Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection

Long Tian, Bowen Xue, Ziyi Wang, Dong Li, Xia Yao, Qiang Cao, Yan Zhu, Weixing Cao, Tao Cheng *

National Engineering and Technology Center for Information Agriculture, MOE Engineering Research Center of Smart Agriculture, MARA Key Laboratory of Crop System Analysis and Decision Making, Jiangsu Key Laboratory for Information Agriculture, Nanjing Agricultural University, One Weigang, Nanjing, Jiangsu 210095, China

ARTICLE INFO

ABSTRACT

Rice blast is considered as the most destructive disease that threatens global rice production and causes severe economic losses worldwide. A detection of rice blast infection in an early manner is vital to limit its expansion and proliferation. However, little research has been devoted to spectral detection of rice leaf blast (RLB) infection, especially at the asymptomatic or early stages. To fill the gap, this study aimed to examine the feasibility of detecting RLB infection from leaf reflectance spectra at asymptomatic, early and mild stages of disease development. Greenhouse experiments were conducted over two consecutive years to collect hyperspectral data (350-2500 nm) on various days after inoculation (DAIs) for the three infection stages. These hyperspectral data were processed to select disease specific spectral features (DSSFs). Such DSSFs were then used to feed the machine learning based sequential floating forward selection (ML-SFFS) methodology for determining the optimal feature combination (OFC) and overall accuracy (OA) in the detection of RLB at various infection stages. The results demonstrated that the rice plants displayed considerable biochemical and spectral variations and this pattern of variations existed consistently during plant-pathogen interactions. A multivariate pool of DSSFs comprising two reflectance bands, fourteen SIs, and five continuous wavelet coefficients, were determined for revealing the dynamic response of RLB infection across two years. The combination of 2 to 4 spectral features selected by the ML-SFFS algorithm was sufficient to identify infected leaves with classification accuracies over 65% and 80% for the asymptomatic and early infection stages, respectively. The OA could rise up to 95% for the mild stage. Compared to the use of all DSSFs with a support vector machine (SVM) classifier, the SVM-based SFFS (SVM-SFFS) algorithm prevailed in the classification accuracy up to 10% over the sampling period. Our results demonstrated the feasibility of accurate classification of RLB infected samples by ML-SFFS. This study suggests that reflectance spectroscopy has great potential in the pre-visual detection of RLB infection and airborne or spaceborne imaging spectroscopy is promising for the mapping of early occurrence and severity levels of RLB infection at large scales.

1. Introduction

Rice serves as the staple food for half of the world's population (Sharma et al., 2012). Rice blast (RB), caused by the fungal pathogen *Magnaporthe oryzae*, is considered as the most devastating and widespread disease in the rice cultivation regions worldwide (Deng et al., 2017). RB outbreaks can occur throughout the entire growing season of rice in cool and rainy summers (Fang et al., 2019; Kobayashi et al., 2001). Since this disease is highly contagious, the abundant conidia from disease lesions could spread quickly from plant to plant by wind or water under the conditions of high humidity and insufficient sunlight (Liu et al., 2019b; Talbot, 2003). Eventually, significant losses in rice production and environmental

---

* Corresponding author.
 E-mail address: tcheng@njau.edu.cn (T. Cheng).

damage could occur as a result of large-scale outbreaks of RB disease, especially in Asia as the largest rice production region in the world (Wilson and Talbot, 2009). It is estimated that the annual loss of rice yield due to the RB disease is enough to feed 60 million people (Dean et al., 2005; Wilson and Talbot, 2009; Yang et al., 2014). Dependent on the infected organs, the fungal pathogen *Magnaporthe oryzae* could cause two major types of pathosystems, namely rice leaf blast (RLB) and rice panicle blast (RPB). The pathogen infection mainly occurs on the leaves at tillering and jointing stages and on the panicles after heading (Kobayashi et al., 2001). Given the direct effect of RPB on grain yield and quality, previous research has investigated its spectral responses and accurate classification through detached samples (Kobayashi et al., 2001; Liu et al., 2010; Wu et al., 2009). However, the pathogens leading to panicle infection result from the spores formed on the leaf lesions (Kobayashi et al., 2001). This pattern of infection strengthens the importance of early detection of RLB due to its close association with RPB. Moreover, when pathogen infection occurs on leaves, the leaf tissue is severely damaged and photosynthesis is adversely affected (Talbot, 2003).

Accurate detection of RLB at the early infection stage is critical for containing the expansion and reducing the potentially larger impacts. Traditional methods for RLB detection mainly rely on visual inspection by experts in the field. This approach requires a large number of experienced professionals and is time consuming and labor intensive. Additionally, it is difficult to detect early RLB infection under field conditions when no symptoms are visible on the leaves. Therefore, a more effective approach is essential for accurately detecting RLB at an early stage. Remote sensing has been proved to be an effective and non-destructive technique for detecting crop diseases across different spatial scales (Franceschini et al., 2019; Huang et al., 2014; Mahlein et al., 2013; Rumpf et al., 2010; Shi et al., 2018b; Zhang et al., 2012b). Although a few studies have examined the estimation of RLB disease severity and the identification of infected leaf samples at late symptomatic stages (Feng et al., 2009; Zhou et al., 2014), little research has been devoted to the spectroscopic detection of RLB at the early infection stage, accompanied with examinations of leaf biochemical properties. Some studies have demonstrated that the early responses of plant biotic stress, which manifested as alterations in plant physiological parameters or photoprotective mechanisms, could be detectable using sensitive spectral indicators (Cheng et al., 2010; Hardisky et al., 1983; Huang et al., 2014; Mahlein et al., 2013; Morel et al., 2018; Zarco-Tejada et al., 2018; Zhou et al., 2018). The minor spectral and physiological alterations could pose great challenges for the unambiguous identification of infected samples at the early stage (Cheng et al., 2010; Mahlein et al., 2013).

Reflectance spectroscopy is favorable for revealing the subtle spectral signals, which are induced by complex plant-pathogen infections and could span a wide range of the electromagnetic spectrum (Mahlein et al., 2018). For some applications, the individual reflectance narrowbands are found to significantly improve the classification of various vegetation types over broadbands (Thenkabail et al., 2004). As a simple and effective tool for characterizing spectral variations, SIs were also used to highlight factors of interest and suppress the influence of other factors and represent a popular tool for plant disease detection (Calderón et al., 2013; De Castro et al., 2015; Huang et al., 2007; Mahlein et al., 2010; Penuelas et al., 1995; Zhang et al., 2012a; Zhang et al., 2019).

Some studies focused on developing disease-specific SIs for various crops (Huang et al., 2014; Mahlein et al., 2013), since the diagnostic spectral responses may be dependent on the disease (Mahlein et al., 2018; Mahlein et al., 2019). In contrast, other studies transformed spectral reflectance in different ways, such as spectral derivative and continuous wavelet transformation, to enhance the subtle spectral signals induced by pathogen infection (Cheng et al., 2010; Luo et al., 2013; Shi et al., 2018a; Zhang et al., 2014; Zhang et al., 2012b). For example, Cheng et al. (2010) developed a novel wavelet-based approach for the spectroscopic detection of green attack damage in lodgepole pine needles. However, difficulties still existed in these studies when distinguishing infected and healthy samples at the early stage using a single type of spectral features.

Besides the individual spectral indicators, physically-based approaches based on radiative transfer models (RTMs) have also gained growing attention in plant disease and pest detection at different scales (Al-Saddik et al., 2018; Badnakhe et al., 2018; Calderón et al., 2013; Hernández-Clemente et al., 2017; Hornero et al., 2020; Li et al., 2020b; Lin et al., 2018; Morel et al., 2018). One of the main merits for RTM-based approaches is that the plant-light interactions could be described using leaf- or canopy-level physical models (Jacquemoud et al., 2009; Jay et al., 2016; van der Tol et al., 2009). Therefore, the inversion of RTMs enables the detection of changes in plant functional traits. For example, recent studies have assessed the relative importance of various plant traits retrieved using RTMs in the detection of *Xylella fastidiosa* (Xf) infection over olive orchards, and identified the anthocyanin content and solar–induced chlorophyll fluorescence (SIF) as the most important traits (Poblete et al., 2020; Zarco-Tejada et al., 2018). Similarly, the anthocyanin (Morel et al., 2018) and chlorophyll (Lin et al., 2019) contents inverted with the RTMs model from hyperspectral images were found to be particularly sensitive to the occurrence of disease at different scales. Although the RTM-based approaches have exhibited superiority in interpretability and mechanistic modeling on the detection of pathogen infection induced spectral alterations, the limitation stemming from the computational burden of model inversion hampers the efficiency of plant disease monitoring.

Plant-pathogen interactions are considered to be complex processes that lead to changes in multiple physiological or biochemical properties of crops. Therefore, the spectral responses to a crop disease could be complex and may not be characterized by a single type of spectral indicators such as SIs. Recently, much attention has been paid to the integration of multiple sensitive spectral indicators as an effective approach for early detection of crop diseases (Behmann et al., 2015; Huang et al., 2014; Mahlein et al., 2019; Mahlein et al., 2013; Rumpf et al., 2010; Shi et al., 2018a; Zarco-Tejada et al., 2018). A simple approach is to combine the most relevant wavelengths with sensitive spectral indices for developing disease specific spectral indices (Huang et al., 2014; Mahlein et al., 2013). Multiple sensitive SIs could also be automatically fed to classification models for disease detection (Behmann et al., 2015; Rumpf et al., 2010; Shi et al., 2018a; Zhang et al., 2012b).

Recently, several machine learning algorithms fed by a pool of remotely sensed plant functional traits were successfully applied to the pre-visual detection of *Xylella fastidiosa* infection symptoms (Poblete et al., 2020; Zarco-Tejada et al., 2018). Other studies have also demonstrated the merits of machine

learning with different sensitive indicators as the input for plant disease detection (Behmann et al., 2015; Chemura et al., 2016; Ghosal et al., 2018; Zhang et al., 2019). Regardless of the number of input variables, these studies supplied all the spectral features or retrieved biochemical/biophysical traits to the models for disease classification or regression. Obviously, they made no attempt to optimize the combination of input variables for more efficient detection of plant diseases nor for better understanding the contribution of pertinent spectral features to disease detection. Previous studies have demonstrated the combination of feature subsets can significantly improve the efficiency of classification models without seriously sacrificing the classification accuracy (Hamed et al., 2020; Hu et al., 2019; Huang et al., 2019). However, many studies have found that the performance of machine learning algorithms could be significantly affected by the number of input variables or spectral features (Fallahpour et al., 2017; Hu et al., 2019; Kaushal and Swarnajyoti, 2018). In particular, machine learning classifiers have been integrated with feature selection algorithms to improve the efficiency and classification performance in fluorescence spectroscopy for nucleotide identification (Huang et al., 2019). Although their machine learning and sequential floating forward selection (ML-SFFS) methodology exhibited the potential for fluorescence spectrum shape analysis, it has not been applied to reflectance spectroscopy of vegetation for disease detection purposes. In addition, the relative importance of each indicator in the input data may vary with disease infection stage (Poblete et al., 2020; Zarco-Tejada et al., 2018). Thus, it remains unclear how the optimal combination of multiple spectral indicators affects the performance of disease detection at different stages of RLB infection.

The overall goal of this research was to develop a classification methodology with ML-SFFS for the spectroscopic detection of RLB infection at various stages by separating healthy leaves from infected leaves. Thus, the specific objectives were: (1) to examine the responses of leaf biochemical properties to RLB infection and the physiological foundation of spectroscopic detection, (2) to determine the disease-specific spectral features (DSSFs) in the form of reflectance bands, SIs and wavelet coefficients based on their spectral separability between healthy and infected leaves, and (3) to evaluate the ML-SFFS algorithms in terms of classification accuracy and optimal feature combination (OFC) at asymptomatic, early and mild infection stages.

