoid and chlorophyll (PSRI, B, G, R, and RGI), three indices related to xanthophyll and photosynthetic efficiency ( $PRI_{515}$ ,  $PRI_{M3}$ , and  $PRI_{M4}$ ), one index related to chlorophyll fluorescence (CUR), and three indices related to anthocyanins (VARI,  $VARI_2$ , and ARIm). From DAI 15, responding VIs included two related to structural traits (NDVI and  $LIC_1$ ) and two to chlorophyll (SIPI and PSND). From DAI 22, responding VIs included the structural-associated OSAVI and chlorophyll-associated TCARI. The response time of VIs highlights a temporal progression in the disease's impact on various physiological and biochemical processes.

Following a multicollinearity diagnosis based on RL, five representative VIs (NDVI, PSRI,  $PRI_n$ , RGI, and  $VARI_2$ ) were selected (Fig. 9). The change trends in NDVI maintaining a close-to-zero difference between DAI 0–DAI 12 and DAI 0 in both infected and healthy areas, indicate that NDVI does not largely change. After DAI 12, the difference in infected areas compared to DAI 0 was consistently higher than in healthy areas, with the infected areas dropping seriously. VARI followed a similar trend to NDVI but shifted earlier by DAI 10. PSRI's variations were minimal until DAI 15, after which the infected plots increased notably. RGI mirrored PSRI's changes but began at DAI 8. Meanwhile,  $PRI_n$  maintaining a close-to-zero difference with DAI 0 in healthy areas, experienced a sharp increase after DAI 15 in the infected areas. Compared to raw spectral reflectance, these VIs provide a relatively stable indication of vegetation status.

### 3.1.4. WFs

From the early to mild infection stages, the WFs are sensitive to leaf spot infection, predominantly spanning the 450–750 nm range, encompassing scales of 3 to 8 (Table 6). Notably, the earliest response of WFs to the disease was observed at DAI 6, specifically with  $WF_{586,6}$ ,  $WF_{590,7}$ , and  $WF_{590-598,8}$ , located within the yellow edge region (550–650 nm). This region is primarily characterized by the overlapping absorption of various pigments, particularly chlorophyll and anthocyanin, indicating pigment dynamics associated with early disease response.

Starting from DAI 12, responding WFs included  $WF_{469-473,3}$ ,  $WF_{465-469,4}$ ,  $WF_{461-469,6}$ ,  $WF_{461-469,7}$ ,  $WF_{461-477,8}$ , and  $WF_{506-510,8}$ , located in the blue region (450–550 nm), primarily governed by chlorophyll absorption. Furthermore, the  $WF_{735,3}$ ,  $WF_{735,4}$ , and  $WF_{735,5}$  were located in the red edge region (700–750 nm), demonstrating sensitivity to rapid photosynthetic rate changes. From DAI 15, the  $WF_{821,5}$ ,

**Table 5**  
Response time of vegetation indices (VIs).

Response time	VIs
DAI 2	--
DAI 4	--
DAI 6	--
DAI 8	$PRI_n$
DAI 10	--
DAI 12	PSRI, $PRI_{515}$ , $PRI_{M3}$ , $PRI_{M4}$ , B, G, R, RGI, CUR, VARI, $VARI_2$ , ARIm
DAI 15	NDVI, $LIC_1$ , SIPI, PSND
DAI 22	OSAVI, TCARI
DAI 24	SPRI, NPQI, NPCI, CTRI, PRI, $PRI_{M1}$ , $PRI_{M2}$ , $BRI_1$ , $BRI_2$ , $LIC_2$ , HI

positioned in the near-infrared region and indicative of sensitive plant structure, began to show a significant difference between the infected and healthy areas. These findings imply a broader impact of the disease, associated with progressive changes in chlorophyll content, photosynthetic efficiency, and plant structural integrity.

After conducting a multicollinearity diagnosis using RL, five representative wavelet features ( $WF_{494,5}$ ,  $WF_{461,6}$ ,  $WF_{469,7}$ ,  $WF_{548,8}$ , and  $WF_{590,8}$ ) were selected (Fig. 10). The trends exhibited by  $WF_{494,4}$ ,  $WF_{461,6}$  and  $WF_{469,7}$  were consistent, displaying a steady increase from DAI 0 to DAI 30 in both infected and healthy areas, with the changes in the infected plots consistently occurring at a faster rate. Conversely, the changes of  $WF_{548,8}$  and  $WF_{590,8}$  in infected areas consistently preceded those in the healthy areas. The observed trends in WFs displayed noisy and discontinuous patterns, particularly in the early stages of disease development. This suggests that while WFs responded rapidly to the onset of disease, their stability remains uncertain during the early stage of infection.

### 3.2. Importance of identified features

According to our experimental trials, we identified 11 key features characterized by high information entropy and low collinearity. Our RL agent meticulously selected these features, ensuring an optimal balance between information richness and feature independence. These included 10 spectral-based features, 5 VIs and 5 WFs, and a biophysical parameter, namely,  $C_{ab}$ . These selected features were utilized as predictors in the final classification model for maize resistance.

