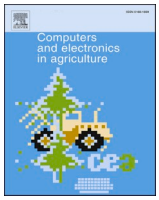
Computers and Electronics in Agriculture 215 (2023) 108366



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag





Monitoring of winter wheat stripe rust by collaborating canopy SIF with wavelet energy coefficients

Kehui Ren [a,b,c], Yingying Dong [a,b,c,*], Wenjiang Huang [a,b,c], Anting Guo [a,b], Xia Jing [d]

[a] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[b] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China
[c] University of Chinese Academy of Sciences, Beijing 100049, China
[d] College of Geometrics, Xi'an University of Science and Technology, Xi'an 710054, China

ARTICLE INFO

Keywords:
Remote sensing
Stripe rust
Solar-induced chlorophyll fluorescence
Continuous wavelet transform
Fractional-order derivative

ABSTRACT

Stripe rust has caused tremendous damage to wheat quality and production. Disease-specific factors revealing group structure and photosynthetic physiology contribute to high-precision monitoring for wheat stripe rust. In this study, we proposed a remote sensing model that collaborates wavelet features (WFs) and solar-induced chlorophyll fluorescence (SIF). First, sensitive features including vegetation indices (VIs), SIF parameters, WFs, and fractional-order derivative spectra (FODs) were screened based on correlation coefficient (CC) analysis and variable importance in projection (VIP). Then, through collaboration among features, six feature sets were received and imported to partial least squares regression (PLSR), back-propagation neural network (BPNN), random forest (RF), and extreme gradient boosting (XGBoost). Finally, monitoring models was evaluated through two methods: holdout cross-validation and 5-fold cross validation to ascertain the optimal feature-algorithm combination. The results demonstrated that the collaboration of canopy SIF with any feature markedly improved the monitoring accuracy due to its responsive nature to the plant's photosynthetic physiology. The model based on XGBoost with WFs-SIF as input features achieved optimal monitoring accuracy, with at least 16.6% increase in R^2 and 32.4% reduction in RMSE compared to the VIs-PLSR model. Correlation analysis of evaluation indexes (R^2 and RMSE) under two cross-validation methods showed determination coefficients of 0.743 and 0.837, indicating mutual validation and high reliability of the conclusions. This study suggests that the collaboration between WFs and SIF exhibits considerable feasibility in high-precision monitoring of stripe rust, providing a novel insight for future field-scale diagnosis of crop diseases.

1. Introduction

Wheat constitutes a pivotal grain crop for China, and its yield holds an intimate correlation to the nation's food security strategy (Han et al., 2020). Stripe rust serves as an airborne fungal disease that is highly prevalent and widespread, triggering symptoms such as leaf yellowing, wilting, and stunting at diverse stages of infection (Huang et al., 2019a). During epidemic years, yield loss in wheat production caused by stripe rust even exceeds 40 %. Hence, obtaining accurate incidence of stripe rust becomes particularly essential for disease control, yield improvement, and minimizing economic losses. Through decades of development, remote sensing has been proven to be efficient in the detection and monitoring of crop diseases, especially hyperspectral data, which can be used to reveal stress status benefiting from rich narrow-band information (Abdulridha et al., 2019).

After being infected by pathogens, crops undergo variations in physiological and biochemical characteristics, as well as surface morphology, resulting in alterations in visible-near infrared reflectance on their leaves (Zhang et al., 2019). Plentiful studies have demonstrated that growth conditions, infection categories and disease severity could be detectable by constructing vegetation indices based on sensitive bands and abnormal changes of spectral response (Ren et al., 2021). For instance, Jiang et al. (2022b) developed a new three-band hyperspectral vegetation index for estimating leaf relative chlorophyll content of mangrove under different severities of pest and disease. Tian et al. (2023) introduced a new RB-specific three-band VI (RIBIs) that not only

* Corresponding author at: Dr. Yingying Dong, Aerospace Information Research Institute, Chinese Academy of Sciences No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China.
E-mail address: dongyy@aircas.ac.cn (Y. Dong).

https://doi.org/10.1016/j.compag.2023.108366
Received 26 June 2023; Received in revised form 26 September 2023; Accepted 24 October 2023
Available online 11 November 2023
0168-1699/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

effectively distinguishes between healthy and rice blast-infected leaves but also allows for precise quantification of disease incidence at the canopy level. Nonetheless, the reflectance spectrum signal is limited in its capacity to disclose the photosynthetic physiological status of plants and is susceptible to noise from soil and shade, thereby curtailing its usability in detecting diseases (Ashourloo et al., 2014). In contrast, SIF flux stems from photosystem II (PSII) regarded as the upstream of entire photosynthesis process, intuitively characterizing crop's physiological changes (Sankaran et al., 2010). Jing et al. (2022) proposed a feature selection and disease index (DI) monitoring model, which integrates the max-relevance and min-redundancy (mRMR) and XGBoost algorithm to address feature redundancy when adopting reflectance spectra and SIF data to detect wheat stripe rust. Du et al. (2023) discovered that SIF responds more strongly to disease severity levels (SL) than normalized difference vegetation index (NDVI) at the leaf scale, and their results confirmed the capability of ΦF-r estimated by the NIRvR method to monitor biologically-driven disease stress.

However, merely utilizing features extracted from the original spectra is unsatisfactory because they are generally incompetent in revealing disease stress information (Zhang et al., 2014). Exploring disease-specific factors that are particularly responsive from full-band spectra, which has become an important research direction in recent years. Continuous wavelet transform (CWT) and fractional-order derivative (FOD) are two universally employed mathematical transformation methods in hyperspectral preprocessing that can attenuate noise and refine spectral information (Jiang et al., 2022a; Mustafa et al., 2022). Shi et al. (2018) innovatively presented a wavelet-based spectral feature set to describe the occurrence process of stripe rust in wheat. Zhang et al. (2023) evaluated canopy spectra under the influence of wheat stripe rust using FOD and constructed a model with higher accuracy and stability compared to the original reflectance spectra. Although both of two techniques are proficient in mining the gradual information within the spectra, they are typically applied individually, and the comparative analysis of their effectiveness in feature extraction is rarely seen. Therefore, it is worth further investigating whether collaborating with SIF on the basis of WFs or FODs can ameliorate the accuracy of stripe rust monitoring.

In addition to the feature types and preprocessing methods, the algorithm also has a considerable impact on the accuracy of disease monitoring. Crop disease remote sensing monitoring models can be broadly divided into two categories: statistical models and machine learning models. Common statistical models used include PLSR (Zhang et al., 2014), binary logistic regression (BLR) (Li et al., 2020), multiple stepwise regression (MSR) (Jones et al., 2010), among others. While these models are computationally simple, they are susceptible to observation noise and restrictively operated in the spatial and temporal dimensions. Machine learning algorithms that can balance training error and generalization ability have been gaining popularity, including fisher's linear discriminant analysis (FLDA) (Huang et al., 2019b), BPNN (Wang et al., 2015), support vector machine (SVM) (Ma et al., 2017), and others. These algorithms possess excellent nonlinear fitting capabilities, which are critical for challenge caused by minor variations in reflectance coefficients. Additionally, ensemble learning algorithms, such as RF and XGBoost, appear more robust against noise and outliers (Bolón-Canedo and Alonso-Betanzos, 2019; Elavarasan and Vincent, 2020). XGBoost, in particular, generates subsequent decision trees based on the residuals of the previous tree, which endows it with superior generalization capability compared to simple averaging of independent trees.

Given the above status, this study extracted and analyzed VIs, WFs, and FODs at the canopy scale, further generating feature sets by collaboration with SIF. Four different algorithms including PLSR, BPNN, RF, and XGBoost were applied to monitoring wheat stripe rust. The specific objectives of this study were to evaluate the potential of canopy SIF for disease monitoring and identify characteristic bands from full-band spectra. Our findings will be conducive to ascertain suitable decomposition scales, fractional orders and the optimal feature-algorithm combination.

2. Materials and methods

2.1. Study area and data acquisition

The experiment was carried out in 2018 at the Chinese Academy of Agricultural Sciences experimental station located in Langfang City, Hebei Province, China (39°30′40″N,116°36′20″E). The study area, situated in the middle east of the North China Plain, has a warm temperate continental monsoon climate. The experimental cultivar was "Mingxian 169″, which is more susceptible to stripe rust, and the disease stress was artificially inoculated. The study area was divided into four groups (Fig. 1), with each group covering an area of 220 m^2. Groups A and D were the healthy control groups, and groups B and C were the infected groups, with each group containing eight plots spaced at least 5 m apart. On April 9, 2018, a spore solution with a concentration of 9 mg/100 ml was evenly sprayed on wheat leaves using a hand-held sprayer. Canopy spectra and DI were collected for 51 wheat samples, including 45 diseased samples and 6 healthy samples, on May 18th, May 24th, and May 30th in 2018.

Canopy spectra were collected using an ASD FieldSpec spectroradiometer (Analytical Spectral Devices, Inc., Boulder, CO, USA) and a QE Pro spectrometer (Ocean Optics Inc., Dunedin, FL, USA) in this experiment. The measurements were conducted between 11:00 and 13:00 Beijing time under sunny, windless, and cloudless conditions. The ASD Field Spec®Pro FR spectrometer covers the spectral range of 350–2500 nm, with spectral resolutions of 3 nm and 10 nm respectively in the range of 350–1000 nm and 1000–2500 nm, and the spectra were resampled to 1 nm bandwidth. The QE 65 pro spectrometer has a spectral resolution of 0.3 nm in the range of 644–806 nm, with a sampling interval of 0.15 nm. The probe was placed vertically downward, at a distance of 1.3 m from the wheat canopy. For each sample, measurements were repeated ten times, and corrected with a $BaSO_4$ calibration plate before and after each measurement. Canopy reflectance was calculated according to Formula 1.

$$R = \frac{L_{target}}{L_{board}} \cdot R_{board} \quad (1)$$

Where R is the reflectance of wheat canopy, L_{target} is the radiance of wheat canopy, L_{board} is the radiance of the calibration plate, and R_{board} is the reflectance of the calibration plate.