2. Material and methods

2.1. Experimental design

2.1.1. Rice planting and artificial inoculation

The greenhouse experiments were conducted in 2018 and 2019 at the Pailou experimental base of Nanjing Agricultural University located in the middle of Nanjing, Jiangsu province, China (118°51′ E, 31°57′N). The rice seeds were first sowed in the field and then transplanted into plastic pots (three plants per pot) on June 25 of 2018 and June 10 of 2019. Three rice varieties were grown in plastic pots (35 cm diameter $\times$ 32 cm height) filled with paddy clay soil. A total of 36 (12 pots per variety) and 24 (8 pots per variety) pots were used for the biochemical and spectral measurements in 2018 and 2019, respectively. Data collected from the leaves of multiple varieties in different years provided an opportunity to evaluating the reproducibility and reliability of the findings on disease detection. To ensure local management practices in the greenhouse

environment, the basal nutrition fertilizers (Nitrogen provided by urea, 150 kg·ha^{-2}; P_2O_5, 135 kg·ha^{-2}; and K_2O, 18.3 kg·ha^{-2}) were applied prior to transplanting and a second nitrogen topdressing (N, 150 kg·ha^{-2}) was applied during the tillering stage. The rice plants were irrigated as needed to ensure that the soil in each pot is always covered by a shallow layer of water. All plants were placed outdoors and were not transferred to the greenhouse until one week before the inoculation treatment. The greenhouse comprised two layers of transparent materials and equipped with air conditioning and humidifying facilities to provide suitable environmental conditions (26–32 °C, relative humidity of over 90% and a photoperiod of 14 h) for the artificial inoculation of RLB pathogen. The inoculation treatments were carried out at the tillering stage dating August 8 of 2018 and July 15 of 2019, respectively.

The artificial inoculation operations were conducted to induce RLB infection on rice leaves in 2018 and 2019. A mixture of various blast pathogens, which were provided by a rice breeding team from State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, was used to inoculate the plants. A conidial suspension with 1×10^5 *Magnaporthe oryzae* conidia per ml was sprayed on the leaves uniformly under greenhouse conditions. After inoculation, all plants were completely covered with black, light-tight plastic materials for 48 h. Temperature (26–32 °C), relative humidity (≥90%) as well as light conditions were strictly controlled to provide suitable conditions for successful leaf infection.

2.1.2. Definition of the disease infection process

The disease severity is defined to characterize the different developmental phases. Previous studies often defined the disease severity at leaf level as the average percentage of infected surface area (Jr et al., 1991; Wolf and Verreet, 2002). However, the disease severity was also assessed by visually inspecting the samples at various scales to better describe the disease infection process (Bock et al., 2020; Zarco-Tejada et al., 2018). Although the infection process was not entirely consistent across all leaves, the overall infection processes were roughly defined by visually inspecting every sample for symptoms of disease lesions in this study. The average disease severity was assessed on four levels according to the size and status of leaf lesions. Specifically, the asymptomatic stage was defined as the period when obvious disease lesions had not yet appeared on infected leaves (Fig. 1A). The leaves that eventually exhibited obvious and stable spindle disease lesions were classified as infected samples although the spots were invisible to the naked eyes at the asymptomatic stage (Fig. 1A). The leaves with a clean surface during all sampling days were classified as healthy samples. The early infection stage indicated that a few watery lesions (usually only one or two) could be observed on infected leaves in the greenhouse, but are often not easily noticed under field conditions (Fig. 1B). The mild infection stage was defined as the period when several small brown spindle lesions could be easily observed on the surface of infected leaves (Fig. 1C). Generally, the leaf surface at this stage appeared normal beyond the necrotic lesions. The severe infection stage was characterized by the multiple distinct fusiform plaques formed on the leaf surface and the signs of wilting and yellowing surrounding the lesions (Fig. 1D). Continuous visual inspection of the same leaf was performed to ensure that no samples were misclassified during the whole test period. Since our focus was on the detection of RLB infection at the asymptomatic, early, and mild stages, samples from the
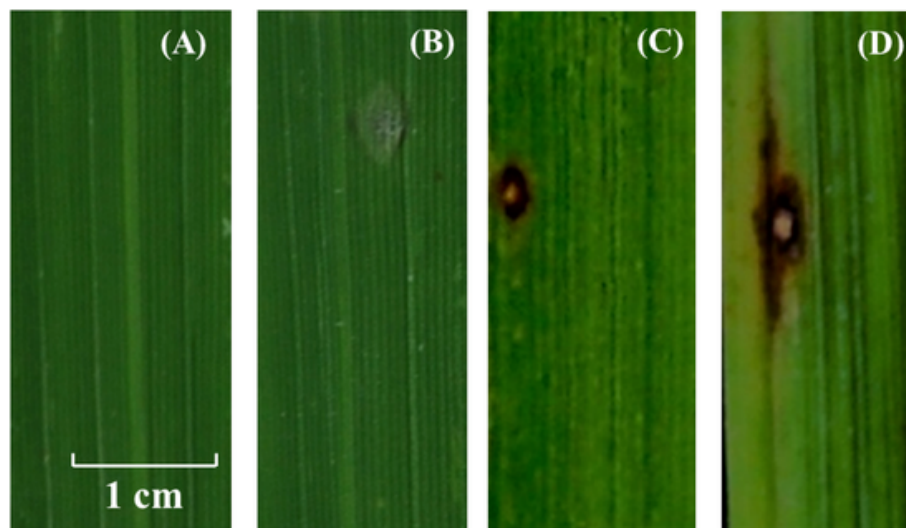
Fig. 1. True colour images of rice leave under various disease severity levels (A. asymptomatic; B. early infection; C. mild infection; D. severe infection) as defined in this study.

severe infection stage were not included for subsequent analysis.

In 2018, the representative plant samples were selected prior to inoculation and then measured from asymptomatic (days after inoculation, DAI, 1, 3, 5, 7, 10) to early infection stages (DAI 12, 17, 21). In 2019, the spectral and biochemical measurements were carried out from the early infection (DAI 8, 11, 16) to the mild infection stage (DAI 20, 28, 35).

2.1.3. Experiment setup

To determine the spectral and biochemical variations on the same infected leaves, we planted 36 pots of three susceptible rice varieties (*Wuyungeng 7, Nipponbare,* and *Nangeng 44*) for data collection from asymptomatic to the early infection stages in 2018. All samples were subjected to the same growth conditions, including environmental setup and nutrient levels. For each plant, three to five fully expanded leaves at the same leaf-order were selected and labeled with colour markers. Continuous biochemical and spectral measurements were then conducted on the same labeled leaves from DAI 1 to DAI 21. At the same time, RGB images of each sample were collected to according to Li et al. (2019). Eventually, infected leaves were observed in 26 of the 36 pots. The average spectral and biochemical parameters were calculated separately over all infected and healthy leaves in each pot for every DAI. Finally, 26 pairs of samples (26 infected and 26 healthy samples) were collected per sampling day (*Table S1*).

In 2019, a total of 24 pots of three susceptible rice varieties (*Wuyungeng 7, Wuyungeng 23,* and *Nangeng 44*) were grown in the same way as in 2018 Experiment for sampling from early to mild infection stages. With regard to rice variety, *Nipponbare* used in 2018 was replaced with *Wuyungeng 23* in 2019 because *Nipponbare* leaves were found to be very fragile during the preparation for biochemical and spectral measurements. In addition, all of the measurements were conducted on the independent leaves on each sampling day from early (Fig. 1B) to mild infection stages (Fig. 1B) for obtaining a broad diversity of infected samples. Three pairs of infected and healthy leaves at the same leaf-order were randomly selected from each pot on each sampling day for spectral and biochemical data collection. Finally, spectral and biochemical data were collected from 45 pairs of infected and healthy samples from 15 pots (no

infected leaves were observed in the other nine pots) (*Table S1*). The dates of measurements in the two years were determined by the expansion of the disease spots on the leaves. Therefore, the dates of taking measurements and the interval between the two measurements were inconsistent between 2018 and 2019 to maintain reasonable workload.

2.2. Data collection

2.2.1. Measurements of leaf reflectance spectra

The FieldSpec 4 Hi-Res spectroradiometer (Analytical Spectral Devices, Boulder, USA) equipped with a leaf clip was used to measure the leaf reflectance for the 2018 and 2019 experiments. The spectral sampling interval of the collected spectra is 1.4 nm in the 350–1000 nm region and 1.1 nm in the 1001–2500 nm region. A contact fiber-optic cable (1.5 m) coupled with an array detector (with a field of view FOV of 25°) of the FieldSpec 4 was utilized to capture light reflected from the target. The leaf clip provides a white reference panel with a reflectance of about 99% and a black background panel with a reflectance of lower than 1%. The black background panel is mainly used as the background for sample testing while the white reference panel is used to perform dark current correction and optimization. With the halogen lamp on, dark current correction and optimization are performed with the probe facing the white reference panel to collect instrument noise and optimize integration time, respectively. These operations were done automatically by the instrument. For each rice leaf, three reflectance spectra were obtained from the adaxial surface at the 1/3, 1/2, and 2/3 distance from the base, respectively. The main veins of rice leaves were excluded from the chamber of leaf clip during spectral collection. All spectra collected from the same leaf were averaged to represent the mean reflectance of this leaf and to minimize the random error during measurements.

2.2.2. Measurements of leaf biochemical variables

To illustrate the biochemical alterations against the fungal invader, four parameters chlorophyll (C_{a+b}), carotenoid (C_{x+c}), anthocyanin (C_{anth}) and water contents (C_w) were obtained after spectral measurements on the same infected and healthy leaves across the two years. C_{a+b} was determined

non-destructively using a Dualex device (Dualex Scientific+, Force-A, Orsay, France). The mean value of three Dualex readings taken from the 1/3, 1/2, and 2/3 distance from the leaf base was considered to be the final measurement. The C_{x+c}, C_{anth} and C_w were estimated from reflectance spectra using the PROCWT model, which couples the PROSPECT-D model (Féret et al., 2017) and continuous wavelet transform (CWT) and was proposed in our recent work (Li et al., 2018). The PROCWT method has advantages of improving the retrieval of leaf biochemical parameters by reducing the specular reflection effect and enhancing the absorption features of chemical constituents. Additionally, the accuracy of PROCWT for rice biochemical contents has been recently validated using laboratory measured leaf chemistry (Li et al., 2018). Therefore, it is well suited for retrieving biochemical contents at high accuracies from bidirectional reflection factor spectra, where were collected with leaf clips as used in this study. To alleviate the ill-posed problems of model inversion, the possible parameter ranges were set according to prior information provided by our previous research. Details about the setup of nominal values and ranges of parameters for PROCWT inversion can be found in Table S2.

2.3. Disease detection methodology

The disease detection methodology was developed based on the hyperspectral and biochemical measurements collected from 26 pairs of identical samples and 45 pairs of different samples over various DAIs of two years. The detection of RLB infection at different stages comprised three main modules: data preparation, feature selection, and leaf classification (Fig. 2).

The significance in the difference of leaf biochemical parameters between the two classes (healthy vs. infected) over different DAIs were assessed using statistical test analysis (Table S3). For each biochemical parameter from the same DAI in the datasets of 2018, a paired Student's t-test (H_0: Infected = Healthy; H_A: Infected $\neq$ Healthy; $p < 0.05$) was performed under normal distribution and a Wilcoxon test under non-normal distribution. In order to evaluate the significant difference in leaf biochemistry between healthy and infected leaves in 2019, an independent t-test (H_0: Infected = Healthy; H_A: Infected $\neq$ Healthy; $p < 0.05$) was applied for samples with uniform variance and normal distribution. The samples that did not meet the requirements were evaluated using a Mann-Whitney U test. All outliers were determined using an interval spanning over the median minus/plus three standard deviations and eliminated before the statistical significance analysis (Leys et al., 2013). The samples with a significant difference in any of the four biochemical parameters were retained for the extraction of DSSFs.