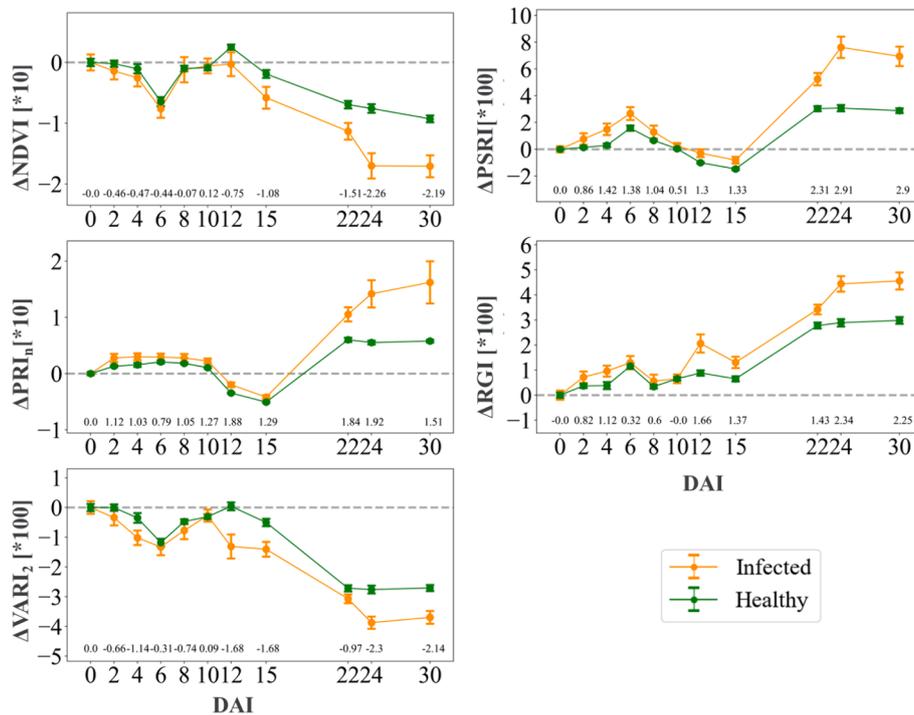
The importance of the chosen features in identifying infected and healthy areas during various disease stages is shown in Fig. 11 a–c. At the early disease stage,  $C_{ab}$  emerged as the feature with the highest importance scores in models discriminating between infected and healthy areas, followed by  $PRI_n$  and PSRI. These features also showed high importance at the mild and severe disease stages.  $VARI_2$ , PSRI, and  $PRI_n$  demonstrated the highest importance scores in models regressing different levels of severity, as illustrated in Fig. 11 d. Notably,  $VARI_2$  displayed a negative relationship with disease severity. At the same time, PSRI and  $PRI_n$  exhibited a positive correlation.

### 3.3. Regression models for quantifying DI

The performances of RF and GBDT regression algorithms for estimating the DI of maize leaf spot, utilizing hyperspectral features in 2022 (DAI 30) and 2021 (DAI 31), are presented in Fig. 12 and Fig. 13, respectively. In 2022, the RF model exhibited an  $R^2$  ranging from 0.51 to 0.62. The GBDT algorithm displayed comparable performance, with an  $R^2$  between 0.71 and 0.79. For the GBDT model, RMSE values ranged from 0.12 to 0.15, and MAE values varied from 0.10 to 0.12. In contrast, the RF model with  $R^2$  values from 0.59 to 0.63 in 2021, while the GBDT model's  $R^2$  spanned from 0.67 to 0.74. The RF model reported RMSE and MAE values from 0.12 to 0.17 and 0.10 to 0.14, respectively, whereas the GBDT model exhibited RMSE values between 0.08 and 0.12 and MAE from 0.07 to 0.10. Optimal performance in both models and years was achieved using spectral features. These results demonstrate a redundancy in integrating biophysical parameters and spectral features for monitoring disease at the initial severe stages.

### 3.4. Classification models for assessing maize resistance

The resistant and susceptible classification models using hyperspectral imagery from different time points in 2022 demonstrated varying levels of OA and kappa (Fig. 14; Table S6–S7). OA ranged from 56.16 % to 86.21 %, and kappa values ranged from 0.39 to 0.76. The models exhibited varying accuracy patterns based on three distinct sets of hyperspectral features. DAI 12 and 37 notably enhanced OA and kappa when integrating biophysical parameters with WFs and VIs. Omitting the  $C_{ab}$  decreased OA by up to 3.45 % and 2.96 %, respectively.



**Fig. 9.** Changes of selected vegetation indices (VIs) in infected (n = 72 samples) and healthy areas (n = 72 samples) on different DAIs. The error bars represent the 95 % confidence interval, and the values included represent the effect size, as determined by Cohen’s d.

**Table 6**  
Response time of wavelet features (WFs).

Response time	WFs
DAI 4	--
DAI 6	WF <sub>586,6</sub> , WF <sub>590,7</sub> , WF <sub>590-598,8</sub>
DAI 8	--
DAI 10	--
DAI 12	WF <sub>469-473,3</sub> , WF <sub>735,3</sub> , WF <sub>465-469,4</sub> , WF <sub>735,4</sub> , WF <sub>465,5</sub> , WF <sub>735,5</sub> , WF <sub>461-469,6</sub> , WF <sub>461-469,7</sub> , WF <sub>461-477,8</sub> , WF <sub>506-510,8</sub>
DAI 15	WF <sub>477-486,3</sub> , WF <sub>473-486,4</sub> , WF <sub>469-481,5</sub> , WF <sub>821,5</sub> , WF <sub>473,6</sub> , WF <sub>510-515,6</sub> , WF <sub>502-510,7</sub> , WF <sub>481-502,8</sub>

Furthermore, removing  $C_{ab}$  and WFs led to additional reductions in OA by up to 9.36 % and 8.87 %, respectively. A decline in this trend was observed at DAI 15. At DAI 22 and 30, the classification models using different feature sets achieved a balanced and best performance, with no significant differences observed among the three cases of features used, with the best precision obtained by combining WFs and VIs.

The 2021 data presented an OA range of 65.31 % to 85.71 % and kappa values from 0.42 to 0.74 (Fig. 15; Table S8–S9). DAI 31 achieved the peak level of accuracy in the study. Paralleling the 2022 findings, DAI 37 of 2021 also exhibited superior OA and kappa when using BSF. The exclusion of the  $C_{ab}$  feature led to a 5.10 % reduction in OA, while excluding both the  $C_{ab}$  and WFs resulted in further OA drops of a maximum of 7.14 %. By DAI 18, the once pronounced complementary strengths of the three feature types began to moderate, and by DAI 25 and DAI 31, this trend disappeared entirely.