To survey the canopy DI, a five-point sampling method was employed, where five symmetrical points were selected within each quadrat with an area of 1 m^2. For each point, ten wheat plants were randomly chosen to investigate the leaf disease. The severity of a single leaf was evaluated based on the National Crop Disease Survey and Prediction Rules (GB/T 15795–2011) and divided into nine gradients ranging from 0 to 8, corresponding to incidence rates of 0 %, 1 %, 10 %, 20 %, 30 %, 45 %, 60 %, 80 %, and 100 %. The number of wheat leaves with different severity was recorded, and the DI for each test quadrat was calculated as follow:

$$DI = \frac{\sum_{i=0}^{8} x_i \cdot g_i}{n \cdot \sum_{i=0}^{8} g_i} \quad (2)$$

Where x_i is the value of each gradient, n is the highest gradient namely eight, and g_i is the number of leaves in each gradient.

2.2. Construction of wheat stripe rust monitoring models

Fig. 2 demonstrates the methodological framework of this study. First, VIs and SIF parameters highly responsive to stripe rust were calculated and screened from the original reflectance. Then, the spectra treated with CWT or FOD were further processed by correlation analysis

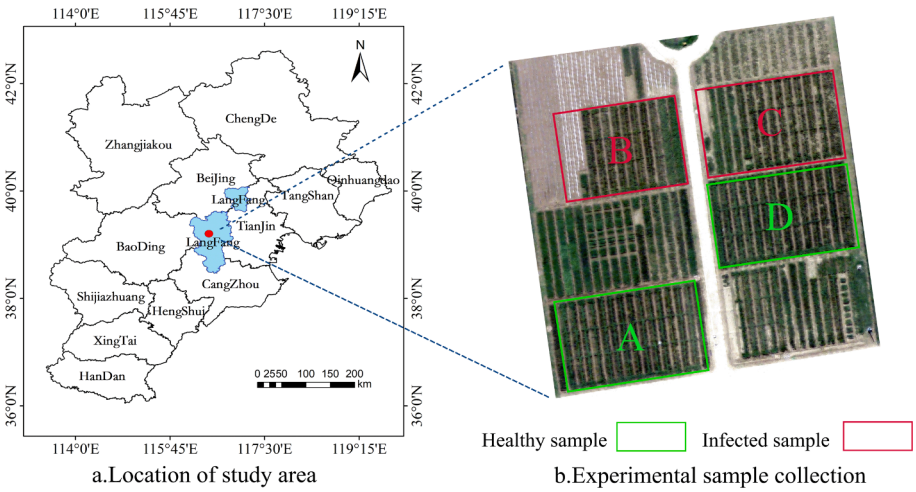


Fig. 1. Overview map of the experimental site.

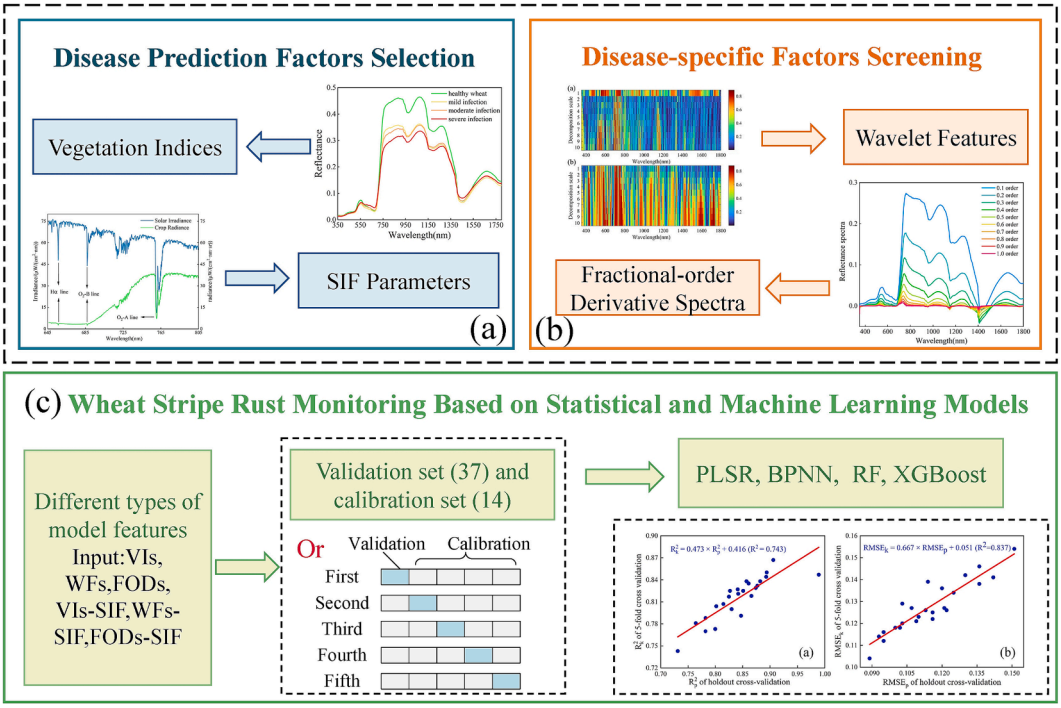


Fig. 2. Methodological framework of the monitoring model for stripe rust.

and variable importance in projection, exploring disease-specific factors (WFs and FODs). Finally, through collaboration among features, six feature sets were received and then imported to monitoring models constructed by multiple algorithms (PLSR, BPNN, RF, and XGBoost). Additionally, this study evaluated and compared various models based on the holdout cross-validation and k-fold cross validation, so as to determine the optimal feature-algorithm combination.

2.2.1. Selection of disease Prediction factors based on original spectra

Vegetation indices established by algebraic operations on sensitive bands can highlight spectral properties of the observed object, achieving disease monitoring from a pathological mechanism perspective (Jing et al., 2022). A total of 17 common spectral indices selected from the literature were calculated as the preliminary characteristics (Zeng et al., 2022). These indices include NPCI, NRI, SIPI, MCARI, and ARI, which are sensitive to pigment levels, as well as WI, NDWI, MSI, and RVSI, which can characterize water content. Moreover, SAVI, TVI, and NDVI can attenuate the impact of soil background on canopy spectra, while PRI and PhRI directly exhibit variations in photosynthetic rate. GNDVI measures vegetation greenness, MSR reflects leaf area, and PSRI can monitor vegetation senescence. Table 1 shows the name and calculation formula of the selected VIs.

Chlorophyll fluorescence immediately reflects the electron transfer rate in the photochemical reaction of photosynthesis (Gu et al., 2019), which is influenced by carbon assimilation in the dark reaction of photosynthesis. Therefore, SIF can serve as an ideal probe for non-destructive monitoring of plant photosynthesis. Among various algorithms for SIF retrieval, 3FLD is considered the most robust (Liu and Liu, 2014), which employs one band inside the Fraunhofer absorption line and two bands outside the line (one on each side) to calculate the chlorophyll fluorescence intensity F_{in}. In this study, the 3FLD algorithm was operated to retrieve the SIF intensity at the O_2-A and O_2-B bands, where the filling effect is prominent. For the O_2-B band, 686.989 nm was employed as the wavelength inside the O_2 feature, and the left and right

Table 1
List of 17 selected spectral vegetation indices.

Index	Description	Formula
NPCI	Normalized Pigment Chlorophyll Index	$(R_{680}-R_{430})/(R_{680}+R_{430})$
MSR	Modified Simple Ratio	$(R_{800}/R_{670}-1)/(R_{800}/R_{670}+1)^{1/2}$
SAVI	Soil Adjusted Vegetation Index	$1.5\times(R_{NIR}-R_R)/(R_{NIR}-R_R+0.5)$
NRI	Nitrogen Reflectance Index	$(R_{570}-R_{670})/(R_{570}+R_{670})$
GNDVI	Greenness Normalized Difference Vegetation Index	$(R_{NIR}-R_G)/(R_{NIR}+R_G)$
TVI	Triangular Vegetation Index	$0.5\times[130\times(R_{NIR}-R_R)-200\times(R_R-R_G)]$
NDVI	Normalized Difference Vegetation Index	$(R_{NIR}-R_R)/(R_{NIR}+R_R)$
WI	Water Index	R_{900}/R_{970}
PSRI	Plant Senescence Reflectance Index	$(R_{680}-R_{500})/R_{750}$
PRI	Photosynthetic Reflectance Index	$(R_{570}-R_{531})/(R_{570}+R_{531})$
NDWI	Normalized Difference Water Index	$(R_{857}-R_{1241})/(R_{857}+R_{1241})$
SIPI	Structural Independent Pigment Index	$(R_{800}-R_{445})/(R_{800}+R_{680})$
MSI	Moisture Stress Index	R_{1599}/R_{819}
MCARI	Modified Chlorophyll Absorption Reflectance Index	$[(R_{701}-R_{671})-0.2\times(R_{701}-R_{549})]/(R_{701}/R_{671})$
ARI	Anthocyanin Reflectance Index	$1/R_{550}-1/R_{700}$
PhRI	Physiological Reflectance Index	$(R_{550}-R_{531})/(R_{550}+R_{531})$
RVSI	Ratio Vegetation Structure Index	$(R_{712}+R_{752})/2-R_{732}$

shoulders were sampled at 686.338 and 688.126 nm, respectively, for the shorter wavelength and the longer wavelength. For the O_2-A band, 760.453 nm was employed as the wavelength inside the O_2 feature, and 758.984 nm and 768.618 nm were the wavelengths on the shoulders to the left and right of the O_2 feature, respectively.