A large number of narrow-band spectral indicators were generated as the candidate features for extracting the DSSFs sensitive to RLB infection, including reflectance bands, SIs *(Table S4)*, and wavelet coefficients. A total of 97 SIs of five categories (*Table S4*) selected from the literature were calculated as the candidates for feature selection. Given the limitation of band number in SIs, the potential of hyperspectral data could not be fully exploited. Therefore, full-band spectral information was also included to explore the subtle spectral variations caused by RLB infection. In addition to selecting features from the reflectance bands, we also performed continuous wavelet transform on the spectra. Sensitive wavelet features
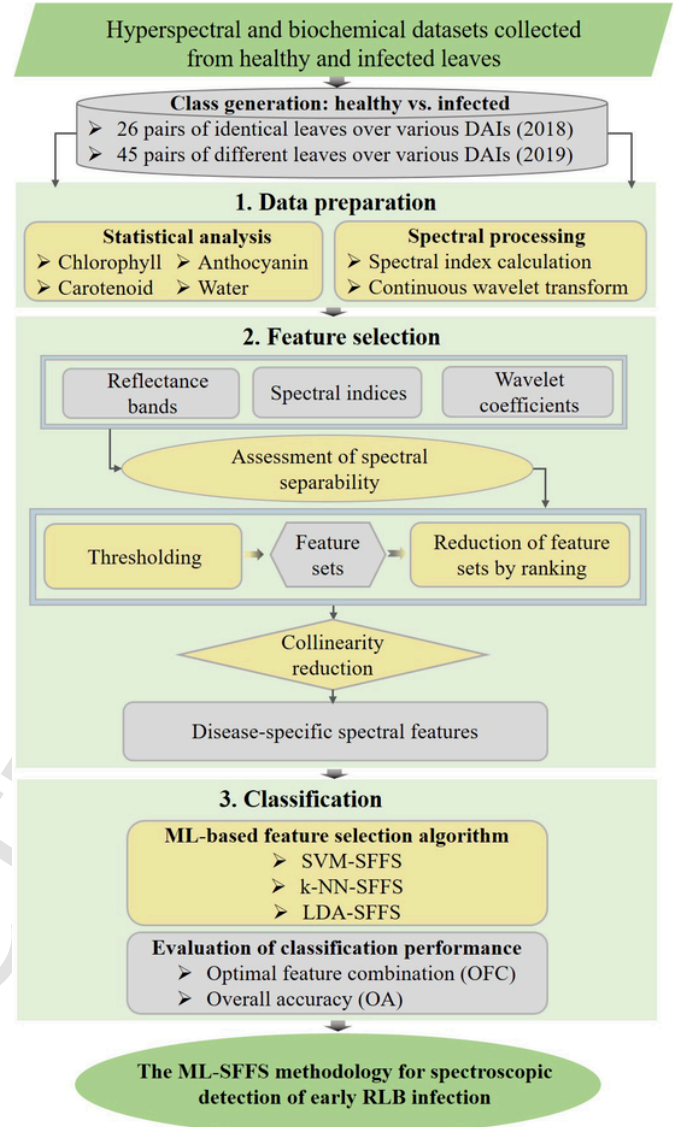

Fig. 2. Schematic illustration of ML-based SFFS classification methodology for the spectroscopic detection of rice blast.

were further extracted from wavelet coefficients over various scales.

When performing continuous wavelet transform (CWT), the reflectance spectrum was converted to sets of coefficients using a mother wavelet basis function. With CWT, each one-dimensional input reflectance spectrum was decomposed into a number of two-dimensional wavelet coefficient spectra by calculating the inner products of the scaled and shifted wavelet and the input spectrum. Each wavelet coefficient or wavelet feature is a function of wavelength and scale (dyadic numbers 2^1, 2^2, 2^3, ..., 2^8 were denoted as Scale 1, Scale 2, Scale 3, ..., Scale 8 for simplicity) and its amplitude is represented by wavelet power. Each scale component of wavelet coefficient is of the same length as the reflectance spectrum. Following Cheng et al. (2010), the second derivative of Gaussian function (Mexican Hat wavelet) was used as the mother wavelet basis and only the wavelet coefficients at Scales 3 to 8 were retained for feature extraction. CWT operations were performed in the IDL8.3 Wavelet Toolkit (Exelis Visual Information Solutions, Boulder, CO, USA).

A large number of redundant features of hyperspectral data were reduced by extracting the consistent feature subsets, which are highly related to RLB infection. The separability of spectral features between infected and healthy leaves can be determined through the thresholding technique (Mahlein et al., 2013). For a specific spectral feature, the mean values of the infected (M_i) and healthy (M_h) samples were separately calculated and the distance between M_i and M_h was divided equally into 100 proportions. Finally, the optimal threshold was defined as the one with the highest classification accuracy among the 100 values (Fig. 3). The separability assessment was applied to all spectral features individually.

After clarifying the separability of each spectral feature, the consistent sensitive features over different DAIs were extracted. For reflectance bands and wavelet features, separability scalograms were generated according to Cheng et al. (2010) for visualizing the distribution of extracted spectral features. Unlike the correlation measure used in Cheng et al. (2010), the separability was used in this study since it matched our purpose of detecting infected samples. Each separability scalograms reports the classification accuracy from 0 to 100% at each wavelength (and scale for wavelet features). The features displaying the highest classification accuracies at the separability scalograms for each DAI were retained as the sensitive features with a cut-off percentage threshold (top 5%). Then the intersection of the retained sensitive features was determined to extract the common features. Only one representative feature with the highest separability within each common feature region was retained to extract the common DSSFs. After the feature extraction operations, a multivariate pool of disease-related spectral indicators comprising a smaller number of reflectance bands, SIs, and wavelet coefficients, were generated as input variables for the disease detection methodology.

The SFFS is a bottom-up search procedure proposed by Pudil et al. (1994) to overcome the problems of computational complexity and to guarantee the optimality of the selected features. The machine learning classification algorithm was integrated into the feature selection model as the criterion function in ML-SFFS methodology (Huang et al., 2019; Kempeneers et al., 2004). Before the classification operation with ML-SFFS, the collinearity of features in the DSSF pool was reduced by the variance inflation factor (VIF) analysis. The features with VIF values less than 10 will be retained and used as the input variables of the ML-SFFS algorithm. Specifically,
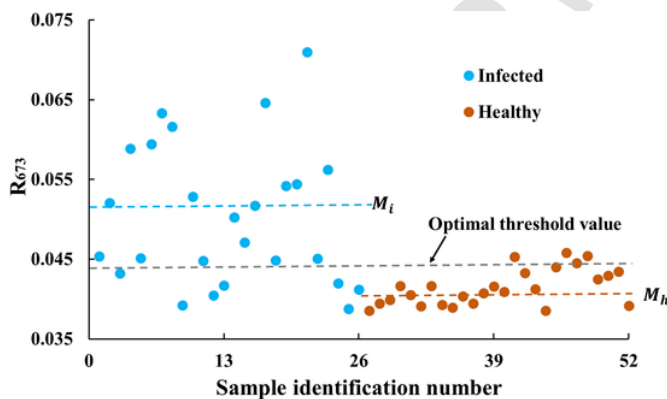
there are three major steps for the ML-SFFS algorithm. Firstly, the most significant features were selected from the input variables to form the feature subset according to the accuracy of the selected features. Secondly, the least significant feature from the generated feature subset was identified. And it would be removed unless it was selected at the first step. Lastly, elimination of the least significant feature was then continuously conducted to see if the OA can be improved at that level. Eventually, the OFC with the highest classification accuracy could be selected by the SFFS algorithm. By means of conditional inclusion and exclusion, this algorithm allowed for correcting wrong decisions made in the previous process so as to approximate the optimal solution as much as possible (Pudil et al., 1994).

Three widely used machine learning classification algorithms LDA, k-NN, and SVM were determined as criterion function to guide the feature selection processes. The LDA, also known as Fisher discriminant (Fisher, 1936), is a discriminant method to search for the optimal projections using the separability between classes (Kempeneers et al., 2004). It assumes that different class variables are normally distributed. Since there are only two classes (infected and healthy), the selected feature set was projected into a one-dimensional space. k-NN is a non-parametric classification method that classifies the unlabeled samples by analyzing its k-nearest neighbors with class labels. The classification performance of k-NN depends significantly on prior knowledge and the metric for calculating the distances between k samples (Weinberger et al., 2009). Specifically, we used the Minkowski distance between three samples to determine the attribution of samples to be classified. SVM is a non-parametric supervised classifier which reduces the misclassification errors on training data by minimizing structural risk strategies. It was originally designed for two-group classification problems by constructing an optimal separating hyperplane to maximizes the margin between classes with a small number of support vectors (training samples). Compared with k-NN and LDA, SVM is a soft classification strategy with stronger generalization ability and better robustness (Cortes and Vapnik, 1995; Pal and Foody, 2010; Wei et al., 2017). A free library (LibSVM) with radial basis function (RBF) was used in our work to perform the SVM classification task (Chang and Lin, 2011). We use a grid search strategy to determine the model parameters C and γ, which are the spread of the RBF kernel and the regularization parameter, respectively.

A 4-fold cross-validation was conducted with 100 repetitions for each ML-SFFS (LDA-SFFS, k-NN-SFFS, SVM-SFFS) classification algorithm. The average of optimal classification accuracies over the 100 repetitions was determined as the OA. Meanwhile, the most frequently selected OFC among the 100 repetitions was retained. The performance of this ML-SFFS methodology was evaluated through OA for the separation of infected and healthy samples at various infection stages. All classifications were performed in the R environment (Version 4.0.1; R Development Core Team). We performed the LDA, k-NN and SVM classifications using the packages "caret"(Kuhn et al., 2020), "kknn"(Schliep et al., 2016) and "e1071"(Meyer et al., 2020) in R, respectively.

3. Results

3.1. Leaf biochemical variation over various days after inoculation

Fig. 4 shows the temporal variations in leaf biochemistry over different DAIs during the years of 2018 and 2019. Signifi-
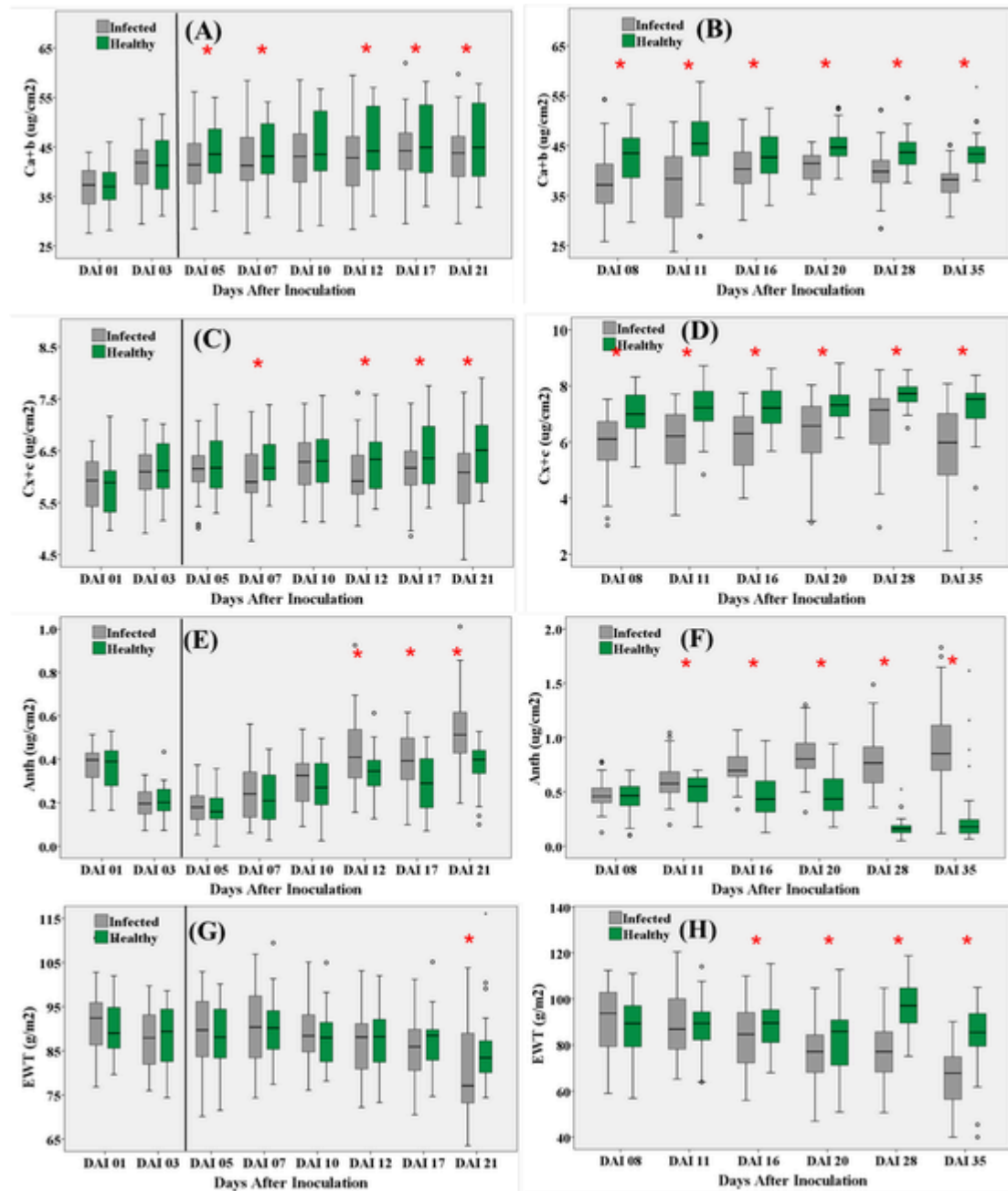

Fig. 3. A conceptual diagram of evaluating the separability of a specific feature (reflectance at 673 nm in DAI 21 in 2018) based on the thresholding technique. The M_i and M_h represents the mean values of the infected and healthy samples, respectively.