#### 4. Discussion

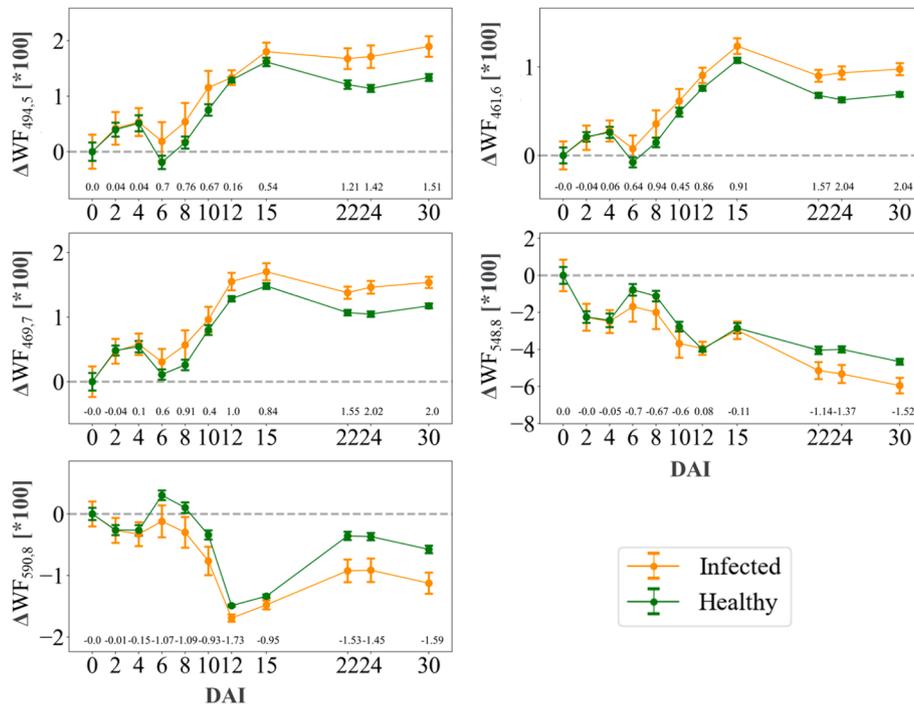
##### 4.1. Sensitivity of normalized features for plant-pathogen interaction

Our results revealed that the normalized features response to leaf spot disease infection display varying temporal dynamics and amplitude. This expected behavior can be attributed to the sensitivity of

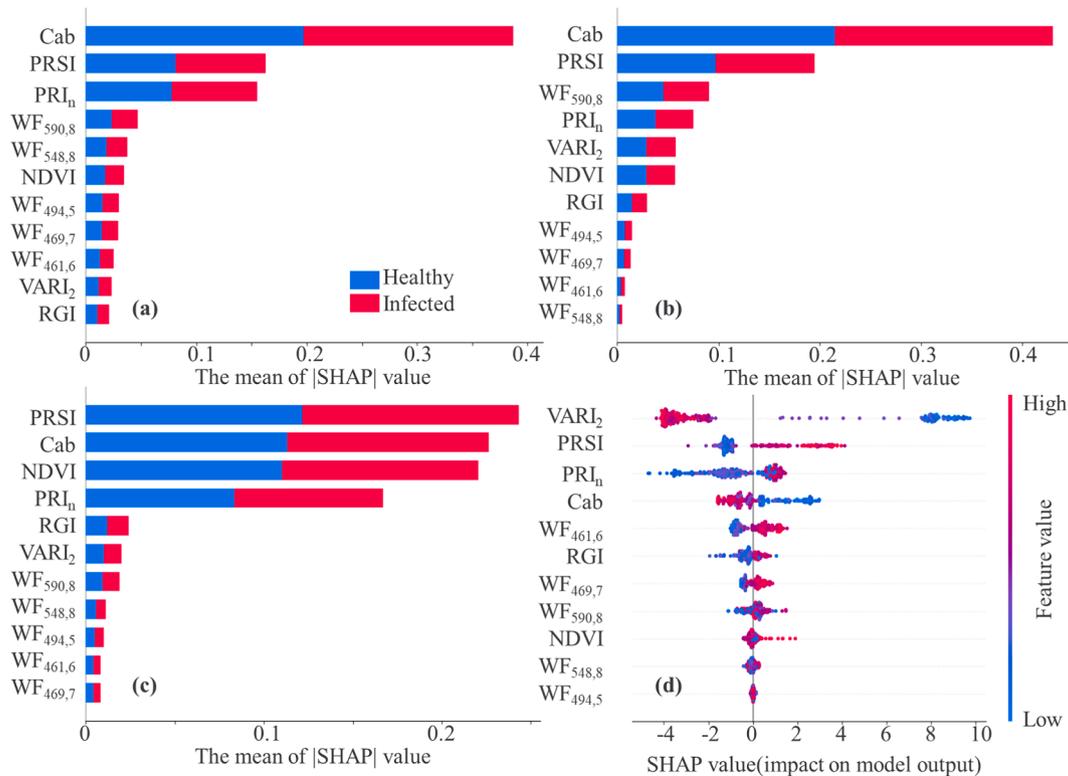
individual features to specific plant adaptation mechanisms, including functional, biochemical, or structural changes brought on by disease infection refer to a review on this topic (Hernandez-Clemente et al., 2019). Notably, we observed a dual response from  $\Delta C_{ab}$ . An initial slow, short-term increase was observed in response to maize growth and exacerbation of leaf spot disease, followed by a longer-term decrease under sustained leaf spot disease infection. Conversely, the  $\Delta PRI_n$  exhibited an immediate increase but did not decrease over time. This complex interplay of features responses to leaf spot disease at canopy scales, operating at different time scales, is why models struggle with cross-temporal application.