$$\omega_{left}=\frac{\lambda_{right}-\lambda_{in}}{\lambda_{right}-\lambda_{left}}$$

$$\omega_{right}=\frac{\lambda_{in}-\lambda_{left}}{\lambda_{right}-\lambda_{left}} \tag{3}$$

$$I_{out}=\omega_{left}\cdot I_{left}+\omega_{right}\cdot I_{right}$$

$$L_{out}=\omega_{left}\cdot L_{left}+\omega_{right}\cdot L_{right} \tag{4}$$

$$F_{in}=\frac{L_{in}\cdot\left(\omega_{left}\cdot I_{left}+\omega_{right}\cdot I_{right}\right)-I_{in}\cdot\left(\omega_{left}\cdot L_{left}+\omega_{right}\cdot L_{right}\right)}{\left(\omega_{left}\cdot I_{left}+\omega_{right}\cdot I_{right}\right)-I_{in}} \tag{5}$$

Where λ_{left}, λ_{in}, and λ_{right} are the left, inner and right bands of the absorption band. ω_{left} and ω_{right} represent the weights of the left and right bands of the Fraunhofer line. I_{left}, I_{in}, I_{right}, and I_{out} are solar irradiance in the left, inside, right bands and outside the Fraunhofer line. L_{left}, L_{in}, L_{right}, and L_{out} represent canopy ascending radiance in the left, inner, right bands and outside the Fraunhofer line.

Furthermore, to mitigate the influence of external factors during different periods on the canopy SIF, this study chose relative SIF as the measure of canopy SIF.

$$F_{rel}=\frac{F_{in}}{I_{in}} \tag{6}$$

Where F_{rel} is the relative intensity of SIF.

In addition to the direct SIF retrieval at O_2-A and O_2-B, there exist two methodologies to quantify the fluorescence emission signals superimposed on the canopy reflectance. The first method is based on optical indices calculated from reflectance bands affected by fluorescence emission and then normalizing these to bands not affected by fluorescence emission. In the second method, the computation of derivative reflectance facilitates the detection of subtle variations caused by fluorescence emission in the red edge (650–800 nm), while concurrently minimizing other confounding effects. To this end, 10 reflectance ratio indices (Dobrowski et al., 2005) and first derivative indices (Zarco-Tejada et al., 2003), including D_{730}/D_{706}, R_{740}/R_{720}, D_{705}/D_{722}, $R^2_{685}/(R_{675}\cdot R_{690})$, R_{740}/R_{800}, $R^2_{683}/(R_{675}\cdot R_{691})$, $(R_{675}\cdot R_{690})/R^2_{683}$, R_{685}/R_{655}, R_{690}/R_{655} and R_{750}/R_{800}, were calculated to characterize fluorescence intensity.

2.2.2. Disease-specific factors screening based on optimized spectra

CWT is a powerful signal processing technique that allows for the decomposition of spectra at continuous wavelengths and scales (ZHANG et al., 2012), generating a two-dimensional wavelet power spectrum. The power spectrum consists of wavelet energy coefficients, which can be viewed as similarity between the wavelet basis function and reflectance.

$$\psi_{s,w}(\lambda)=\frac{1}{\sqrt{s}}\psi\left(\frac{\lambda-\omega}{s}\right) \tag{7}$$

Where $\psi_{s,w}(\lambda)$ is the wavelet basis function after deformation, both s and ω are positive real numbers, respectively representing scale factor and displacement factor (band).

$$W_f(\mathrm{s},\omega)=<f,\ \psi_{s,w}>=\int_{-\infty}^{+\infty}f(\lambda)\cdot\psi_{s,w}(\lambda)\cdot\mathrm{d}\lambda \tag{8}$$

Where $f(\lambda)$ is the reflectance spectra in the wavelength range of 350–1800 nm. $W_f(s_i,\omega_j)$ represents the wavelet energy coefficients, which contains two dimensions, one is the scale factor (i = 1, 2, 3,…, m), and the other is the band (j = 1, 2, 3,…, n).

Research has shown that decomposing vegetation spectra with wavelet basis functions, specifically db5 (Liu et al., 2011) and mexh (Koger et al., 2003), can alleviate noise while enhancing singular information. In this study, the original spectra were subjected to CWT using both db5 and mexh basis functions, with scales ranging from 2^1 to 2^{10}, respectively.

Additionally, FOD can perform fine interpolation between the original spectra and integral-derivative (Hong et al., 2019), serving as an effective preprocessing method for hyperspectral data. Among various expressions of FOD, Grunwald-Letnikov is extensively implemented due to its relatively simple calculation. Therefore, this study adopted this expression, specifically defined as:

$$d^qf(x)=\lim_{h\to\infty}\frac{1}{h^q}\sum_{m=0}^{(t-a)/h}(-1)^m\frac{\Gamma(q+1)}{m!\Gamma(q-m+1)}f(x-mh) \tag{9}$$

Where d represents the derivative function, q denotes the order, h is the step length, t and a are the upper and lower bounds of FOD. The Gamma (Γ) function is:

$$\Gamma(\beta)=\int_0^{\infty}e^{-t}t^{\beta-1}dt=(\beta-1)! \tag{10}$$

Where β is an arbitrary variable, especially the derivative order.

Let f(λ) be one-dimensional spectrum, and [a, t] be the wavelength interval, where λ ∈ [a,t]. Since the resampling interval of ASD Field Spec ®Pro FR spectrometer is 1 nm, we set h = 1, n = (t − a)/h = t − a. Then, the difference expression of FOD can be derived from Equation (9) and

(10) as follows:

$$\frac{d^v f(\lambda)}{d\lambda^v} \approx f(\lambda) + (-v)f(\lambda-1) + \frac{(-v)(-v+1)}{2} f(\lambda-2) + \cdots + \frac{\Gamma(-v+1)}{n!\Gamma(-v+n+1)} f(\lambda-n) \tag{11}$$

When v is a fraction, the above equation represents the Grunwald-Letnikov form (Xue, 2019). In this study, the interval was set from 0.1 to 2.0 orders (0.1 is the interval step), and FOD was implemented in MATLAB 2021b.

To avoid model estimation distortion caused by multicollinearity among disease-specific factors, it is necessary to calculate VIP to quantify the explanatory power of WFs and FODs on DI (Chong and Jun 2005), and screen out input variables with larger contributions. Assume that there is a dependent variable y and a set of independent variables x_1, $x_2, \cdots\cdots, x_k$. For the j-th independent variable, its VIP can be obtained as follows:

$$VIP_j = \sqrt{\frac{k}{\sum_{h=1}^{m} r^2(y, c_h)} \sum_{h=1}^{m} r^2(y, c_h)\omega_{hj}^2} \tag{12}$$

Where k represents the number of independent variables, c_h is the principal component extracted by relevant independent variables, and ω_{hj} is the contribution of the independent variable to the principal component. $r(y, c_h)$ is the correlation coefficient between the dependent variable and the principal component, indicating the explanatory ability of the principal component to y. The explanatory ability of x_j on y is transmitted by the principal component c_h. If c_h has a strong explanation for y, and x_j is closely related to c_h, then x_j can be considered to play an important role in the interpretation of y.

2.2.3. Model algorithms and accuracy evaluation

To investigate effectiveness of canopy SIF in probing wheat stripe rust, six feature sets were constructed, considering VIs, FODs and WFs as elementary variables and whether to collaborate with SIF. Ensemble learning algorithms, such as RF and XGBoost were carried out to model, and their performance was compared with the statistical model PLSR and deep learning algorithm BPNN. In this study, a grid search method (GS) was deployed for hyperparameter tuning. These models were trained and validated employing two approaches, holdout cross-validation and k-fold cross validation, with evaluation indexes including determination coefficient (R^2), root mean squared error (RMSE), and relative percent deviation (RPD). 51 samples were stochastically divided into five folds, with four folds used as the training set and one fold used as the validation set. Then, 5 iterations were performed so that each fold of samples can serve as the validation set. Finally, the average of the evaluation indexes were calculated to depict the potential of monitoring models. For each feature-algorithm combination, 100 rounds of 5-fold cross validation were randomly executed, and the result with the highest accuracy was selected as the final evaluation metric.

3. Results

3.1. Disease response of vegetation indices and SIF parameters

The initial feature set derived from original spectra consisted of 17 vegetation indices and 12 canopy SIF parameters, which were analyzed with DI individually, and correlation coefficients between them are presented in Fig. 3. It was observed that winter wheat affected by stripe rust exhibited significant responses in photosynthetic rate, pigment content and leaf area. NPCI, as a normalized index that can estimate chlorophyll concentration, had the highest correlation coefficient with DI, reaching 0.862. Monitoring indicators such as MSR, SAVI, and NRI can effectively reveal the stress status. That's because once infected with stripe rust, wheat will consume more nutrition and moisture, with its chloroplasts damaged and photosynthesis declining. Subsequent studies selected 13 vegetation indices showing statistical significance (P-value < 0.001), including NPCI, MSR, and SAVI, as input features for the model.