Fig. 4. Comparison of leaf biochemical parameters between the infected and healthy leaves (A, B: chlorophyll content, C, D: carotenoid content, E, F: anthocyanin content, G, H: water content) measured in the datasets of 2018 (left column) and 2019 (right column) on various DAIs. In each boxplot, the top edge, black line, and the bottom edge of the box represent the upper (Q3), median (Q2), and lower (Q1) quartiles, respectively. The whiskers represent the maximum (Q3 + 1.5*IQR) and minimum (Q1–1.5*IQR) valid values defined by interquartile ranges (IQR = Q3-Q1), respectively. The circles and dots outside the boxplot represent outliers and extreme outliers, respectively. The vertical solid lines in black denote the first day when visible disease spots appeared on infected leaves. The red asterisk marks at the top of the pairs of boxplots represent the significant differences in the biochemical variables between healthy and infected samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cant differences regarding biochemical parameters between infected and healthy leaves were observed at different infection stages in both years. Note that the sampling DAIs with no significant differences in biochemical parameters would have been discarded in subsequent spectral analysis. Compared with healthy ones, infected leaves exhibited significant decreases in total chlorophyll, carotenoid, and water contents, and a contradictory trend in anthocyanin content. The first significant

deficit in the chlorophyll and carotenoid contents of the infected leaves occurred on DAI 5 and 7, respectively. These biochemical variations were prior to the appearance of disease spots on most of the tested leaves at the early infection stage in both years (Fig. 5). Significant differences in anthocyanin and water contents were absent at the early infection stage. The critical photosynthetic pigments (chlorophyll and carotenoid) were most sensitive to the early infection of RLB disease in
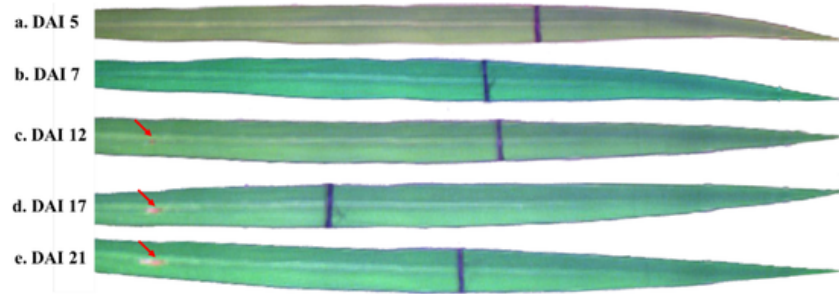
Fig. 5. Spatiotemporal dynamics of RLB disease lesions on the rice leaves from asymptomatic (a-b) to the early infection stage (c-e) of 2018. The red arrow points to the watery disease lesion. The black line cross the leaf is the fine wire made of plastic that was used to fasten the leaf for taking pictures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both years, but the response of water and anthocyanin contents lagged behind. Note that the development of disease lesions on the leaves in the greenhouse was relatively slow due to the resistance of rice varieties and environmental factors. In addition, continuous visual inspection revealed that the lesions remained stagnant after reaching a certain morphology and rarely caused the death of the entire leaf.

3.2. Selection of disease-specific spectral features (DSSFs) based on spectral separability

The classification scalograms produced with the reflectance data collected in 2018 and 2019 are displayed in Fig. 6. The

individual bands highly sensitive to RLB infection from the asymptomatic to early infection stages of 2018 were observed mostly in the VIS (400–700 nm) and a SWIR (1850–2450 nm) region (Fig. 6a-e). Specifically, the individual bands located near the red region were consistently found from DAI 7 to 21 and the SWIR bands located near the water absorption center (1945 nm) could be observed from DAI 12 to 21. Fig. 6f-k displays the highly sensitive individual bands to RLB infection from the early to mild infection stages of 2019, which are mainly observed in the VIS region. Several spectral bands located in the red region were determined consistently throughout all DAIs. On the contrary, the sensitive SWIR bands could only be observed for late DAIs at the mild infection stage.
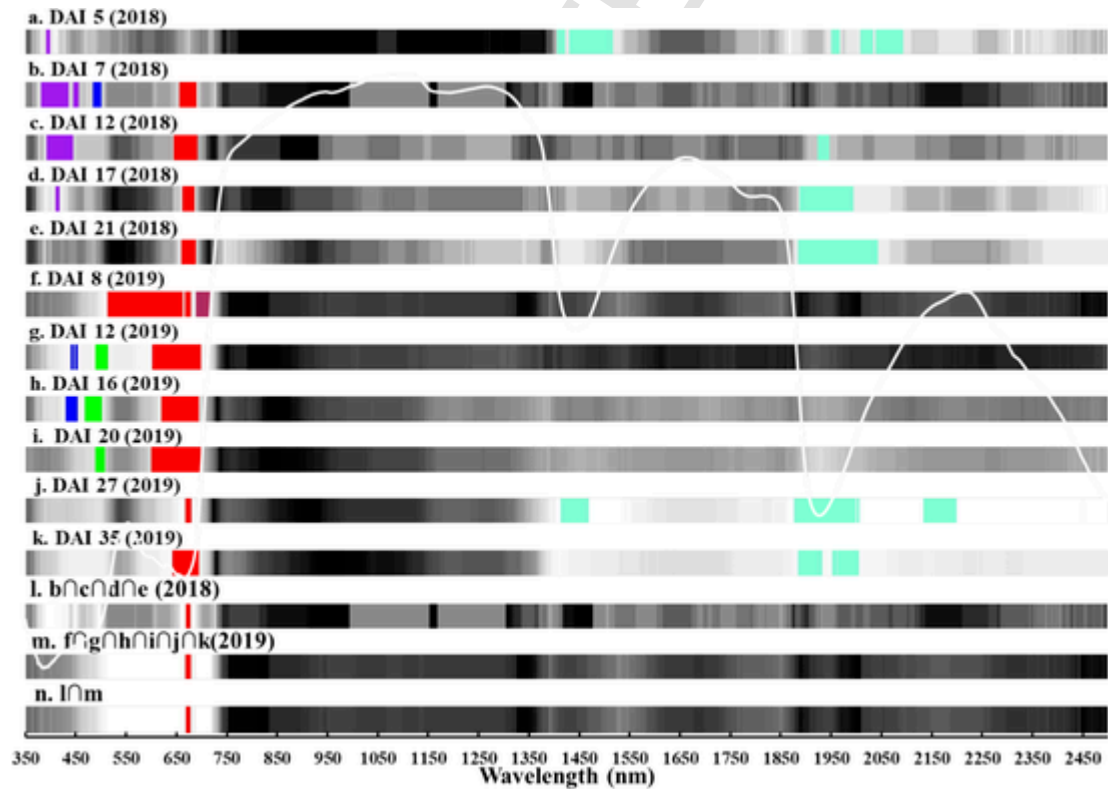

Fig. 6. Classification scalograms derived with reflectance spectra of the leaves collected from DAI 5 to 21 in 2018 (a-e) and DAI 8 to 35 in 2019 (f-k), as well as the intersection of sensitive bands determined from the individual datasets of 2018 (l) and 2019 (m), and both of them (n). The X-axis is the spectral wavebands from 350 to 2500 nm. The grayscale brightness of a scalogram represents the magnitude of classification accuracy (The stronger the brightness, the higher the classification accuracy). The patches in colour in a-k represent the most sensitive features that produced the top 5% highest accuracies among all features from the scalogram. The patches in red in l-n represent the intersection of the most sensitive features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Common sensitive bands were determined from the datasets of 2018 (Fig. 6l), 2019 (Fig. 6m), and both of them (Fig. 6n). For the intersection of sensitive features from different DAIs, the individual bands selected from the early infection stage in 2018 (Fig. 6l) aligned perfectly with those of 2019 (Fig. 6m). The sensitive bands near 1945 nm were not overlapping but appeared on several DAIs at the early infection stage of 2018 (Fig. 6c-e) and the mild infection stage of 2019 (Fig. 6j-k). Therefore, several consistent individual bands located at the chlorophyll absorption center (near 673 nm) and the water absorption center (1945 nm) were selected from different DAIs in both years. The former is highly associate with the chlorophyll content and is generally used for detecting the chlorophyll content while the latter is highly related to water absorption.

Out of the 97 SIs (*Table S4*), ten were determined as the most sensitive to RLB infection for 2018 and 2019 (Table 1). Six of them were common to both datasets, namely CCI, PSSRa, PSNDc, PRI_{m2}, SRPI, and HI_2013. The chlorophyll-related indices accounted for the largest proportion (4 out of 10 in 2018 and 7 out of 10 in 2019), followed by carotenoid-related indices (4 out of 10 in 2018 and 4 out of 10 in 2019). Two water-related (NDWI and WI) and two health-related (HI_2014 and HI_2013) indices were selected for 2018 but only HI_2013 for 2019. On the contrary, the anthocyanin-related index (ARI) and xanthophyll-related index (PRI) was selected only in 2019. All of the band combinations comprising the selected SIs were located in the VNIR region, except for the 1240 nm band for NDWI.

Most of the SIs common to both years were highly sensitive to chlorophyll or carotenoid content, or both of them. The chlorophyll/ carotenoid index (CCI) that is sensitive to the ratio of chlorophyll to carotenoid content performed best in 2018. The three simple ratio SIs (PSSRa, PSNDc, and SRPI) which are highly related to chlorophyll content share a similar spectral band near 650–680 nm. By contrast, only one carotenoid-related index (PRI_{m2}) was selected for both years. Furthermore, the plant health-related index (HI_2013) derived for identifying sugar beet diseases also showed good separability among healthy and infected leaves.

Fig. 7 shows the classification scalograms derived with wavelet coefficients in 2018 and 2019 datasets. Fig. 7a-e highlights the wavelet features strongly sensitive to RLB infection from the asymptomatic stage to the early infection stage of 2018. Consistent features can be observed in the VNIR (400–

1050 nm) and SWIR (1850–2450 nm) regions at the scales of 3 to 8 from DAI 12 to 21. The sensitive features within the SWIR region became concentrated and obvious gradually. Most of the sensitive features occurred at scales of 3 to 6 except for those located in the red edge region, which was present at all scales examined.

The classification scalograms produced with the wavelet coefficients of 2019 are displayed in Fig. 7f-k. The wavelet features sensitive to the RLB infection from the early infection to mild infection stages in 2019 mainly covered the VNIR region at scales of 3 to 7. Several sensitive features concentrated in the yellow edge region (550–650 nm) at the scales of 4 and 5 were consistently observed on all sampling days. In addition, consistent sensitive features were observed in the NIR region at the scales of 4 to 6 from DAI 16 to 35. Sensitive features that occurred in the red edge region in 2019 were located at shorter wavelengths than those in 2018 and were only observed at the early infection stage (DAI 8 to 16). For the SWIR region, sparse features were identified at the mild infection stage (DAI 27 and 35). Therefore, only the leaves from the last DAI 3 and 4 in 2018 and 2019 were involved to extract common features.

Fig. 7l-n displays the overlapping sensitive wavelet features extracted from 2018 (Fig. 7l), 2019 (Fig. 7m), and both years (Fig. 7n). The common wavelet features sensitive to RLB infection for both years mainly formed three intervals in the 550–1000 nm range. Only one common wavelet feature ($WF_{608,5}$) located in the yellow edge region (550–650 nm), which is mainly dominated by the overlapping absorption of several pigments (particularly chlorophyll and anthocyanin), was defined as the optimal wavelet features for the detection of RLB infection. Moreover, two overlapping features ($WF_{850,4}$ & $WF_{866,5}$,) located near 850 nm in the NIR region were consistently found to be the optimal wavelet feature from the early to mild infection stage. The sensitive features occurred in the NIR region are generally considered to be closely related to the leaf internal structure. Another two common wavelet feature ($WF_{939,5}$ & $WF_{960,6}$) located near 950 nm, whose reflectance is dominated by both the water absorption centered at 970 nm and the leaf internal structure, were also consistently found for all DAIs over both years. However, sensitive features that span multiple scales occurred in the red edge region (700–750 nm) and the right shoulder of the water absorption center (near 1980 nm) were only obtained from 2018.