In fact, from DAI 6 to DAI 10, scattered lesions were already visible to the naked eye while maize was still at the growth stage (tassel). Our results showed that the  $C_{ab}$ , retrieved by the radiative transfer model, indicates a slight separation of infected and healthy canopies. It suggests slightly slower chlorophyll accumulation and limited growth rates in the infected areas. In addition, an increase of  $PRI_n$  in the infected area could be driven by physiological changes, likely signifying a slight reduction of stomatal aperture compared to the healthy area. The yellow edge region wavelet features (WF<sub>586,6</sub>, WF<sub>590,7</sub>, and WF<sub>590-598,8</sub>) as remote sensing proxy sensitive to various pigment variations like  $C_{ab}$  and Anth. Notably, the  $\Delta WF_{598,8}$  time series showed differences in DAI 6 between infected and healthy areas, indicating pigment dynamics due to plant response to leaf spot disease. These dynamics of features support the idea that plants could have reacted physiologically and biochemically to the evolving leaf spot infection at the early stages.

From DAI 12, scattered lesions began extending into contiguous ones as the maize entered the reproductive stage (silking). More features began responding to the leaf spot disease infection. For instance, the  $PRI_{M4}$ , a proxy for short-term changes in photosynthetic activity and less sensitivity to structural effects, showed a larger reduction in infected than healthy areas. Higher changes of CUR in infected areas than in healthy areas could be due to the physiological impacts and changes in chlorophyll fluorescence. The PSDN showed a decline in both infected and healthy areas but a more pronounced decrease in the infected area, indicating a quicker rate of plant senescence. Additionally, blue region



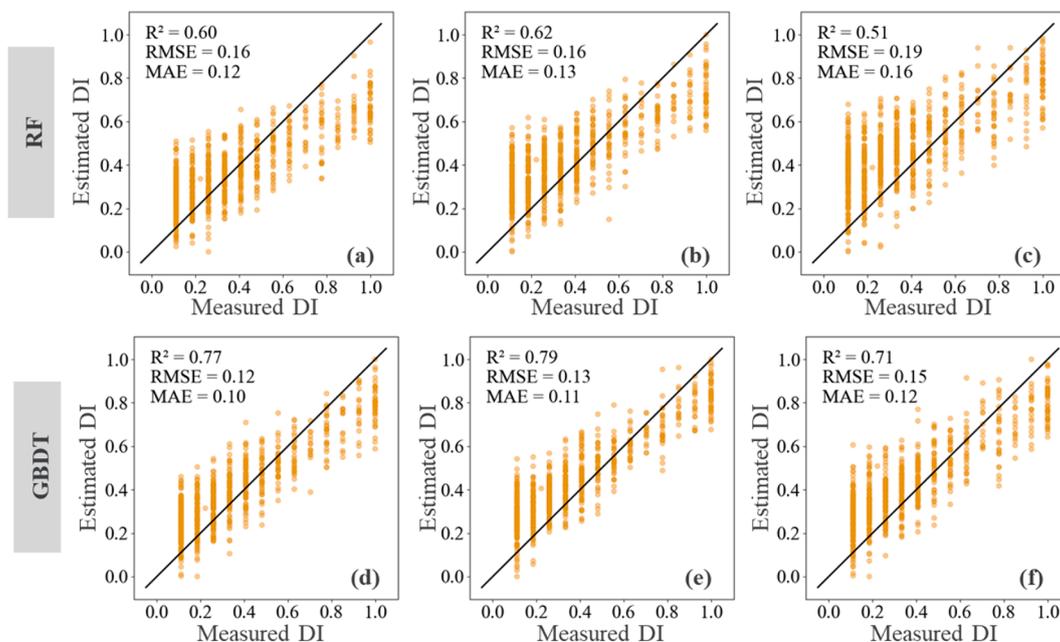
**Fig. 10.** Changes of selected wavelet features (WFs) in infected (n = 72 samples) and healthy areas (n = 72 samples) on different DAIs. The error bars represent the 95 % confidence interval, and the values included represent the effect size, as determined by Cohen's d.



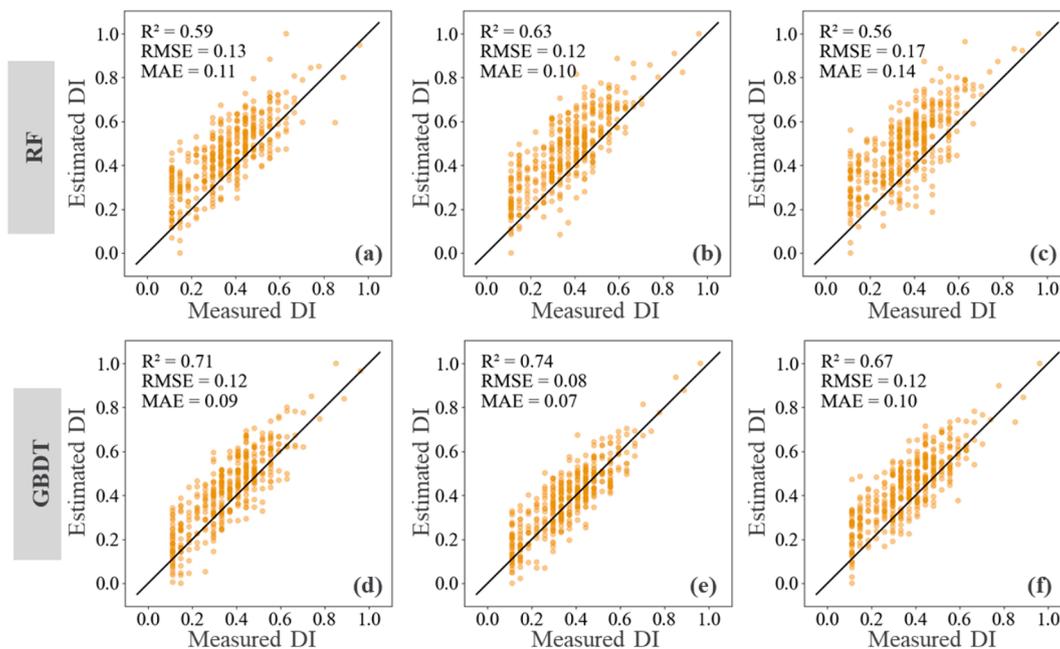
**Fig. 11.** Feature importance ranking by SHAP value. (a)–(c) represents SHAP value when classifying infected and healthy areas at each leaf spot disease stage: (a) early infection, (b) mild infection, and (c) severe infection. (d) depicts SHAP value when regressing infected plots from DAI 0 to DAI 30 with increasing maize leaf spot disease. The feature importance is ordered along the y-axis, and the contributions of features (SHAP value) to the model are on the x-axis. Positive SHAP values refer to a more severe leaf spot disease.