In addition, SIF parameters also displayed sensibility to disease stress, with a significantly negative correlation found between DI and

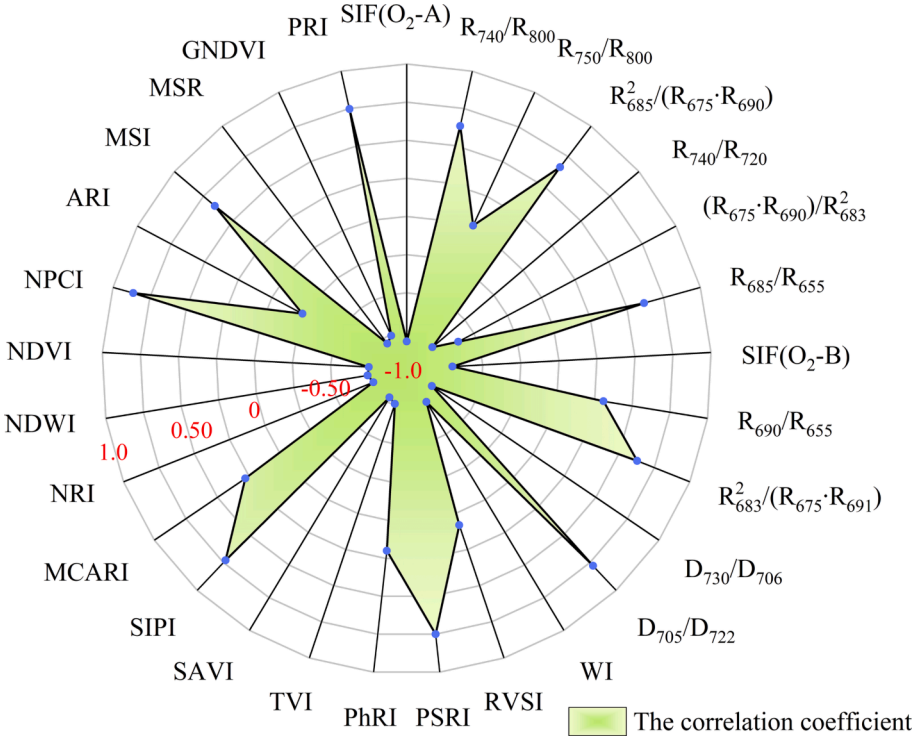


Fig. 3. Correlation between vegetation indices or SIF parameters and DI, $R_{0.1\%}[51] = 0.447$.

the relative SIF estimated by means of the 3FLD method in both O_2-A and O_2-B. Notably, the O_2-A band portrayed more distinct dark line feature and stronger fluorescence, with a correlation coefficient of −0.819 with DI, which was 17.2 % higher than that of the O_2-B band (-0.699). This negative correlation is attributed to the dominant reduction in chlorophyll content and photosynthetic activity with the worsening of the disease, leading to a decrease in $APAR_{chl}$, which is the primary driver of SIF. Besides, 8 reflectance ratio indices and first derivative indices, including D_{730}/D_{706}, R_{740}/R_{720} and D_{705}/D_{722}, have been traced to pass significance testing (P-value < 0.001). On the whole, a total of 10 SIF parameters can be included as modeling factors.

3.2. Sensitivity study of WFs and FODs to stripe rust

Fig. 4 illustrates the original spectra captured at different disease severity, along with the curve depicting the correlation between DI and reflectance. As demonstrated in Fig. 4a, the reflectance of infected wheat was marginally higher than that of healthy wheat in the visible region (560–700 nm), while it was evidently lower in the near-infrared region (750–1350 nm). Correspondingly, it was observed that DI exhibited positive correlation with reflectance in the range of 550–714 nm but negative correlation in the range of 714–1407 nm (Fig. 4b). The spectral discrepancy within the visible region might be associated with the content of pigments, and damage to chlorophyll caused by stripe rust prompted an increase in reflectance and a blue shift phenomenon. On the other hand, the cell structure of the leaves was a critical factor determining reflectance in the near-infrared region, whose trauma caused less reflection. It was discovered that 187 bands passed the positive significance test (P-value < 0.001), with the highest correlation coefficient being 0.674. Conversely, 608 bands showed a highly significant negative correlation (P-value < 0.001), with a maximum correlation coefficient of −0.709.

The original spectra were subjected to continuous wavelet transformation with the db5 and mexh basis functions at scales 1–10. Based on the wavelet energy coefficients and DI, the correlation matrices |R| were generated in Fig. 5. As shown in Fig. 5a, the db5 basis function at scale 1 produced the largest number of sensitive features, specifically 842 bands. These bands highly sensitive to DI were primarily located at the blue, orange, and red edges, as well as in the near-infrared region, indicating declines in the pigment and water content after infecting with stripe rust. In contrast, the mexh basis function generally outperformed db5 (Fig. 5b), with the number of bands passing significance testing (P-value < 0.001) increasing from 244 to 1046 as the scale raised from 1 to 10. Notably, four spectral intervals (420–439 nm, 693–770 nm, 1563–1614 nm, and 1691–1739 nm) exhibited correlation coefficients greater than 0.80. In summary, regardless of the basis function applied, CWT contributed to increasing the quantity of sensitive bands and detecting more responsive spectral features.

Fig. 6 displays the maximum correlation coefficients and corresponding bands between DI and canopy spectra for different scales and basis functions. When applying db5, the maximum correlation coefficient peaked at 0.889 with scale 6 and a wavelength of 701 nm. If employing mexh, the maximum correlation coefficient would peak at 0.877 with scale 1 and a wavelength of 714 nm. Regardless of the basis

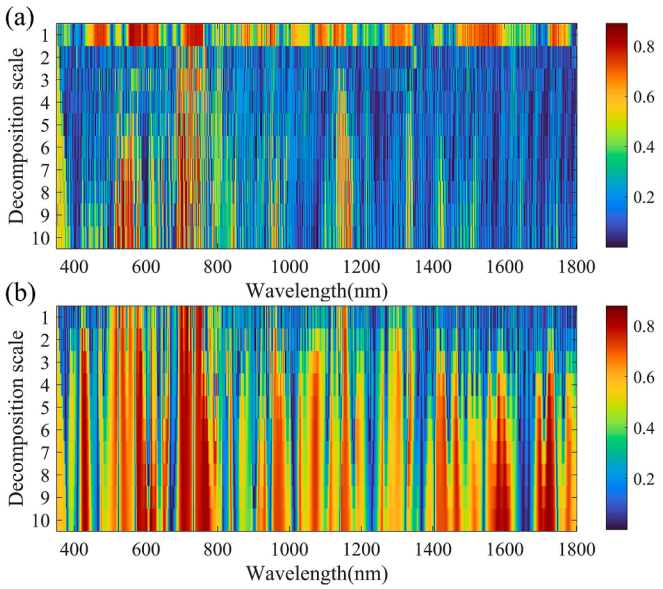


Fig. 5. Disease-sensitive wavelet regions extracted from spectra. The correlation coefficient matrix utilizing basis functions (a) db5 and (b) mexh.

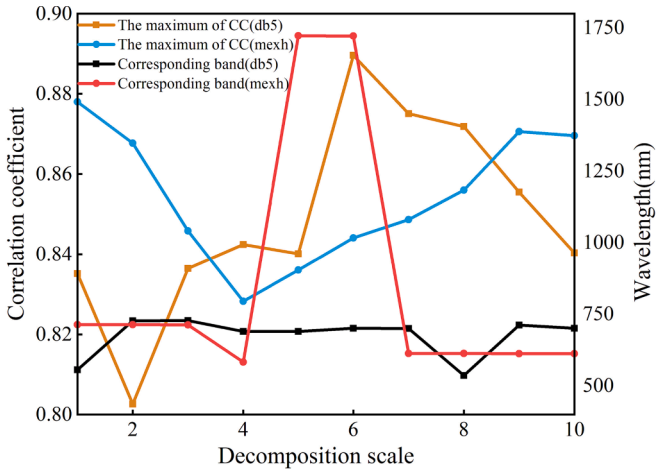


Fig. 6. The maximum correlation coefficients and corresponding bands between DI and canopy spectra for different scales and basis functions.

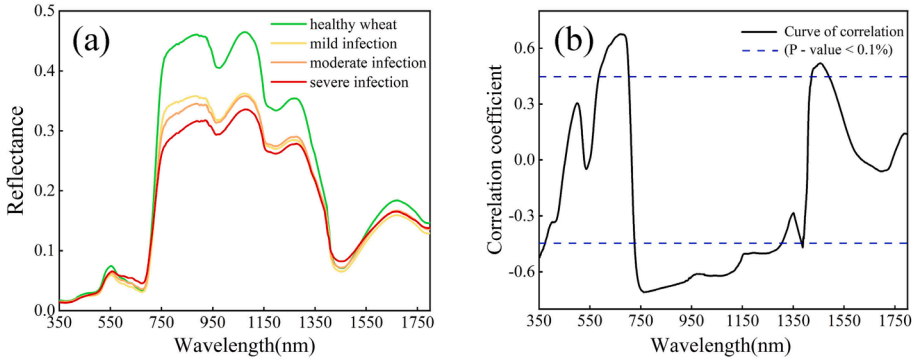


Fig. 4. Analysis based on canopy spectra. (a) the original spectra under different disease severity; (b) the curve of correlation coefficient between DI and reflectance.