3.3. Detection of RLB infection using ML-SFFS classification techniques

The OFCs and the corresponding classification accuracy generated by feeding the selected features, which were retained by performing VIF analysis on the DSSFs in Table 2, to each ML-SFFS classification technique are listed in Table 3. For each DAI the mean classification accuracy obtained by SVM-SFFS was consistently higher than that by LDA-SFFS or k-NN-SFFS. This superiority was highly significant over LDA-SFFS for the asymptomatic stage of 2018, and over k-NN-SFFS for 2018 and the early infection stage of 2019. LDA-SFFS performed significantly better than k-NN-SFFS for the early infection stage of 2018 and DAI 8 of 2019 but was evenly matched with k-NN-SFFS for the rest DAI. The performance of both LDA-SFFS and k-NN-SFFS was capped by SVM-SFFS for all DAIs. The maximum classification accuracies obtained by SVM-SFFS were over 80% and 83% for the early infection stage of 2018 and 2019, respectively. It rose to 98.65% for the mild infection

Table 1
List of the most sensitive SIs producing the top 10 highest classification accuracies for the separation of infected and healthy leaves in 2018 and 2019, respectively.

Rank	Top 10 VIs for 2018	Top 10 VIs for 2019
1	CCI*	PRI_{m2}*
2	HI_2014	HI_2013*
3	NDWI	SRPI*
4	WI	CCI*
5	PRI_{m1}	PRI × CI
6	PSSRa*	PRI
7	PSNDa*	ARI
8	PRI_{m2}*	PSSRb
9	SRPI*	PSSRa*
10	HI_2013*	PSNDa*

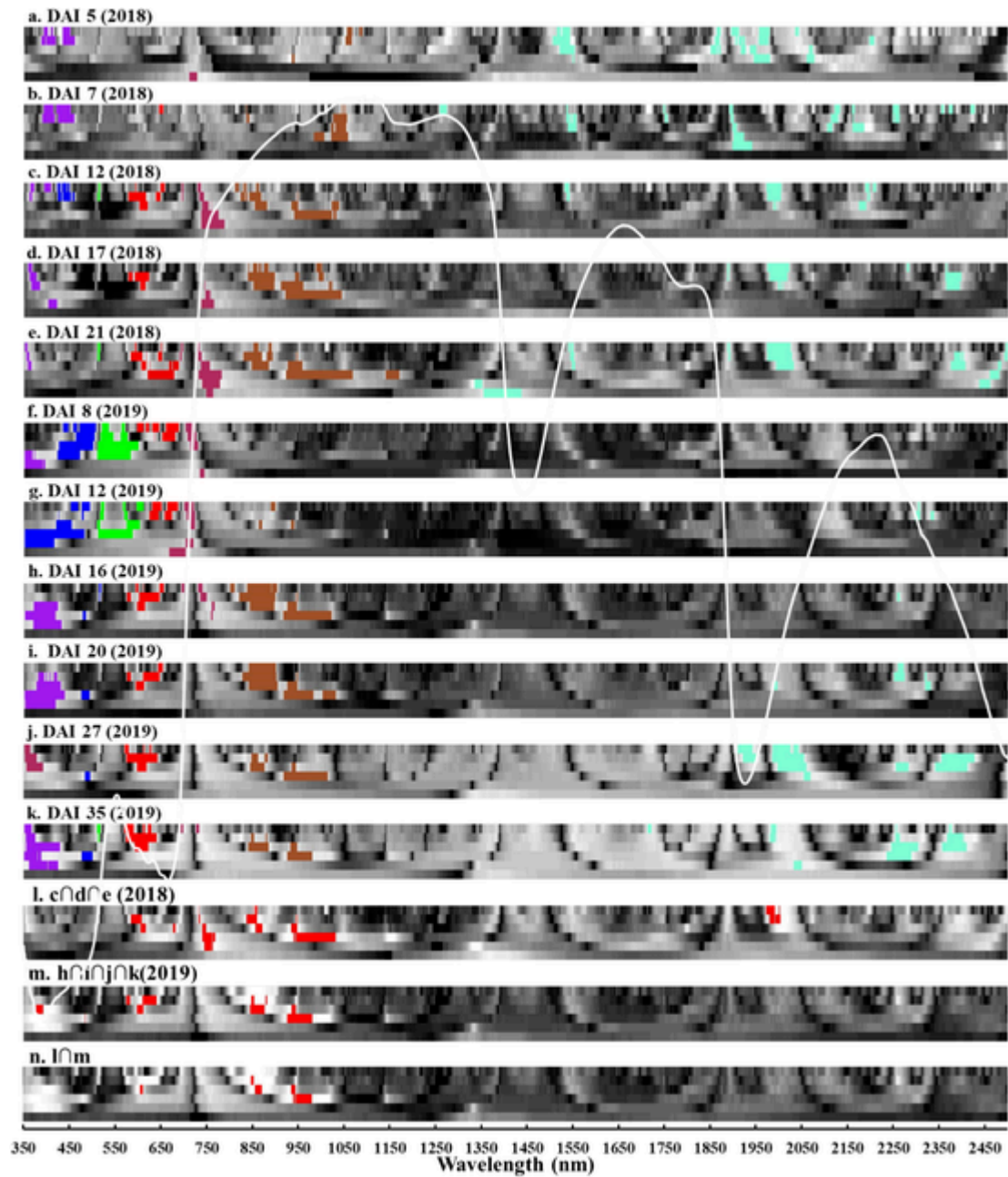Note: The asterisk marks the sensitive VIs common to two years.

Fig. 7. Classification scalograms derived with wavelet coefficients of the leaves collected from DAI 5 to 21 in 2018 (a-e) and from DAI 8 to 35 in 2019 (f-k), as well as the intersected features, extracted from 2018 (l) and 2019 (m), and both of them (n). The X-axis is the spectral wavebands from 350 to 2500 nm, and Y-axis represents the third to eighth wavelet scale. The grayscale brightness of a scalogram represents the magnitude of classification accuracy (The stronger the brightness, the higher the classification accuracy). The patches in colour in a-k represent the most sensitive wavelet features that produced the top 5% highest accuracies among all features from the scalogram. The patches in red in l-n represent the intersection of the most sensitive features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

stage of 2019. Compared with the SVM classification accuracies obtained using all DSSFs, the SVM-SFFS classification accuracies were over 10% higher for DAI 5, 7 and 12 of 2018 and approximately 5% higher for DAI 17 and 21 of 2018 and DAI 8 and 11 of 2019. However, the superiority was insignificantly for the mild infection stage of 2019 (Fig. 8).

Concerning the OFC, two to four features were retained after the SFFS procedure for each classifier and each DAI (Table 3). This number was far less than the total feature number (21) of the multivariate pool in Table 2. For each DAI the three-

feature OFCs appeared most frequently among all classification results, while two-feature and four-feature OFCs were more likely to be observed in the results obtained by LDA-SFFS and SVM-SFFS, respectively. Additionally, no significant collinearity could be observed among the features contained in each OFC. Although the OFCs selected by each procedure were not exactly the same across different DAIs, some features appeared frequently in different OFCs. For example, F6 ($WF_{936,5}$) for SVM-SFFS occurred more frequently in all DAIs than other features.

Table 2
A summary of the selected disease-specific spectral features (DSSFs).

Feature type	Feature code	Spectral feature
Spectral band	F1	R_{673}
	F2	R_{1945}
Wavelet coefficient	F3	$WF_{850,4}$
	F4	$WF_{608,5}$
	F5	$WF_{866,5}$
	F6	$WF_{936,5}$
	F7	$WF_{960,6}$
Spectral index	F8	CCI
	F9	HI_2014
	F10	NDWI
	F11	WI
	F12	PRI_{m1}
	F13	PSSRa
	F14	PSNDa
	F15	PRI_{m2}
	F16	HI_2013
	F17	SRPI
	F18	PRI × CI
	F19	PRI
	F20	ARI
	F21	PSSRb

4. Discussion

4.1. Physiological interpretation of the sensitive spectral features

Given that the variation of various plant physiological parameters could induce strong responses in specific spectral ranges, the unbalance of the spectral features selected from the VNIR and SWIR range could be attributed to the sensitivity of different physiological parameters to disease infection. A recent study conducted by Zhang et al. (2019) also confirms that the VNIR spectral features were most frequently used in the detection of plant diseases and pests. The sensitivity of spectral features in the VNIR region was linked to leaf internal structure and rapid degradations and biosynthesis of pigments, including chlorophyll, carotenoids, and anthocyanin (Jacquemoud et al., 2009; Ustin et al., 2009). Our results suggest that the significant variations in pigment contents of infected leaves resulted from the passive (for chlorophyll, carotenoid) and active (for anthocyanin) responses of plants to the RLB infection (Fig. 4). Moreover, continuous measurements at differ-ent DAIs (DAI 6 in 2018 and DAI 8 in 2019) over two years support the reliability and reproducibility of findings on the biochemical analysis. The consistent and rapid response of these pigment contents to disease infections directly leads to the selection of a large number of spectral features in the VNIR region. On the contrary, the lagged response of water content to the RLB infection is the main explanation that only scattered spectral features were captured in the SWIR region for late sampling DAIs. It could probably be explained by the late occurrence of water stress induced by RLB infection. This speculation is supported by the significant differences in water content only for late DAIs (Fig. 4 *G-H*).

Further analysis of the sensitive features could provide a deeper understanding of the spectroscopic detection of RLB. Specifically, the identified sensitive individual band R_{673} coincided with the chlorophyll absorption center, which means that changes in chlorophyll content caused by RLB infection could be captured directly with this band. The wavelet feature ($WF_{607,5}$) characterized the change in the local spectral shape of the yellow region 550–650 nm, which is mainly dominated by the overlapping absorption of several pigments, particularly chlorophyll and anthocyanin (Féret et al., 2017). As for the occurrence of a large proportion of chlorophyll and carotenoid-related SIs (CCI, SRPI, HI_2013, PSSRa, PSNDc, and PRI_{m2}) could be attributed to the high sensitivity of these two pigment molecules to RLB infection. The superior performance of the CCI and SRPI, which are found to be highly sensitive to both chlorophyll and carotenoid in the literature (Gamon et al., 2016; Penuelas et al., 1995), implies that the combination of multiple visible bands holds the potential to improve the spectral separability. Similar operations have been applied by Mahlein et al. (2013) and Huang et al. (2014) to develop sensitive SIs (HI_2013 and HI_2014) for discrimination between different diseases, respectively. The considerable spectral separability of infected from healthy leaves with HI_2013 indicates that the infection of different diseases may induce similar physiological responses from plants. The remaining SIs (PSSRa, PSNDc, and PRI_{m2}) sensitive to only a specific pigment were also found to be sufficient to detect the pigment variation induced by RLB infection. This could be explained by the strong correlation between chlorophyll and carotenoid contents in crop leaves (Feret et al., 2011).

Unlike the early responses of pigment-related SIs, the absence of sensitive features in the SWIR region at early infection stage is consistent with the lack of statistical difference in water content between infected and healthy samples. A recent study by Wahabzada et al. (2015) demonstrated that the loss
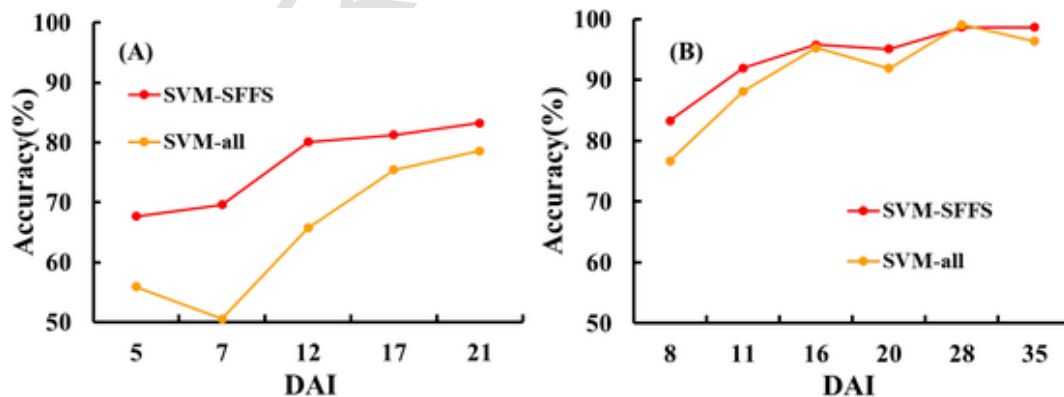

Fig. 8. Comparison of SVM classification accuracies between with (SVM-SFFS) and without (SVM-all) the SFFS procedure for the datasets of 2018 (A) and 2019 (B).

Table 3
Assessment of the ML-SFFS techniques in OA for classifying healthy and infected samples by 4-fold cross-validation with 100 replications.