wavelet features and red edge region wavelet features (WF<sub>735,3</sub>, WF<sub>735,4</sub>, and WF<sub>735,5</sub>) showed notable differences between leaf spot infected and healthy canopies. The blue region is related to pigment degradation,

while the red edge region is sensitive to chlorophyll concentration and quick photosynthetic rate changes (Carter and Knapp, 2001; Hernández-Clemente et al., 2016).



**Fig. 12.** Scatter plots between the measured DIs and the best estimated DIs by RF (upper) and GBDT (bottom) in 2022. The models were trained on different sets of features: (a) and (d) utilize biophysical parameters in combination with spectral features (BSF), (b) and (e) rely on spectral features, including wavelet features and vegetation indices (SF), and (c) and (f) focus on vegetation indices alone (VI), respectively.



**Fig. 13.** Scatter plots between the measured DIs and the best estimated DIs by RF (upper) and GBDT (bottom) in 2021. The models were trained on different sets of features: (a) and (d) utilize biophysical parameters in combination with spectral features (BSF), (b) and (e) rely on spectral features, including wavelet features and vegetation indices (SF), and (c) and (f) focus on vegetation indices alone (VI), respectively.

Contrasting with prior research, our findings provide analytical insight into the biophysical and spectral feature responses throughout the progression of leaf spot infection. While several of our observations align with prior studies, some findings diverge. Interestingly, none of the spectral reflectance features demonstrated significant continuous changes in response to the disease over the observation period. However, several studies have identified the green, red, and near-infrared spectral regions as sensitive to various plant diseases (Bauriegel et al., 2011; Garcia-Ruiz et al., 2013; Huang et al., 2012). Both results seem to contradict but can be explained by the susceptibility and instability of

spectral reflectance to environmental factors at different time scales. To our knowledge, this is the first time that the stability of spectral reflectance in disease monitoring under natural conditions has been experimentally verified using canopy time-series data.

Displaying the changes in biophysical parameters, spectral reflectance, VIs, and WFs is crucial for understanding the disease dynamics comprehensively. This comprehensive understanding is fundamental for improving disease monitoring strategies. In practical terms, it is essential to select features that show a consistent response to disease for effective monitoring tasks. Specifically, for early-stage disease

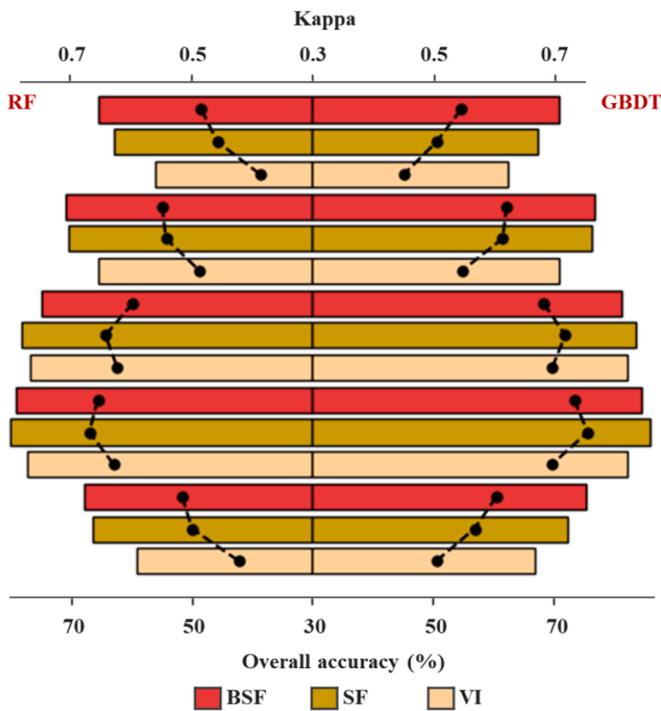


Fig. 14. Performances of classification models in 2022. Overall accuracy (bars) and Kappa (bullets) metrics of RF and GBDT algorithms classifying maize plants resistant to leaf spot disease from susceptible. The models were trained using biophysical parameters in combination with spectral features (BSF), spectral features including wavelet features and vegetation indices (SF), and vegetation indices alone (VI), respectively. The models used hyperspectral imageries from DAI 12, DAI 15, DAI 22, DAI 30, and DAI 37, 2022, respectively.