Table 2
Detailed information of WFs after filtering.

The basis function	Scale	Wavelength/ nm	VIP value	The correlation coefficient
db5	6	701	1.039	−0.889
db5	6	700	1.033	−0.883
mexh	1	714	1.026	0.877
db5	7	700	1.022	−0.875
db5	8	536	1.018	0.871
mexh	9	612	1.017	0.870
mexh	10	612	1.016	0.869
db5	7	701	1.014	−0.868
mexh	2	714	1.013	0.867
mexh	1	715	1.010	0.865
db5	8	535	1.008	0.862
mexh	10	613	1.006	0.861
mexh	2	713	1.006	0.860
mexh	9	613	1.005	0.860
mexh	2	715	1.005	0.860

function and decomposition scale, the bands corresponding to the maximum correlation coefficient are generally distributed near the red edge. Two correlation matrices of size 10 × 1451 had been obtained, which were comprised of 29,020 correlation coefficients. Among them, 300 correlation coefficients were selected according to the descending order (the threshold of top 1 % was 0.814). Of these values, 76 came from db5 and 224 from mexh.

The WFs were ranked in a descending order, based on the absolute values of correlation coefficients. Subsequently, the top 50 features were selected as independent variables, whereas DI was chosen as the dependent variable for VIP calculation in the SIMCA 14.1 software. The sensitive WFs with VIP values in the top 15 were employed as input parameters for the disease monitoring model. More details can be found in Table 2, where all the selected variables' VIP values were greater than 1, indicating their strong explanatory power for DI.

To investigate the spectral profiles of wheat under various derivative orders and disease severity, 51 samples were categorized into three groups based on DI: "healthy and mild samples" (0 ≤ DI < 20 %), "moderate samples" (20 % ≤ DI < 45 %), and "severe samples" (45 % ≤ DI < 100 %). The average reflectance was calculated and fractional-order derivatized from 0.1 to 2.0 orders, as depicted in Fig. 7. With lower-order derivatives, the spectra exhibit relatively smooth contour and gradual alterations in reflectance. As the order increases, the outline surrounding the peak noticeably changes, with the values of FOD approaching 0, ultimately characterizing the disappearance of baseline drift and mixed overlap peaks. Moreover, it is evident that the range of reflectance spectra narrows as the disease of wheat worsens. Fig. 7 illustrates that fractional-order derivative, in contrast to conventional integral-derivative, provides a more detailed characterization of the spectral profile, enabling a better exploration of spectral intricacies.

Correlation analysis was implemented to disclose the linear association between spectra and DI, generating the correlation coefficient curve as drawn in Fig. 8. Fig. 9a illustrates the relationship between derivative order and the number of bands that pass significance testing (P-value < 0.001). Overall, as the derivative order increases, the number of characteristic bands initially climbs up and then diminishes. The quantity of characteristic bands peaks at a derivative order of 0.8 and then decreases to its minimum value at an order of 2.0. With the increase in derivative order from 0.1 to 0.8, the count of characteristic bands (P-value < 0.001) escalates from 742 to 918, while statistically significant correlations are observed at 442–501 nm and 1724–1800 nm. As the derivative order raises from 0.9 to 2.0, the curve manifests greater volatility, accompanied by a precipitous decline in the number of characteristic bands.

The maximum correlation coefficients and corresponding bands between DI and canopy spectra for various derivative orders are given in Fig. 8b. As a general trend, the value of maximum correlation coefficient enhances as the derivative order rises. At the derivative order of 1.2, the maximum correlation coefficient reaches its peak at 0.888, which is 25.2 % higher than the original spectra (0.709), 6.3 % higher than the spectra in 1.0 order derivative (0.835), and 2.5 % higher than the spectra in 2.0 order derivative (0.866). This suggests that original spectra and integral-derivative are not conducive in exploring detailed information concerning the disease, whereas FOD contributes to a better

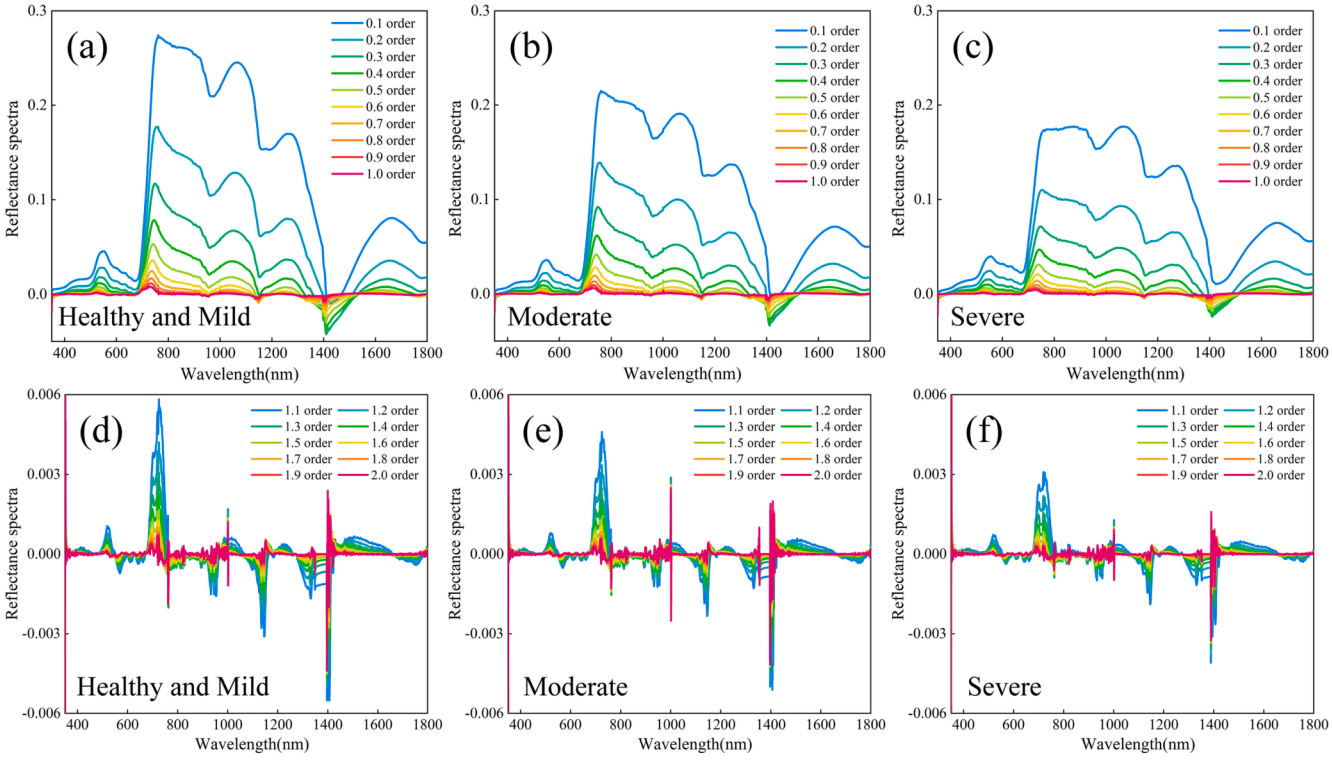


Fig. 7. The gradual transition of wheat spectra with different disease severity in 0.1–2.0 orders derivative.