Healthy vs infected	Infection stage	DAI	LDA		k-NN (k = 3)		SVM	
			Optimal feature combination	Mean accuracy	Optimal feature combination	Mean accuracy	Optimal feature combination	Mean accuracy
2018	Asymptomatic	5	F2, F7	60.58%	F2, F4, F9	58.83%	F7, F8, F9	**67.67%**
		7	F6, F9, F17	60.33%	F7, F9	61.67%	F3, F6, F9	**69.58%**
	Early	12	F4, F6, F8	75.75%	F6, F11, F19	71.17%	F6, F8, F12	**80.08%**
		17	F1, F8, F12	80.25%	F6, F7	68.25%	F1, F2, F8, F12	**81.25%**
		21	F2, F4, F5	81.33%	F2, F4, F21	70.33%	F2, F4, F5	**83.25%**
2019	Early	8	F5, F7, F18	81.91%	F1, F8, F12	72.09%	F5, F6, F17, F18	**83.32%**
		12	F2, F3, F6, F17	89.05%	F2, F3, F6, F17	87.45%	F2, F6, F17	**92.00%**
		16	F3, F21	90.73%	F2, F3, F21	87.50%	F2, F6, F9, F21	**95.77%**
	Mild	20	F4, F5	93.50%	F5, F7, F18	92.73%	F5, F17	**95.08%**
		27	F4, F5, F10	93.05%	F3, F4, F10	95.32%	F3, F4, F5, F9	**98.65%**
		35	F10, F18	96.55%	F3, F6, F10	94.82%	F3, F6, F10	**98.65%**

Note: The best performance for each DAI is highlighted in bold.

of water occurred only at later infection stage during rust pathogenesis on barley, which is highly consistent with our results on water variation induced by RLB infection. However, a previous study has demonstrated that the significant deficit in water content could be observed during the fungus infection caused by the mountain pine beetle (MBP) at the green attack stage (Cheng et al., 2010). This contradictory finding may be due to the biological processes in the pathogenesis that varied considerably under different biotic stresses. For example, the water-conducting xylems were blocked through the MBP infection, which seriously affects the water transport system of the host trees (Solheim, 1995; Yamaoka et al., 1990). For the *Magnaporthe oryzae* infection in this research, the interactions between the pathogen and host may inhibit various biochemical processes in cells (e.g. pigment synthesis) (Ulferts et al., 2015). This suggests that the leaf water content may not be significantly affected in a short period due to the powerful water delivery system of rice .The underlying physiological mechanism of specific disease stress is crucial for understanding how pathogens infect the leaves and the reason for the determination of sensitive spectral features. Relevant studies have examined the biochemical variations during the disease infection process to guide the determination of disease-specific features (Cheng et al., 2010; Zarco-Tejada et al., 2018). Therefore, the determination and interpretation of sensitive spectral features during plant pathogenesis could be linked to physiological analysis. The lagged response of anthocyanin and water content, which are key probes for detecting multiple disease stresses (Cheng et al., 2010; Morel et al., 2018), suggests that the traditional approaches based on these indicators may be insufficient for the early detection of RLB infection. On the contrary, particular attention could be paid to the utilization of the chlorophyll and carotenoid-sensitive indicators. To the best of our knowledge, this study represents the first leaf-level report on revealing the physiological and spectral responses of rice plants to the disease with intensive samplings over the time course of RLB infection.

4.2. Importance of the near-infrared domain for disease detection

Several significant and consistent wavelet features detected in the NIR (750–1000 nm) region indicates that the internal structure of leaves may change during plant-pathogen interactions. Micrographs of RLB mycelia interacting with rice cells at early (Fig. 9A) and mild (Fig. 9B) infection stages could be used to illustrate the damage of RLB pathogen to leaf cellular structure. This could provide a straightforward understanding of the effects of the pathogen on the leaf internal structure. When rice tissues are subjected to fungus infestation, the leaf epidermal cells are commonly destroyed to facilitate the invasion of mycelium into the leaf (Fig. 9A). Then the movement of the fungus from cell-to-cell occurred by means of penetrating the barriers among cells (Fig. 9B) (Wilson and Talbot, 2009). The collapse of physical structure among cells during this process may be the main factor leading to spectral variation in the NIR region, because the leaf optical properties in this region are primarily governed by the intercellular leaf structure (Jacquemoud and Féret, 1990; Ollinger, 2011). According to the research by Morel et al. (2018), some heterogeneity will be introduced to the internal structure of the leaves when the inter-cellular space is penetrated by fungus. However, their result was obtained by simulating the leaf internal structure parameter (N) through the PROCOSIN model and lacked convincing support from experimental measurements. The microscopic analysis of cell-pathogen interaction in this study provides solid evidence to support the fact that the fungus infestation can result in the collapse of the intercellular structure. Readers are referred to Li et al. (2020c) for more
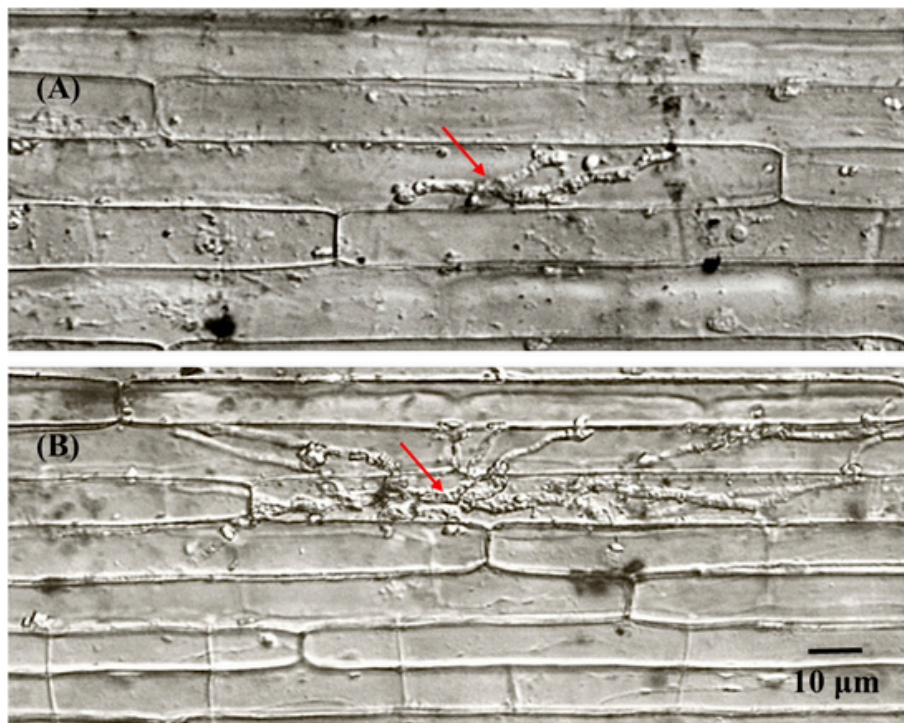
Fig. 9. Micrograph of RB mycelia (denoted by red arrows) interacting with rice cells at the early (A) and mild (B) infection stages, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

experiment details about conidia invasive growth on rice tissue.

Four consistent wavelet features with considerable separability among infected and healthy samples were near 850 nm ($WF_{850,4}$, $WF_{866,5}$) and 950 nm ($WF_{936,5}$, $WF_{960,6}$), which characterized the subtle variation of local spectral slope in the NIR region. The overall reflectance slope in the NIR region (760–900 nm) of the infected leaves increased relative to that of the healthy leaves, which is more obvious in the dataset of 2019 (Fig. 10). This finding is in agreement with our previous research on the identification of RLB lesions based on close-range hyperspectral images (Tian et al., 2018). Although the slope variation in the NIR region (760–900 nm) is almost imperceptible, the subtle variation in this range induced by RLB infection is consistently captured by the wavelet features $WF_{850,4}$ and $WF_{866,5}$. The other two wavelet features $WF_{936,5}$ and $WF_{960,6}$ carried information on local spectral variation over a NIR region (900–1000 nm) that is mainly

dominated by both the water absorption centered at 970 nm and the leaf internal structure (Cheng et al., 2014; Jacquemoud and Féret, 1990; Penuelas et al., 1997). Therefore, the spectral variation over this NIR region was enhanced by the mixed impacts of the destruction of the mesophyll cell structure and the deficit of water content due to RLB infection. A similar wavelet feature ($WF_{950,4}$) was also derived by Cheng et al. (2010) to distinguish control and infested trees at the green attack stage following a significant water deficit. This study represents the first report that emphasizes the potential of the subtle spectral variation in the NIR region for revealing the complex plant internal structure alterations induced by RLB infection.

Among the three spectral analytical approaches, only wavelet analysis captured the subtle variation of spectral shape in the NIR region. Several disease specific spectral indices were developed by Mahlein et al. (2013) and Huang et al. (2014) for the detection of crop diseases using feature selec-
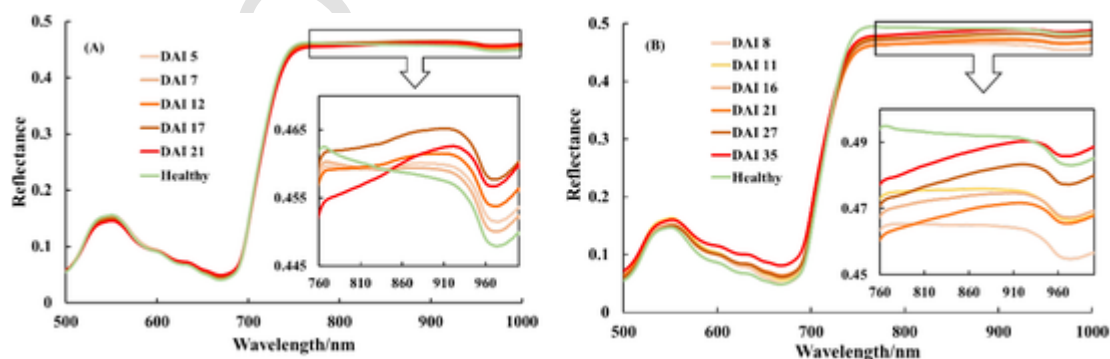

Fig. 10. Comparison of mean reflectance in the NIR region (760–1000 nm) between infected and healthy samples from datasets of (A) 2018 and (B) 2019.

tion algorithms. However, those studies only identified the sensitive bands in the region of 400–750 nm and their methods are still insufficient to distinguish infected samples at early stages. Although the merits of CWA in quantifying the disease severity of plant diseases have been investigated in a few studies (Cheng et al., 2010; Shi et al., 2018a; Zhang et al., 2014), the variation of local spectral shape in the NIR region remains insufficiently mined in the existing literature due to the complexity of its spectral response. The NIR region plays an important role in detecting various plant pests and diseases (Zhang et al., 2019). However, it was primarily used to construct SIs or as input features to machine learning methods in the form of original reflectance values (De Castro et al., 2015; Liu et al., 2019a). Our results demonstrated that the unique subtle disease information in the NIR region could be enhanced and fully exploited by performing the wavelet decomposition, which greatly facilitates the detection of RLB induced subtle signals.

4.3. Advantages of using ML-SFFS classification models

The proposed ML-SFFS classification methodology goes beyond previous disease classification models by optimizing the combination of sensitive features for improved performance. The pronounced increase in OA with DAI suggests that an OFC consisting of 2 to 4 features could provide a higher OA than all DSSFs even with a lower computational cost (Fig. 8). Although individual spectral features (De Castro et al., 2015; Huang et al., 2014; Mahlein et al., 2013; Penuelas et al., 1995; Shi et al., 2018b) and ML (Badnakhe et al., 2018; Feng et al., 2020; Rumpf et al., 2010)have been used in previous studies for the detection of plant disease with promising results, the unambiguous separation of infected samples is challenging at the early infection stage due to comparatively low spectral changes in reflectance spectra (Cheng et al., 2010; Mahlein et al., 2013). On the other hand, most previous studies employed complex classification strategies for disease detection and had only a few attempts to optimize the OFC for improved classification performance (Al-Saddik et al., 2018; Hornero et al., 2020; Liu et al., 2019a; Rumpf et al., 2010; Shi et al., 2018a; Zarco-Tejada et al., 2018). By optimizing the combination of sensitive features, the weak spectral signal to disease stress at the early infection stage could be effectively enhanced. Additionally, the ML-SFFS methodology has considerable advantages regarding the computational efficiency over the physical-based RTMs. Feature extraction and selection from hyperspectral data can greatly improve computational efficiency and highlight the important features for the construction of classification methodology. Several studies have emphasized the limitation of computational time required for model inversion, which restricts the large-scale applicability of the RTMs in disease detection(Hornero et al., 2020; Rivera et al., 2015). On the contrary, feature selection algorithms have proved to be an effective strategy to retaining useful information while decreasing the computational time (Huang et al., 2019; Kaushal and Swarnajyoti, 2018). The application of the VIF analysis and ML-SFFS algorithm in this study allowed not only the reduction of collinearity among predictor variables, but also the relief of computational burden.