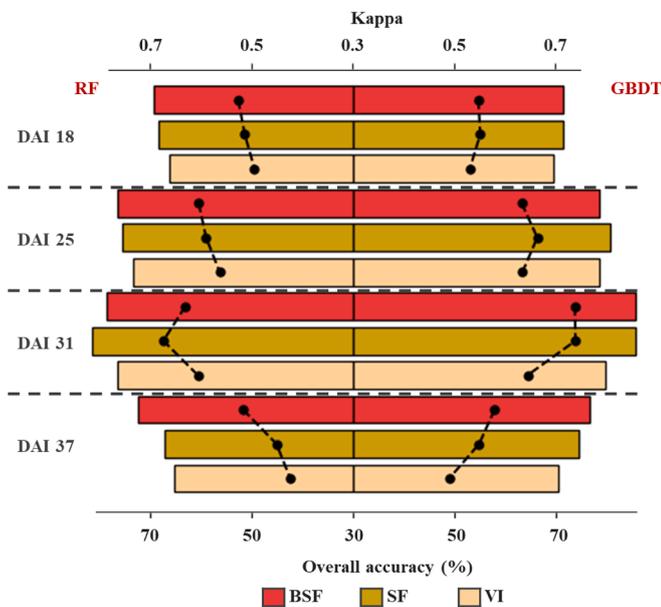


Fig. 15. Performances of classification models performance in 2021. Overall accuracy (bars) and Kappa (bullets) metrics of RF and GBDT algorithms classifying maize plants resistant to leaf spot disease from susceptible. The models were trained using biophysical parameters in combination with spectral features (BSF), spectral features including wavelet features and vegetation indices (SF), and vegetation indices alone (VI), respectively. The models used hyperspectral imageries from DAI 12, DAI 15, DAI 22, DAI 30, and DAI 37, 2021, respectively.

monitoring, features that respond before DAI 10 should be prioritized. For mild-stage monitoring, features that respond before DAI 15 should be selected. For severe stage monitoring, choosing features that exhibit continuous and stable responses is important.

#### 4.2. Performance of hyperspectral features

The ranking of feature importance provides insights into the ability of each feature to distinguish between healthy and infected areas. The  $C_{ab}$  and PRSI consistently stand out as the most influential features throughout every stage of infection. Their prominent SHAP values underscore their essential role in differentiating between healthy and infected maize areas. As infection severity escalates, there is a notable reduction in chlorophyll concentration, aligning with patterns of decline observed in other stress experiments conducted under controlled conditions (Damm et al., 2022; Hornero et al., 2021) and field surveys (Camarero et al., 2012). Notable, the importance of PRSI even exceeded  $C_{ab}$ , reinforcing its utility as a sensitive index to variations in pigment composition, including  $C_{ab}$  and  $C_{ar}$ . Additionally, VIs such as NDVI,  $PRIn$ , RGI, and  $VARI_2$  gain increased importance in severe infection, indicating their potential to reflect compounded stress impacts on plant physiology as the disease advances.

The performances of our maize resistance classification models demonstrated both complementarity and redundancy among the biophysical parameters and spectral features. These diverse spectral features exhibited complementary information at the early stages of disease infection, improving accuracy when integrated. Specifically, models incorporating  $C_{ab}$ , VIs, and WFs outperformed models that solely relied on spectral features or VIs, with a maximum OA difference of 9.36%. This finding highlighted the added value of integrating different types of features during the early disease stages (Poblete et al., 2021; Zarco-Tejada et al., 2018). However, as the disease progressed, the complementarity of these features gradually diminished, indicating a redundancy of information at the mild and initial severe disease stages. This observation aligns with previous studies that have reported that a reduced set of selected features can achieve optimal results in disease classification models (Poblete et al., 2020; Wang et al., 2019).

Our study provided a novel explanation for this contradiction. The complementary advantages of multiple features are effective for monitoring challenges during the early stages of disease classification and the latter part of the severe stage when disease intersects with crop senescence. As the disease becomes more distinguishable at the mild and initial severe stages, the classification difficulty decreases, leading to the emergence of redundant properties among the different features. This understanding of the contradictory nature of the influence of multiple features on disease classification models contributes to further research.

In addition, the performance of classification models showed inconsistency. At the early stages of the disease (DAI 12), the models performed very well in identifying the resistant genotypes, but they struggled to correctly identify susceptible genotypes, resulting in high precision and lower recall. This is because the susceptible genotype showed weak symptoms, making it easy to misclassify as resistant. Conversely, in the severe stages of the disease (DAI 37), the models exhibited high recall but lower precision. This indicates that while the models could successfully identify susceptible genotypes, they mistakenly classified some resistant genotypes as susceptible. This is because the susceptible genotype undergoes accelerated senescence as maize was in the mid-milk stage, making it easier to identify with high recall. Meanwhile, resistant genotypes also showed signs of senescence and were easily misclassified as susceptible, leading to lower precision.

#### 4.3. Limitations of this study

The findings of this study, substantiated by numerous field experiments, leave open questions regarding their applicability to other maize leaf diseases. However, maize is also susceptible to other foliar diseases

such as rust, blight, and anthracnose. The potential applicability of our findings to these additional diseases has not been studied and necessitates further research. Additionally, the specificity of the identified features in distinguishing between different diseases must be confirmed in future studies.

The mismatch between field survey data and remote sensing data has always been a source of uncertainty in disease monitoring. To address this, multi-scale analysis should be employed to analyze the disconnect between various scales and explore the factors causing this disconnect. Moreover, the current methodology of evaluating maize resistance through subjective assessment by researchers introduces potential biases. To address this, future work should consider the adoption of automated image analysis algorithms. Such technological advancements promise a more objective and practical approach to quantifying disease impact on leaf tissue, substantially diminishing subjectivity and bolstering the efficacy of monitoring processes for crop resistance.