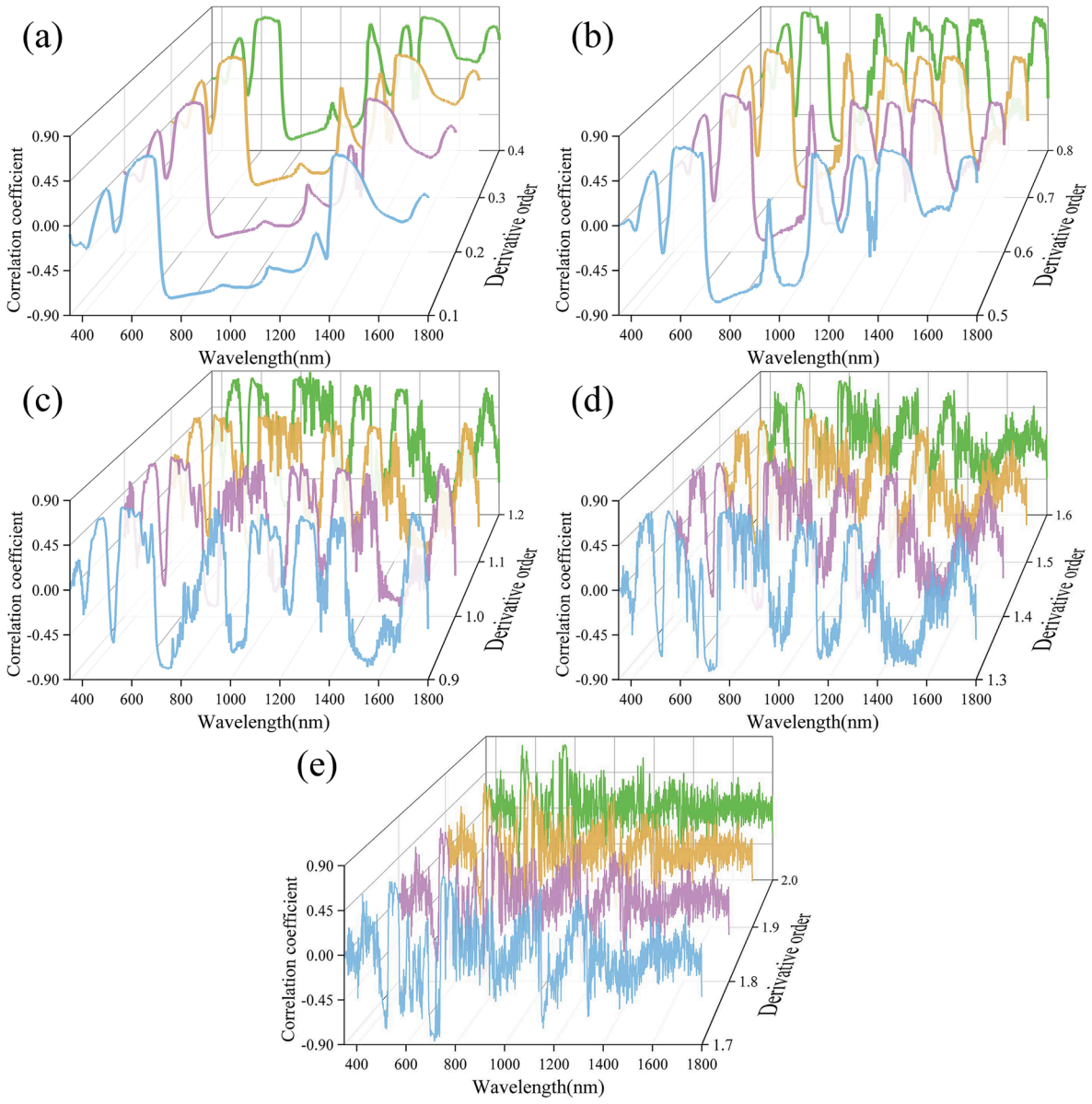


Fig. 8. The curve of correlation coefficient between FOD spectra and DI.

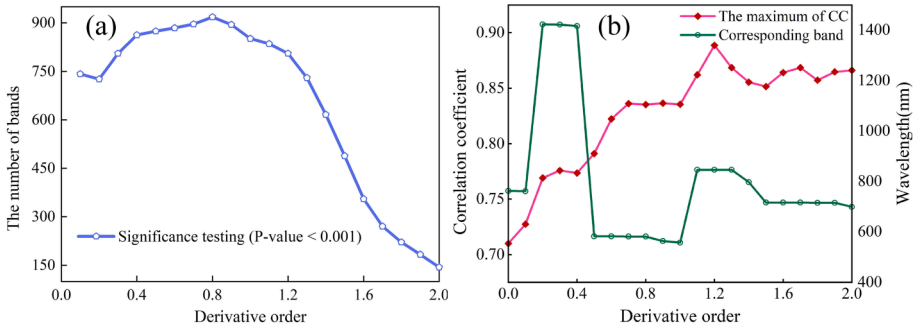


Fig. 9. (a) the association between derivative order and the quantity of sensitive bands (P-value < 0.001); (b) the maximum correlation coefficients and corresponding bands between DI and canopy spectra for various derivative orders.

Table 3
Detailed information of FODs after screening.

Derivative order	Wavelength/nm	VIP value	The correlation coefficient
1.1	846	1.320	0.862
1.1	847	1.231	0.844
1.2	846	1.177	0.888
1.3	846	1.166	0.868
1.2	847	1.073	0.855
1.8	716	1.070	−0.856
2.0	715	1.017	−0.863
1.7	716	1.010	−0.868
1.6	715	0.996	−0.834
1.4	717	0.994	−0.833
1.6	724	0.992	−0.833
1.4	716	0.991	−0.835
1.5	797	0.989	0.849
1.9	715	0.988	−0.864
2.0	699	0.986	−0.866

estimate of DI. Furthermore, the distribution of sensitive bands remains consistent, primarily centered around 581 nm, 716 nm, 762 nm, 846 nm, and 1422 nm.

Similarly, the disease-specific FODs with VIP values in the top 15 can be obtained, and Table 3 shows the corresponding derivative order, wavelength and so on.

3.3. Comparison of model results between two cross validation methods

3.3.1. Model evaluation based on holdout cross-validation

A total of 51 samples were randomly partitioned into two groups, resulting in a calibration dataset consisting of 37 samples (33 diseased and 4 healthy) and a validation dataset comprising 14 samples (12 diseased and 2 healthy). Statistical analysis with Origin software revealed that the minimum DI for all samples was 0, the maximum was 0.952, the mean was 0.532, the standard deviation was 0.317, and the coefficient of variation was 59.586 %. Importantly, there was no significant difference in mean and standard deviation between the calibration and validation datasets, indicating that they came from the same distribution and were comparable (Table 4).

Table 4
Statistical descriptions of DI on the total, calibration and validation datasets.

Sample Classification	Sample Number	Max	Min	Mean	Standard Deviation	Coefficient of Variation/%
Total dataset	51	0.952	0	0.532	0.317	59.586
Calibration dataset	37	0.952	0	0.571	0.318	55.691
Validation dataset	14	0.951	0	0.524	0.302	57.633

Table 5
Model evaluation based on holdout cross-validation.

Algorithm	Input features	Evaluation indexes		
		R_p^2	$RMSE_p$	RPD
PLSR	VIs	0.731	0.151	1.98
	VIs-SIF	0.782	0.136	2.24
	FODs	0.815	0.125	2.97
	FODs-SIF	0.841	0.116	3.10
	WFs	0.802	0.130	2.43
	WFs-SIF	0.827	0.121	2.74
BPNN	VIs	0.764	0.142	2.17
	VIs-SIF	0.825	0.122	2.82
	FODs	0.842	0.116	2.84
	FODs-SIF	0.861	0.109	3.03
	WFs	0.851	0.113	2.83
	WFs-SIF	0.857	0.110	2.79
RF	VIs	0.782	0.136	2.04
	VIs-SIF	0.830	0.120	2.16
	FODs	0.874	0.103	2.98
	FODs-SIF	0.893	0.095	2.75
	WFs	0.882	0.100	2.81
	WFs-SIF	0.894	0.095	2.77
XGBoost	VIs	0.800	0.130	1.98
	VIs-SIF	0.847	0.114	2.25
	FODs	0.866	0.107	2.37
	FODs-SIF	0.877	0.102	2.69
	WFs	0.898	0.093	2.80
	WFs-SIF	0.906	0.089	3.02

Table 6
Model evaluation based on 5-fold cross validation.

Algorithm	Input features	Evaluation indexes	
		R_k^2	$RMSE_k$
PLSR	VIs	0.743	0.154
	VIs-SIF	0.788	0.138
	FODs	0.807	0.134
	FODs-SIF	0.827	0.122
	WFs	0.804	0.129
	WFs-SIF	0.825	0.127
BPNN	VIs	0.781	0.141
	VIs-SIF	0.817	0.126
	FODs	0.821	0.125
	FODs-SIF	0.835	0.121
	WFs	0.825	0.126
	WFs-SIF	0.838	0.123
RF	VIs	0.770	0.146
	VIs-SIF	0.800	0.136
	FODs	0.829	0.120
	FODs-SIF	0.844	0.116
	WFs	0.838	0.118
	WFs-SIF	0.850	0.112
XGBoost	VIs	0.773	0.142
	VIs-SIF	0.791	0.139
	FODs	0.818	0.127
	FODs-SIF	0.832	0.118
	WFs	0.847	0.114
	WFs-SIF	0.867	0.104

In this study, 24 types of disease monitoring model were tested by holdout cross-validation and evaluated through multiple indexes, as shown in Table 5. Based on statistical results, the following conclusions can be drawn: 1. Regardless of the algorithm, models that collaborated with SIF perform better monitoring accuracy compared to those just utilizing a single type of feature. 2. Among the various algorithms applied, models constructed with WFs as independent variables generally demonstrate better than those using FODs and VIs, with VIs exhibiting the worst. This discovery remains valid when the input parameter is combined with the canopy SIF, respectively. 3. When assessing the efficiency of diverse algorithms, it was found that XGBoost and RF have the excellent capability for disease monitoring, while PLSR performs worst. The superlative algorithm for different feature types varied. When the input features were VIs or VIs-SIF, the model constructed by XGBoost algorithm was best, with R^2 of 0.800 and 0.847, RMSE of 0.130 and 0.114, as well as RPD of 1.98 and 2.25. When the input features were FODs or FODs-SIF, the model constructed by RF algorithm was optimal, with R^2 of 0.874 and 0.893, RMSE of 0.103 and 0.095, as well as RPD of 2.98 and 2.75. When the input features were WFs or WFs-SIF, the model constructed by XGBoost algorithm was best, with R^2 of 0.898 and 0.906, RMSE of 0.093 and 0.089, as well as RPD of 2.80 and 3.02.

Synthetically examining the impact of algorithms and feature categories on monitoring competency, it can be concluded that selecting XGBoost algorithm and WFs-SIF as input variables is the optimal solution. Compared to utilizing the PLSR algorithm with only VIs as input parameters, its R^2 shows an improvement of 23.9 %, RMSE drops by 41.0 % and RPD increases by 52.5 %. It is obvious that the appropriate combination of algorithms and feature can enormously upgrade the accuracy of disease monitoring models.