Another strength of the ML-based classification methodology is to ensure the rationality and interpretability of the OFCs selected from DSSFs by the SFFS procedure. For example, the features selected from the reflectance data may perform well

but could be case-specific and lack interpretability. However, all the OFCs selected from DSSFs were directly linked to the RLB infection and the corresponding physiological and morphological variations occurring to infected leaves, which permits the transferability and generalization of this methodology to other conditions (Zarco-Tejada et al., 2018). Generally, multiple features related to several plant traits were involved in an OFC, which better reveals the complex physiological processes associated with RLB infection. Note that the OFCs are inconsistent over different DAIs and we cannot claim that a specific feature combination is more significant in the differentiation of RLB infection. This inconsistency could be attributed to the variation in spectral features responding to the physiological progression of RLB infection. Recently, the relative importance of sensitive spectral features for disease detection has received attention in the literature (Poblete et al., 2020). By performing the DSSF selection and SFFS operation, the most informative features were emphasized and the remainder were suppressed. Thus, a large amount of less informative and redundant spectral features were filtered out after these operations, which effectively balanced computational efficiency and spectral separability. Similar strategies have been applied to reveal the most effective plant traits (Zarco-Tejada et al., 2018) and to evaluate the contribution of different spectral indicators (Poblete et al., 2020) in the detection of pre-visual symptoms of *Xylella fastidiosa* infection.

4.4. Potential applications and limitations

The chlorophyll content (C_{a+b}) and several DSSFs (Table 2) that were characterized as the most effective and consistent proxies to reveal RLB infection provide considerable transferability and potential application for early detection of RLB infection to other datasets, rice species and environmental conditions. One particular interest is the chlorophyll content, which exhibited a considerable response to the occurrence of RLB infection at the asymptomatic stage. Accurate estimation of leaf chlorophyll content at various scales has been proved to be feasible in recent studies (Gitelson and Solovchenko, 2017; Gitelson et al., 2019; Li et al., 2020a; Xu et al., 2019), which facilitates the possibility to detect the subtle chlorophyll variation induced by RLB infection. Moreover, the systematic influence of pathogen infestation on the topmost layer could be amplified at the canopy level due to the reabsorption between multiple leaf layers, which will provide support for up-scaling application of the detection of RLB infection. Similarly, several spectral features derived from the VNIR region also exhibited promising applications on RLB detection for their potential link with physiological effects induced by RLB infection, such as R_{673}, CCI, PRI_{m2}, PSSRa and PSNDa (Blackburn, 1998; Gamon et al., 2016; Gamon et al., 1992). Additionally, other sensitive spectral features (e.g, SRPI and HI_2013) have already demonstrated the superiority on the evaluation and classification of various plant disease (Mahlein et al., 2013; Penuelas et al., 1995).

Although the ML-SFFS methodology performs well in the classification of RLB infected samples over various infection stages, the limitations with the ambiguity of the sensitivity of sensitive features prevent their potential application in the early detection of RLB using multispectral sensors on different platform. Resent researches showed that the relative importance of each input indicator may vary with disease infection stage (Poblete et al., 2020; Zarco-Tejada et al., 2018). Given that the sensitive features were all selected from hyper-

spectral data, the demonstrated relative importance of sensitive features may serve as a practical guide for the RLB detection when limited spectral information is available. On the other hand, it could also provide new insight into the complex plant-pathogen interactions to assess the contribution of each sensitive indicators in the detection of RLB at various infection stages.

5. Conclusions

This study demonstrated the combination of a few DSSFs enabled the detection of RLB infection from the asymptomatic to mild infection stages and the hyperspectral detection was well supported by physiological variations. Non-destructive detection of RLB infection at an asymptomatic and early stage would be of great significance to prevent a devastating outbreak. Our work made full use of the abundant spectral information in the two-year datasets from three perspectives, including individual reflectance bands, SIs, and wavelet coefficients. A total of 21 DSSFs were consistently determined to be the most significant features for revealing the occurrence of RLB infection. In addition, significant biochemical variations occurred in both years during plant-pathogen interactions and were physically linked to the selected DSSFs. While feeding the selected DSSFs to the ML-SFFS classification methodology, the OFC for each DAI was determined and the OA was greatly improved. The encouraging classification accuracy of over 80% for the early infection stage suggested the feasibility of consistent hyperspectral signals in the non-destructive detection of RLB infection. The following conclusions could be drawn from the results:

- We have illuminated the responses of leaf spectral and physiological properties to RLB infection and determined the sensitivity of different biochemical parameters across the RLB infection process. Specifically, significant deficits in total chlorophyll, carotenoid, and water contents were observed for infected leaves and a contradictory trend in anthocyanin content. Compared with anthocyanin and water contents, chlorophyll and carotenoid contents had a superior sensitivity to the occurrence of RLB infection. Moreover, this pattern of variation existed consistently during plant-pathogen interactions.
- Several DSSFs were defined to distinguish infected leaves from healthy ones. Two reflectance bands, 14 VIs, and five wavelet feature proved to be the most significant sensitive indicators to the differentiation of infected leaves. More importantly, four wavelet features located in the NIR region were first reported and could be interpreted by intercellular structure collapse. In general, the selected DSSFs were dominantly concentrated in the VNIR domain.
- The overall classification performance of the ML-SFFS methodology was significantly improved by inputting the selected DSSFs instead of all features. It was shown that the combination of two to four DSSFs was sufficient to identify infected samples with OAs of over 66%, 80% and 95% at the asymptomatic, early and mild infection stages, respectively. Compared to the OAs obtained without feature selection, the OA of SVM-SFFS was 10% higher for most DAIs of 2018 and marginally different for 2019.

Future work could be focused on assessing the contribution of each DSSF to the RLB infection detection, or developing more simplified strategies (e.g., specific spectral index) for the detection over different infection stages. When airborne or spaceborne hyperspectral data are available at sufficient spatial and temporal resolutions the disease detection methodology has a considerable potential for the mapping of early occurrence and severity levels of RLB infection or other cereal crop diseases at large scales.

CRediT authorship contribution statement

Long Tian: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Bowen Xue:** Resources. **Ziyi Wang:** Resources. **Dong Li:** Methodology, Formal analysis. **Xia Yao:** Resources, Supervision. **Qiang Cao:** Resources, Funding acquisition. **Yan Zhu:** Conceptualization, Funding acquisition. **Weixing Cao:** Conceptualization, Funding acquisition. **Tao Cheng:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

None

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112350.

References

Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS J. Photogramm. Remote Sens. 88, 48–59.

Al-Saddik, H., Laybros, A., Billiot, B., Cointault, F., 2018. Using image texture and spectral reflectance analysis to detect yellowness and esca in grapevines at leaf-level. Remote Sens. 10, 618.

Badnakhe, M.R., Durbha, S.S., Jagarlapudi, A., Gade, R.M., 2018. Evaluation of Citrus gummosis disease dynamics and predictions with weather and inversion based leaf optical model. Comput. Electron. Agric. 155, 130–141.

Behmann, J., Mahlein, A.-K., Rumpf, T., Roemer, C., Pluemer, L., 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis. Agric. 16, 239–260.

Blackburn, G.A., 1998. Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches. Remote Sens. Environ. 66, 273–285.

Bock, C.H., Barbedo, J.G.A., Del Ponte, E.M., Bohnenkamp, D., Mahlein, A.-K., 2020. From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. Phytopathol. Res. 2.

Calderón, R., Navas-Cortés, J.A., Lucena, C., Zarco-Tejada, P.J., 2013. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sens. Environ. 139, 231–245.

Chang, C.-C., Lin, C.-J., 2011. Libsvm. ACM Trans. Intell. Syst. Technol. 2, 1–27.

Chemura, A., Mutanga, O., Dube, T., 2016. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis. Agric. 18, 859–881.

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.

Cheng, T., Riaño, D., Ustin, S.L., 2014. Detecting diurnal and seasonal variation in canopy water content of nut tree orchards from airborne imaging spectroscopy data using continuous wavelet analysis. Remote Sens. Environ. 143, 39–53.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.

De Castro, A.I., Ehsani, R., Ploetz, R., Crane, J.H., Abdulridha, J., 2015. Optimum spectral and geometric parameters for early detection of laurel wilt disease in avocado. Remote Sens. Environ. 171, 33–44.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H.Q., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W.X., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E., Birren, B.W., 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature 434, 980–986.

Deng, Y.W., Zhai, K.R., Xie, Z., Yang, D.Y., Zhu, X.D., Liu, J.Z., Wang, X., Qin, P., Yang, Y.Z., Zhang, G.M., Li, Q., Zhang, J.F., Wu, S.Q., Milazzo, J., Mao, B.Z., Wang, E.T., Xie, H., Tharreau, D., He, Z.H., 2017. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. Science 355, 962–965.

Fallahpour, S., Lakvan, E.N., Zadeh, M.H., 2017. Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem. J. Retail. Consum. Serv. 34, 159–167.

Fang, N., Wei, X., Shen, L., Yu, Y., Li, M., Yin, C., He, W., Guan, C., Chen, H., Zhang, H., Bao, Y., 2019. Fine mapping of a panicle blast resistance gene Pb-bd1 in Japonica landrace Bodao and its application in rice breeding. Rice (N Y) 12, 18.

Feng, L., Chai, R.Y., Sun, G.M., Wu, D., Lou, B.G., He, Y., 2009. Identification and classification of rice leaf blast based on multi-spectral imaging sensor. Spectrosc. Spectr. Anal. 29, 2730–2733.

Feng, X., Zhan, Y., Wang, Q., Yang, X., Yu, C., Wang, H., Tang, Z., Jiang, D., Peng, C., He, Y., 2020. Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. Plant J. 101, 1448–1461.

Feret, J.-B., Francois, C., Gitelson, A., Asner, G.P., Barry, K.M., Panigada, C., Richardson, A.D., Jacquemoud, S., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sens. Environ. 115, 2742–2750.

Féret, J.B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215.

Fisher, R.A., 1936. The use of multiple measures in taxonomic problems. Ann. Eugenics 7, 179–188.

Franceschini, M.H.D., Bartholomeus, H., van Apeldoorn, D.F., Suomalainen, J., Kooistra, L., 2019. Feasibility of unmanned aerial vehicle optical imagery for early detection and severity assessment of late blight in potato. Remote Sens. 11, 224.

Gamon, J.A., Penuelas, J., Field, C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41, 35–44.

Gamon, J.A., Huemmrich, K.F., Wong, C.Y.S., Ensminger, I., Garrity, S., Hollinger, D.Y., Noormets, A., Peñuelas, J., 2016. A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. Proc. Natl. Acad. Sci. 113, 13087–13092.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. Proc. Natl. Acad. Sci. U. S. A. 115, 4613–4618.

Gitelson, A., Solovchenko, A., 2017. Generic algorithms for estimating foliar pigment content. Geophys. Res. Lett. 44, 9293–9298.

Gitelson, A., Viña, A., Solovchenko, A., Arkebauer, T., Inoue, Y., 2019. Derivation of canopy light absorption coefficient from reflectance spectra. Remote Sens. Environ. 231, 111276.

Hamed, A.M.S., Mehdi, M., Asghari, B.B., 2020. A feature extraction method based on spectral segmentation and integration of hyperspectral images. Int. J. Appl. Earth Obs. Geoinf. 89, 102097.

Hardisky, M.A., Klemas, V., Smart, R.M., 1983. The influence of soil-salinity, growth form, and leaf moisture on the spectral radiance of spartina–alterniflora canopies. Photogramm. Eng. Remote. Sens. 49, 77–83.

Hernández-Clemente, R., North, P.R.J., Hornero, A., Zarco-Tejada, P.J., 2017. Assessing the effects of forest health on sun-induced chlorophyll fluorescence using the FluorFLIGHT 3-D radiative transfer model to account for forest structure. Remote Sens. Environ. 193, 165–179.

Hornero, A., Hernández-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.

Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P., Wu, W., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. Int. J. Appl. Earth Obs. Geoinf. 80, 218–229.