## 5. Conclusion

Hyperspectral technology provides diverse avenues to explore plant-pathogen relationships on large scales, yet the complexity of leaf spot-associated plant responses remains a significant problem. This study develops a systematic approach for identifying hyperspectral features sensitive to leaf spot disease. Our experimental results illustrate that temporal response analysis enables the analysis of inherent spectral characteristics of plants and pathogens. We conclude that canopy-derived  $C_{ab}$ ,  $PRIn$ , and PRSI are significant features at all leaf spot disease infection stages. The effectiveness of our classification models highlights the complementary information from biophysical and spectral-based features during the early and severe disease infection stages, suggesting the information redundancy at the mild and initial severe stages of disease infection. This conflict was confirmed for the first time through continuous observation data collection. Our research methodology and the important features derived from this study could guide other disease monitoring research.

## CRedit authorship contribution statement

**Yali Bai:** Writing – review & editing, Writing – original draft. **Chenwei Nie:** Supervision. **Xun Yu:** Data curation. **Mingyue Gou:** Data curation. **Shuaibing Liu:** Data curation. **Yanqin Zhu:** Data curation. **Tiantian Jiang:** Data curation. **Xiao Jia:** Data curation. **Yadong Liu:** Data curation. **Fei Nan:** Data curation. **Liming Li:** Data curation. **Bedir Tekinerdogan:** Supervision. **Yang Song:** Writing – review & editing. **Qingzhi Liu:** Writing – review & editing, Methodology. **Xiuliang Jin:** Writing – review & editing, Project administration, Funding acquisition.

## 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.

## Data availability

Data will be made available on request.

## Acknowledgment

This research was supported by Nanfan special project, CAAS (YBXM2401, YBXM2402), National Natural Science Foundation of China (42071426) Research and application of key technologies of smart brain for farm decision-making platform (2021ZXJ05A03), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2024.109350>.

## References

- Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U., Herppich, W., 2011. Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers Electronics in Agriculture* 75, 304–312.
- Behmann, J., Steinrücken, J., Plümer, L., 2014. Detection of early plant stress responses in hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* 93, 98–111.
- Camarero, J.J., Olano, J.M., Alfaro, S.J.A., Fernández-Marín, B., Becerril, J.M., García-Plazaola, J.I., 2012. Photoprotection mechanisms in *Quercus ilex* under contrasting climatic conditions. *Flora-Morphology, Distribution, Functional Ecology of Plants* 207, 557–564.
- Camino, C., Calderon, R., Parnell, S., Dierkes, H., Chemin, Y., Roman-Ecija, M., Montes-Borrego, M., Landa, B.B., Navas-Cortes, J.A., Zarco-Tejada, P.J., Beck, P.S.A., 2021. Detection of Xylella fastidiosa in almond orchards by synergic use of an epidemic spread model and remotely sensed plant traits. *Remote Sens. Environ.* 260, 112420.
- Carter, G.A., Knapp, A.K., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.* 88, 677–684.
- Charach, A., Dashti, B., Carson, P., Booker, L., Lim, C.G., Lillie, E., Yeung, E., Ma, J., Raina, P., & Schachar, R. (2011). *AHRQ Comparative Effectiveness Reviews. Attention Deficit Hyperactivity Disorder: Effectiveness of Treatment in At-Risk Preschoolers; Long-Term Effectiveness in All Ages; and Variability in Prevalence, Diagnosis, and Treatment.* Rockville (MD): Agency for Healthcare Research and Quality (US).
- Chen, G., He, Y.N., De Santis, A., Li, G.S., Cobb, R., Meentemeyer, R.K., 2017. Assessing the impact of emerging forest disease on wildfire using Landsat and KOMPSAT-2 data. *Remote Sens. Environ.* 195, 218–229.
- Cheng, T., Rivard, B., Sánchez-Azofeifa, G.A., Feng, J., Calvo-Polanco, M., 2010. Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation. *Remote Sens. Environ.* 114, 899–910.
- Damm, A., Cogliati, S., Colombo, R., Fritsche, L., Genangeli, A., Genesio, L., Hanus, J., Peressotti, A., Rademski, P., Rascher, U., Schuettemeyer, D., Siegmann, B., Sturm, J., Miglietta, F., 2022. Response times of remote sensing measured sun-induced chlorophyll fluorescence, surface temperature and vegetation indices to evolving soil water limitation in a crop canopy. *Remote Sens. Environ.* 273, 112957.
- Dhimi, N., Kim, S.K., Paudel, A., Shrestha, J., Rijal, T., 2015. A review on threat of gray leaf spot disease of maize in Asia. *Journal of Maize Research and Development* 1, 71–85.
- Feng, J., Li, D., Gu, J., Cao, X., Shang, R., Zhang, X., Jiao, L., 2022. Deep reinforcement learning for semisupervised hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–19.
- García-Ruiz, F., Sankaran, S., Maja, J.M., Lee, W.S., Rasmussen, J., Ehsani, R., 2013. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers Electronics in Agriculture* 91, 106–115.
- Gonçalves, R., Figueiredo, J., Pedro, E., Meirelles, W., Leite, R., Sauer, A., Paccola, L., 2013. Etiology of Phaeosphaeria leaf spot disease of maize. *J. Plant Pathol.* 95, 559–569.
- Hernandez-Clemente, R., Kolari, P., Porcar-Castell, A., Korhonen, L., Möttus, M., 2016. Tracking the seasonal dynamics of boreal forest photosynthesis using EO-1 Hyperion reflectance: Sensitivity to structural and illumination effects. *IEEE Transactions on Geoscience and Remote Sensing* 54 (9), 5105–5116.
- Hernandez-Clemente, R., Hornero, A., Mottus, M., Penuelas, J., Gonzalez-Dugo, V., Jimenez, J.C., Suarez, L., Alonso, L., Zarco-Tejada, P.J., 2019. Early diagnosis of vegetation health from high-resolution hyperspectral and thermal imagery: lessons learned from empirical relationships and radiative transfer modelling. *Current Forestry Reports* 5, 169–183.
- Hornero, A., Hernandez-Clemente, R., North, P.R.J., Beck, P.S.A., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Monitoring the incidence of Xylella fastidiosa infection in olive orchards using ground-based evaluations, airborne imaging spectroscopy and Sentinel-2 time series through 3-D radiative transfer modelling. *Remote Sens. Environ.* 236, 111480.
- Hornero, A., Zarco-Tejada, P.J., Quero, J.L., North, P.R.J., Ruiz-Gomez, F.J., Sanchez-Cuesta, R., Hernandez-Clemente, R., 2021. Modelling hyperspectral- and thermal-based plant traits for the early detection of Phytophthora-induced symptoms in oak decline. *Remote Sens. Environ.* 263, 112570.
- Huang, W., Lamb, D.W., Niu, Z., Zhang, Y., Liu, L., Wang, J., 2007. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precis. Agric.* 8, 187–197.
- Huang, J., Liao, H., Zhu, Y., Sun, J., Sun, Q., Liu, X., 2012. Hyperspectral detection of rice damaged by rice leaf folder (*Cnaphalocrocis medinalis*). *Computers Electronics in Agriculture* 82, 100–107.
- Huang, L., Liu, Y., Huang, W., Dong, Y., Ma, H., Wu, K., Guo, A., 2022. Combining random forest and XGBoost methods in detecting early and mid-term winter wheat stripe rust using canopy level hyperspectral measurements. *Agriculture-Basel* 12, 74.
- Kokaly, R.F., Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67, 267–287.
- Kumar, R., Hooda, K., Olakh, D., Kaur, H., Malik, V.K., Kumar, S., 2014. Research Article Reaction of QPM inbred lines against Maydish Leaf Blight (MLB) and Charcol Rot. *Electronic Journal of Plant Breeding* 4, 1280–1283.