3.3.2. Comparative analysis between two different validations

Due to the small sample size in this study, fixed calibration and validation sets might lead to overfitting. Therefore, further evaluation was conducted based on 5-fold cross validation (Table 6). The comparison of Tables 5 and 6 displays high consistency between the conclusions obtained from two validation methods. The unique distinction is that when adopting VIs or VIs-SIF as input features, model through BPNN algorithm achieved the supreme accuracy, with R^2 of 0.781 and 0.817, as well as RMSE of 0.141 and 0.126. In addition, when the input features were FODs or FODs-SIF, RF algorithm was still optimal, with R^2 of 0.829 and 0.844, as well as RMSE of 0.120 and 0.116. When the input features were WFs or WFs-SIF, XGBoost algorithm was still optimal, with R^2 of 0.847 and 0.867, as well as RMSE of 0.114 and 0.104. In conclusion, when WFs-SIF serves as the input variable and XGBoost algorithm is applied, the accuracy is still the highest. Compared to the model taking PLSR algorithm and only VIs as input parameters, its R^2 increases by 16.6 % and RMSE decreases by 32.4 %.

4. Discussion

Spectral response of crop diseases forms the basis of remote sensing monitoring. Previous studies have indicated that the spectrum of crop under stress is the function of changes in pigments, moisture, morphology, structure, etc. As Fig. 3 shows, stripe rust results in discernable difference of wheat canopy spectra, which becomes more pronounced as the disease worsens. Similar observations were also reported by Guo et al. (2021) and Wang et al. (2016). In this study, 17 vegetation indices were initially selected as candidate features, and following correlation analysis, 13 sensitive features, such as NPCI, MSR, and SAVI, were identified through significant testing (P-value < 0.001). These indices are strongly correlated with photosynthesis efficiency, pigment content and LAI, which can reflect changes in the physicochemical components of diseased crops. Based on the 3FLD algorithm, relative SIF was extracted in the O_2-A and O_2-B bands, whose correlation coefficient with DI reached −0.819 and −0.699. Furthermore, 8 reflectance ratio indices and first derivative indices also exhibited high sensitivity (P-value < 0.001) towards disease infection. When crops are subjected to disease stress, chlorophyll fluorescence undergoes attenuation prior to those in chlorophyll content, illustrating high accuracy and sensitivity in quantifying vegetation stress and physiological variations (Kimm et al., 2021).

In this study, correlation analysis was performed between the original reflectance and DI across the entire spectral region. A total of 795 bands demonstrated significant correlation (P-value < 0.001), with the highest correlation coefficient of −0.709 observed at 762 nm near the red edge, which is consistent with the finding of Zheng et al. (2018). However, the original spectra's response to disease stress is relatively limited, manifested as low correlation with DI. To improve the sensitivity of spectra in detecting rust stress while mitigating the impact of uncontrollable factors in observations, mathematical transformations such as CWT and FOD were employed. Applying the db5 as basis function in CWT, the number of sensitive bands reaches a maximum of 842 when the scale is 1; the maximum correlation with DI is obtained at 701 nm with scale of 6, yielding a value of 0.889. Comparatively, the mexh basis function results in a peak of 1046 sensitive bands at scale of 10, and the highest correlation coefficient (0.877) occurs at 714 nm with scale of 1. The comparison shows that mexh is more helpful in highlighting the beneficial spectral information. This may be due to the fact that the shape of the basis function mexh is closer to the peaks and valleys in the reflectance spectra and exhibits a strict symmetry, which partly reduces phase distortion during signal decomposition and reconstruction. The vanishing moment order of the wavelet basis db5 may be excessively high, resulting in a weakened time-domain tight support, deteriorating real-time performance in neighboring bands during decomposition, which is not conducive to mining disease-specific factors. The FOD results indicate that this technique facilitates gradual changes in spectral shape, captures small differences, and enhances features, with the maximum correlation coefficient of 0.888 achieved at a derivative order of 1.2, an improvement of 0.179 compared to the original spectra.

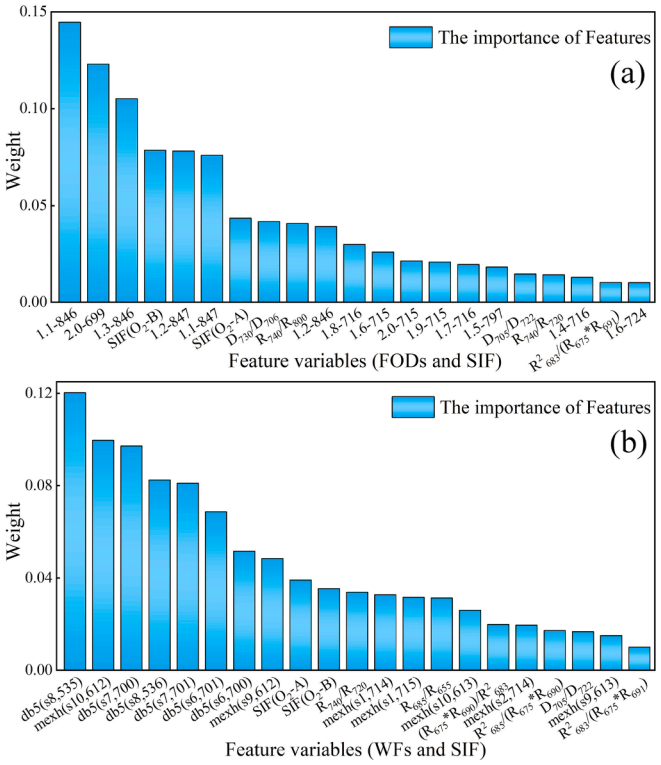


Fig. 10. The importance distributions of multiple features. (a) RF algorithm employing FODs-SIF as input variables; (b) XGBoost algorithm employing WFs-SIF as input variables.

During model evaluation based on holdout cross-validation, RF achieved the highest validation accuracy ($R_p^2 = 0.893$, $RMSE_p$=0.095, RPD = 2.75) among models using FODs-SIF as independent variables; while XGBoost exhibited the highest validation accuracy ($R_p^2 = 0.906$, $RMSE_p$=0.089, RPD = 3.02) among models using WFs-SIF as independent variables. Compared to PLSR and BPNN, ensemble learning algorithms based on decision trees, such as RF and XGBoost, possess feature selection mechanisms and allow flexible control over tree depth and complexity. Furthermore, diversity-enhancing strategies and cost function regularization terms enable them to effectively address nonlinear problems and mitigate overfitting phenomena. Both algorithms are ensemble learning methods that rely on decision trees, and feature importance can be evaluated by quantifying the contribution of input variables to DI, as illustrated in Fig. 10. If RF algorithm was employed, the spectral feature at 846 nm obtained the highest importance score of 0.1447 when the derivative order was 1.1. This is because vegetation internal structure and biomass can enormously influence the canopy reflectance in the near-infrared region, while disease invasion may result in the collapse of physical structures among cells. Additionally, the relative SIF at O_2-B and O_2-A is also crucial in predicting DI. This underscores the fact that stripe rust infestation brings about modifications in the photosynthetic physiology and vegetation structure of winter wheat. These alterations subsequently result in fluctuations in parameters such as $fAPAR_{chl}$、ϕF and f_{esc}. Ultimately, the aforementioned process endows the observed SIF with the ability to serve as a health probe (Xu et al., 2021). On the other hand, the WFs at 535 nm, 612 nm, and 700 nm are the primary control factors for disease monitoring when using the XGBoost algorithm. The reflection properties at 535 nm are predominantly governed by chlorophyll, whose degradation in wheat due to rust fungus infection can considerably impact the

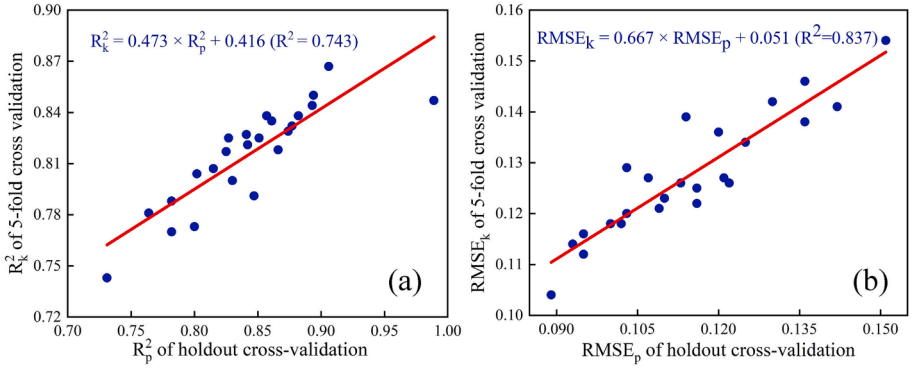


Fig. 11. Linear regression of evaluation indexes from two validation methods. (a) R^2; (b) RMSE.

reflectance within this spectral band. 612 nm is located in the yellow edge (550–650 nm), which is mainly dominated by the overlapping absorption of several pigments (particularly chlorophyll and anthocyanin). 700 nm is within the red edge, which is an essential indicator band for describing pigment status and health level.

To corroborate the consistency of the two validation methods and to reinforce the credibility of the study findings, a correlation analysis was conducted between the evaluation indexes, setting R_p^2 and $RMSE_p$ of holdout cross-validation as independent variables, R_k^2 and $RMSE_k$ of 5-fold cross validation as dependent variables. As shown in Fig. 11, both categories of indexes meet the standard of high correlation. The formulas after fitting are: $R_k^2 = 0.473 \times R_p^2 + 0.416$, $RMSE_k = 0.667 \times RMSE_p + 0.051$. Determination coefficients are 0.743 and 0.837 respectively, which suggests that the results of two accuracy test are mutually verified and the conclusions drawn around the optimal feature-algorithm combination are quite reliable.