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, W., Guan, Q., Luo, J., Zhang, J., 2014. New optimized spectral indices for identifying and monitoring winter wheat diseases. IEEE J.Select. Top. Appl. Earth Observ. Remote Sens. 7, 2516–2524.

Huang, Y., Li, Z., Risinger, A.L., Enslowd, B.T., IV, Gonge, J., Yang, Y., Schanze, K.S., 2019. Fluorescence Spectral Shape Analysis for Nucleotide Identification.

Jacquemoud, S., Féret, J.B., 1990. PROSPECT: a model of leaf optical properties spectra. Remote Sens. Environ. 75, 9–91.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT + SAIL models: a review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66.

Jay, S., Bendoula, R., Hadoux, X., Féret, J.-B., Gorretta, N., 2016. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. Remote Sens. Environ. 177, 220–236.

Jr, F.W.N., Teng, P.S., Shokes, F.M., 1991. Disease assessment terms and concepts. Plant Dis. 75, 1187–1188.

Kaushal, B., Swarnajyoti, P., 2018. An unsupervised technique for optimal feature selection in attribute profiles for spectral-spatial classification of hyperspectral images. ISPRS J. Photogramm. Remote Sens. 138, 139–150.

Kempeneers, P., De Backer, S.B., Debruyn, W., Scheunders, P., 2004. Wavelet based feature extraction for hyperspectral vegetation monitoring. In: Bruzzone, L. (Ed.), Image and Signal Processing for Remote Sensing Ix. pp. 297–305.

Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., Torigoe, Y., 2001. Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathology 91, 316–323.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hun, T., Team., A.R.C, 2020. Misc Functions for Training and Plotting Classification and Regression Models. v6. pp. 0–86.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766.

Li, D., Cheng, T., Jia, M., Zhou, K., Lu, N., Yao, X., Tian, Y., Zhu, Y., Cao, W., 2018. PROCWT: coupling PROSPECT with continuous wavelet transform to improve the retrieval of foliar chemistry from leaf bidirectional reflectance spectra. Remote Sens. Environ. 206, 1–14.

Li, D., Tian, L., Wan, Z., Jia, M., Yao, X., Tian, Y., Zhu, Y., Cao, W., Cheng, T., 2019. Assessment of unified models for estimating leaf chlorophyll content across directional-hemispherical reflectance and bidirectional reflectance spectra. Remote Sens. Environ. 111240.

Li, D., Chen, J.M., Zhang, X., Yan, Y., Zhu, J., Zheng, H., Zhou, K., Yao, X., Tian, Y., Zhu, Y., Cheng, T., Cao, W., 2020a. Improved estimation of leaf chlorophyll content of row crops from canopy reflectance spectra through minimizing canopy structural effects and optimizing off-noon observation time. Remote Sens. Environ. 248, 111985.

Li, X., Huang, H., Shabanov, N.V., Chen, L., Yan, K., Shi, J., 2020b. Extending the stochastic radiative transfer theory to simulate BRF over forests with heterogeneous distribution of damaged foliage inside of tree crowns. Remote Sens. Environ. 250, 112040.

Li, Y., Liu, X., Liu, M., Wang, Y., Zou, Y., You, Y., Yang, L., Hu, J., Zhang, H., Zheng, X., Wang, P., Zhang, Z., 2020c. Magnaporthe oryzae auxiliary activity protein MoAa91 functions as chitin-binding protein to induce Appressorium formation on artificial inductive surfaces and suppress plant immunity. mBio 11.

Lin, Q.A., Huang, H.G., Yu, L.F., Wang, J.X., 2018. Detection of shoot beetle stress on Yunnan pine forest using a coupled LIBERTY2-INFORM simulation. Remote Sens. 10, 19.

Lin, Q.N., Huang, H.G., Wang, J.X., Huang, K., Liu, Y.Y., 2019. Detection of pine shoot beetle (PSB) stress on pine forests at individual tree level using UAV-based hyperspectral imagery and lidar. Remote Sens. 11, 20.

Liu, Z.Y., Wu, H.F., Huang, J.F., 2010. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. Comput. Electron. Agric. 72, 99–106.

Liu, L., Dong, Y., Huang, W., Du, X., Luo, J., Shi, Y., Ma, H., 2019a. Enhanced regional monitoring of wheat powdery mildew based on an instance-based transfer learning method. Remote Sens. 11, 298.

Liu, M., Zhang, S., Hu, J., Sun, W., Padilla, J., He, Y., Li, Y., Yin, Z., Liu, X., Wang, W., Shen, D., Li, D., Zhang, H., Zheng, X., Cui, Z., Wang, G.-L., Wang, P., Zhou, B., Zhang, Z., 2019b. Phosphorylation-guarded light-harvesting complex II contributes to broad-spectrum blast resistance in rice. Proc. Natl. Acad. Sci. 116, 17572–17577.

Luo, J., Huang, W., Zhao, J., Zhang, J., 2013. Detecting aphid density of winter wheat leaf using hyperspectral measurements. IEEE J.Select. Top. Appl. Earth Observ. Remote Sens. 6, 690–698.

Mahlein, A.K., Steiner, U., Dehne, H.W., Oerke, E.C., 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. Precis. Agric. 11, 413–431.

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 Sens. Environ. 128, 21–30.

Mahlein, A.K., Kuska, M.T., Behmann, J., Polder, G., Walter, A., 2018. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. Annu. Rev. Phytopathol. 56, 535–558.

Mahlein, A.K., Kuska, M.T., Thomas, S., Wahabzada, M., Behmann, J., Rascher, U., Kersting, K., 2019. Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: seamless interlocking of phytopathology, sensors, and machine learning is needed!. Curr. Opin. Plant Biol. 50, 156–162.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., Lin, C., 2020. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. v1.7–4 (CRAN).

Morel, J., Jay, S., Feret, J.B., Bakache, A., Bendoula, R., Carreel, F., Gorretta, N., 2018. Exploring the potential of PROCOSINE and close-range hyperspectral imaging to study the effects of fungal diseases on leaf physiology. Sci. Rep. 8, 15933.

Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. New Phytol. 189, 375–394.

Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Trans. Geosci. Remote Sens. 48, 2297–2307.

Penuelas, J., Filella, I., Lloret, P., Mun⁻Oz, F., & Vilajeliu, M., 1995. Reflectance assessment of mite effects on apple trees. Int. J. Remote Sens. 16, 2727–2733.

Penuelas, J., Pinol, J., Ogaya, R., Filella, I., 1997. Estimation of plant water concentration by the reflectance water index WI (R900/R970). Int. J. Remote Sens. 18, 2869–2875.

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.

Pudil, P., Novovicova, J., Kittler, J., 1994. Floating Search Methods in Feature Selection. *Pattern Recognit. Lett,* 15, 1119-1125.

Rivera, J., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G., 2015. An emulator toolbox to approximate radiative transfer models with statistical learning. Remote Sens. 7, 9347–9370.

Rumpf, T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., Pluemer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput. Electron. Agric. 74, 91–99.

Schliep, K., Hechenbichler, K., Lizee, A., 2016. kknn: Weighted k-Nearest Neighbors. V1.3.1.

Sharma, T.R., Rai, A.K., Gupta, S.K., Vijayan, J., Devanna, B.N., Ray, S., 2012. Rice blast management through host-plant resistance: retrospect and prospects. Agricult. Res. 1, 37–52.

Shi, Y., Huang, W., González-Moreno, P., Luke, B., Dong, Y., Zheng, Q., Ma, H., Liu, L., 2018a. Wavelet-based rust spectral feature set (WRSFs): a novel spectral feature set based on continuous wavelet transformation for tracking progressive host–pathogen interaction of yellow rust on wheat. Remote Sens. 10, 525.

Shi, Y., Huang, W., Ye, H., Ruan, C., Xing, N., Geng, Y., Dong, Y., Peng, D., 2018b. Partial least square discriminant analysis based on normalized two-stage vegetation indices for mapping damage from rice diseases using planetscope datasets. Sensors (Basel) 18.

Solheim, H., 1995. Early stages of blue-stain fungus invasion of lodgepole pine sapwood following mountain pine-beetle attack. Can. J. Botany-Revue Canadienne De Botanique 73, 70–74.

Talbot, N.J., 2003. On the trail of a cereal killer: exploring the biology of Magnaporthe grisea. Annu. Rev. Microbiol. 57, 177–202.

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., Wan, Z.F., Li, D., Jiang, J.L., Yao, X., Cao, Q., Tian, Y.C., Zhu, Y., Cao, W.X., Cheng, T., Ieee, 2018. Detecting rice blast disease using model inverted biochemical variables from close-range reflectance imagery of fresh leaves. In: Igarss 2018–2018 Ieee International Geoscience and Remote Sensing Symposium. Ieee, New York, pp. 2749–2756.

Ulferts, S., Delventhal, R., Splivallo, R., Karlovsky, P., Schaffrath, U., 2015. Abscisic acid negatively interferes with basal defence of barley against Magnaporthe oryzae. BMC Plant Biol. 15, 7.

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., Zarco-Tejada, P., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ. 113, S67–S77.

van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 3109–3129.

Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., Kersting, K., 2015. Metro maps of plant disease dynamics-automated mining of differences using hyperspectral images. PLoS One 10.

Wei, C., Huang, J., Wang, X., Blackburn, G.A., Zhang, Y., Wang, S., Mansaray, L.R., 2017. Hyperspectral characterization of freezing injury and its biochemical impacts in oilseed rape leaves. Remote Sens. Environ. 195, 56–66.

Weinberger, K., Blitzer, J., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor Classificatio. J. Mach. Learn. Res. 10, 207–244.

Wilson, R.A., Talbot, N.J., 2009. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. Nat. Rev. Microbiol. 7, 185–195.

Wolf, P.F.J., Verreet, J.A., 2002. An integrated pest management system in Germany for the control of fungal leaf diseases in sugar beet: the IPM sugar beet model. Plant Dis. 86, 336–344.

Wu, D., Cao, F., Zhang, H., Sun, G.M., Feng, L., He, Y., 2009. Study on disease level classification of Rice panicle blast based on visible and near infrared spectroscopy. Spectrosc. Spectr. Anal. 29, 3295–3299.

Xu, X.Q., Lu, J.S., Zhang, N., Yang, T.C., He, J.Y., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C., 2019. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. ISPRS J. Photogramm. Remote Sens. 150, 185–196.

Yamaoka, Y., Swanson, R.H., Hiratsuka, Y., 1990. Inoculation of lodgepole pine with 4 blue-stain fungi associated with mountain pine-beetle, monitored by a heat pulse velocity (HPV) instrument. Can. J. Forest Research-Revue Can. Rech. Forest. 20, 31–36.

Yang, W., Zhang, H., Li, M., Wang, Z., Zhou, J., Wang, S., Lu, G., Fu, F., 2014. Early diagnosis of blast fungus, Magnaporthe oryzae, in rice plant by using an ultra-sensitive electrically magnetic-controllable electrochemical biosensor. Anal. Chim. Acta 850, 85–91.

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. Previsual symptoms of Xylella

fastidiosa infection revealed in spectral plant-trait alterations. Nat. Plants 4, 432–439.

Zhang, J., Pu, R., Huang, W., Lin, Y., Luo, J., Wang, J., 2012a. Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses. Field Crop Res. 134, 165–174.

Zhang, J.C., Pu, R., Wang, J., Huang, W., 2012b. Detecting powdery mildew of winter wheat using leaf level hyperspectral measurements. Comput. Electron. Agric. 85, 13–23.

Zhang, J., Yuan, L., Pu, R., Loraamm, R.W., Yang, G., Wang, J., 2014. Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat. Comput. Electron. Agric. 100, 79–87.

Zhang, J.C., Huang, Y.B., Pu, R.L., Gonzalez-Moreno, P., Yuan, L., Wu, K.H., Huang, W.J., 2019. Monitoring plant diseases and pests through remote sensing technology: a review. Comput. Electron. Agric. 165, 14.

Zhou, L.N., Yu, H.Y., Zhang, L., Ren, S., Sui, Y.Y., Yu, L.J., 2014. Rice blast prediction model based on analysis of chlorophyll fluorescence Spectrum. Spectrosc. Spectr. Anal. 34, 1003–1006.

Zhou, R.Q., Jin, J.J., Li, Q.M., Su, Z.Z., Yu, X.J., Tang, Y., Luo, S.M., He, Y., Li, X.L., 2018. Early detection of Magnaporthe oryzae-infected barley leaves and lesion visualization based on hyperspectral imaging. Front. Plant Sci. 9, 1962.