- Langenkämper, D., Mogstad, A.A., Hansen, I.M., Baussant, T., Bergsagel, Ø., Nilssen, I., Frost, T.K., Nattkemper, T.W., 2022. Exploring time series of hyperspectral images for cold water coral stress response analysis. *PLoS One* 17, e0272408.
- Luo, L., Chang, Q., Wang, Q., Huang, Y., 2021. Identification and severity monitoring of maize dwarf mosaic virus infection based on hyperspectral measurements. *Remote Sens. (Basel)* 13, 4560.
- Mahlein, A.K., Kuska, M.T., Behmann, J., Polder, G., Walter, A., 2018. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annual review of phytopathology* 56 (1), 535–558.
- Mahlein, A.K., Rumpf, T., Welke, P., Dehne, H.W., Pluemer, L., Steiner, U., Oerke, E.C., 2013. Development of spectral indices for detecting and identifying plant diseases. *Remote Sens. Environ.* 128, 21–30.
- Martinez, E.L., Fernandez, F.J.B., 2019. Chapter 4 - Economics of production, marketing and utilization. In: Serna-Saldivar, S.O. (Ed.), *Corn*, Third Edition. AACCI International Press, Oxford, pp. 87–107.
- Mou, L., Saha, S., Hua, Y., Bovolo, F., Bruzzone, L., Zhu, X.X., 2022. Deep reinforcement learning for band selection in hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14.
- Nie, C., Shi, L., Li, Z., Xu, X., Yin, D., Li, S., Jin, X., 2023. A comparison of methods to estimate leaf area index using either crop-specific or generic proximal hyperspectral datasets. *Eur. J. Agron.* 142, 126664.
- Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Detection of *Xylella fastidiosa* infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. *ISPRS J. Photogramm. Remote Sens.* 162, 27–40.
- Poblete, T., Navas-Cortes, J.A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., Landa, B.B., Zarco-Tejada, P.J., 2021. Discriminating *Xylella fastidiosa* from *Verticillium dahliae* infections in olive trees using thermal- and hyperspectral-based plant traits. *ISPRS J. Photogramm. Remote Sens.* 179, 133–144.
- Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Van Der Meer, B., 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* 91, 354–376.
- Tian, L., Xue, B., Wang, Z., Li, D., Yao, X., Cao, Q., Zhu, Y., Cao, W., Cheng, T., 2021. Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection. *Remote Sens. Environ.* 257, 112350.
- Wang, D., Vinson, R., Holmes, M., Seibel, G., Bechar, A., Nof, S., Tao, Y., 2019. Early detection of tomato spotted wilt virus by hyperspectral imaging and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN). *Sci. Rep.* 9, 4377.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402.
- Yan, J., Tan, B.C., 2019. Maize biology: From functional genomics to breeding application. *J Integr Plant Biol* 61, 654–657.
- Zarco-Tejada, P.J., González-Dugo, V., Berni, J.A.J., 2012. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sens. Environ.* 117, 322–337.
- Zarco-Tejada, P.J., Camino, C., Beck, P.S.A., Calderon, R., Hornero, A., Hernandez-Clemente, R., Kattenborn, T., Montes-Borrego, M., Susca, L., Morelli, M., Gonzalez-Dugo, V., North, P.R.J., Landa, B.B., Boscia, D., Saponari, M., Navas-Cortes, J.A., 2018. Prevalent symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations. *Nat. Plants* 4, 432–439.
- Zhou, L., Nie, C., Su, T., Xu, X., Song, Y., Yin, D., Liu, S., Liu, Y., Bai, Y., Jia, X., Jin, X., 2023. Evaluating the canopy chlorophyll density of maize at the whole growth stage based on multi-scale UAV image feature fusion and machine learning methods. *Agriculture* 13, 895.