In this study, CWT and FOD were performed for the full-band spectra, and disease-specific spectral feature sets were obtained based on two feature screening strategies, which fully exploited the physiological and biochemical properties of the stressed crops. Additionally, the inversion of SIF efficiently revealed changes in crop photosynthetic physiology. Ultimately, the advantages of reflectance spectroscopy and chlorophyll fluorescence for remote sensing monitoring of stripe rust were effectively combined based on a collaborated scheme.

Although this study has yielded satisfactory results in accurately monitoring wheat stripe rust, there remain several challenges that require attention in future research. First, fungal infection leads to cellular damage, subsequently altering physiological and biochemical components, directly impacting the canopy fluorescence emission rate (ϕF). In winter wheat subjected to prolonged disease infection, leaf morphology, leaf inclination distribution, and other factors also undergo changes, which affects the fluorescence escape fraction (f_{esc}) and the fraction of PAR absorbed by chlorophyll ($fAPAR_{chl}$). Therefore, it is necessary to distinguish between physiological and non-physiological responses of winter wheat to stripe rust stress. Second, the canopy spectra were measured in a vertical downward manner around noon, without taking into account the influence of direction during SIF retrieval. In tower-based or satellite observations, viewing angles and solar zenith angles often vary, and further research is required to investigate whether the conclusions can be extrapolated under other situations. Furthermore, it is imperative in future research to intensify the integration of pathological and remote sensing mechanisms, while considering meteorological information, in order to improve the adaptability of the model in complex farmland environments.

5. Conclusion

This study proposes a reliable method for monitoring wheat stripe rust by combining WFs with canopy SIF, which could not only accurately capture disease-specific factors but also comprehensively consider photosynthetic and physiological information. First, VIs and SIF parameters that could perceptively reveal disease stress were extracted based on correlation coefficient analysis. Next, the full-band spectra were mathematically transformed at distinct scales and orders using CWT and FOD. Comparatively, CWT proved to be more advantageous in refining spectral information and uncovering sensitive features. Two types of features (WFs and FODs) were retrieved after performing variable importance in projection, which exhibited high explanatory power for DI and weak multicollinearity. Finally, through collaboration among features, six feature sets were received and then imported to monitoring models constructed by multiple algorithms (PLSR, BPNN, RF, and XGBoost).

Holdout cross-validation and 5-fold cross validation, were employed to validate the accuracy and determine the optimal feature-algorithm combination. The results demonstrate that: (1) the collaboration of canopy SIF with any feature markedly improved the monitoring accuracy due to its responsive nature to the plant's photosynthetic physiology. (2) the model based on XGBoost with WFs-SIF as input features achieved optimal monitoring accuracy, with at least 16.6 % increase in R^2 and 32.4 % reduction in RMSE compared to the VIs-PLSR model. (3) correlation analysis of evaluation indexes (R^2 and RMSE) under two cross-validation methods showed determination coefficients of 0.743 and 0.837, indicating mutual validation and high reliability of the conclusions. Overall, the collaboration between WFs and SIF shows considerable feasibility in high-precision monitoring of wheat stripe rust. Considering the scale effect of SIF and incorporating pathological mechanisms and meteorological elements will be important components of our future research in this domain.

CRediT authorship contribution statement

Kehui Ren: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization. **Yingying Dong:** Investigation, Resources, Data curation, Writing – review & editing, Project administration. **Wenjiang Huang:** Writing – review & editing, Supervision, Funding acquisition. **Anting Guo:** Methodology, Investigation, Resources, Data curation. **Xia Jing:** Methodology, Investigation, Resources, Data curation.

Declaration of Competing Interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was funded by the National Key R&D Program of China (2021YFB3901300), Alliance of International Science Organizations (Grant No. ANSO-CR-KP-2021-06), National Natural Science Foundation of China (32271986), Chinese Academy of Sciences (183611 KYSB20200080) and SINO-EU, Dragon 5 proposal (ID 57457). The authors would like to thank Chao Ruan and Xia Jing for collecting the field measurement data at Langfang City.

References

Abdulridha, J., Ehsani, R., Abd-Elrahman, A., Ampatzidis, Y., 2019. A remote sensing technique for detecting laurel wilt disease in avocado in presence of other biotic and abiotic stresses. Computers and Electronics in Agriculture 156, 549–557.

Ashourloo, D., Mobasheri, M.R., Huete, A., 2014. Developing two spectral disease indices for detection of wheat leaf rust (Pucciniatriticina). Remote Sensing 6, 4723–4740.

Bolón-Canedo, V., Alonso-Betanzos, A., 2019. Ensembles for feature selection: A review and future trends. Information Fusion 52, 1–12.

Chong, I.-G., Jun, C.-H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems 78, 103–112.

Dobrowski, S., Pushnik, J., Zarco-Tejada, P.J., Ustin, S.L., 2005. Simple reflectance indices track heat and water stress-induced changes in steady-state chlorophyll fluorescence at the canopy scale. Remote Sensing of Environment 97, 403–414.

Du, K., Jing, X., Zeng, Y., Ye, Q., Li, B., Huang, J., 2023. An Improved Approach to Monitoring Wheat Stripe Rust with Sun-Induced Chlorophyll Fluorescence. Remote Sensing 15, 693.

Elavarasan, D., Vincent, D.R., 2020. Reinforced XGBoost machine learning model for sustainable intelligent agrarian applications. Journal of Intelligent & Fuzzy Systems 39, 7605–7620.

Gu, L., Han, J., Wood, J.D., Chang, C.Y.Y., Sun, Y., 2019. Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions. New Phytologist 223, 1179–1191.

Guo, A., Huang, W., Dong, Y., Ye, H., Ma, H., Liu, B., Wu, W., Ren, Y., Ruan, C., Geng, Y., 2021. Wheat yellow rust detection using UAV-based hyperspectral technology. Remote Sensing 13, 123.

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J., 2020. Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing 12, 236.

Hong, Y., Liu, Y., Chen, Y., Liu, Y., Yu, L., Liu, Y., Cheng, H., 2019. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. Geoderma 337, 758–769.

Huang, L.-S., Jiang, J., Huang, W.-J., Ye, H., Zhao, J., Ma, H., Ruan, C., 2019a. Wheat yellow rust monitoring method based on Sentinel-2 image and BPNN model. Trans. Chin. Soc. Agric. Eng 35, 178–185.

Huang, L., Wu, Z., Huang, W., Ma, H., Zhao, J., 2019b. Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine. Applied Sciences 9, 3894.

Jiang, X., Zhen, J., Miao, J., Zhao, D., Shen, Z., Jiang, J., Gao, C., Wu, G., Wang, J., 2022b. Newly-developed three-band hyperspectral vegetation index for estimating leaf relative chlorophyll content of mangrove under different severities of pest and disease. Ecological Indicators 140, 108978.

Jiang, G., Zhou, K., Wang, J., Sun, G., Cui, S., Chen, T., Zhou, S., Bai, Y., Chen, X., 2022a. Estimation of Rock Copper Content Based on Fractional-order Derivative and Visible Near-infrared–Shortwave Infrared Spectroscopy. Ore Geology Reviews 105092.

Jing, X., Zou, Q., Yan, J., Dong, Y., Li, B., 2022. Remote Sensing Monitoring of Winter Wheat Stripe Rust Based on mRMR-XGBoost Algorithm. Remote Sensing 14, 756.

Jones, C., Jones, J., Lee, W., 2010. Diagnosis of bacterial spot of tomato using spectral signatures. Computers and Electronics in Agriculture 74, 329–335.

Kimm, H., Guan, K., Jiang, C., Miao, G., Wu, G., Suyker, A.E., Ainsworth, E.A., Bernacchi, C.J., Montes, C.M., Berry, J.A., 2021. A physiological signal derived from sun-induced chlorophyll fluorescence quantifies crop physiological response to environmental stresses in the US Corn Belt. Environmental Research Letters 16, 124051.

Koger, C.H., Bruce, L.M., Shaw, D.R., Reddy, K.N., 2003. Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (Ipomoea lacunosa) in soybean (Glycine max). Remote Sensing of Environment 86, 108–119.

Li, X., Yang, C., Huang, W., Tang, J., Tian, Y., Zhang, Q., 2020. Identification of cotton root rot by multifeature selection from sentinel-2 images using random forest. Remote Sensing 12, 3504.

Liu, M., Liu, X., Wu, L., Duan, L., Zhong, B., 2011. Wavelet-based detection of crop zinc stress assessment using hyperspectral reflectance. Computers & Geosciences 37, 1254–1263.

Liu, X., Liu, L., 2014. Assessing band sensitivity to atmospheric radiation transfer for space-based retrieval of solar-induced chlorophyll fluorescence. Remote Sensing 6, 10656–10675.

Ma, H., Huang, W., Jing, Y., Dong, Y., Zhang, J., Nie, C., Tang, C., Zhao, J., Huang, L., 2017. Remote sensing monitoring of wheat powdery mildew based on AdaBoost model combining mRMR algorithm. Transactions of the Chinese Society of Agricultural Engineering 33, 162–169.

Mustafa, G., Zheng, H., Khan, I.H., Tian, L., Jia, H., Li, G., Cheng, T., Tian, Y., Cao, W., Zhu, Y., 2022. Hyperspectral Reflectance Proxies to Diagnose In-Field Fusarium Head Blight in Wheat with Machine Learning. Remote Sensing 14, 2784.

Ren, Y., Huang, W., Ye, H., Zhou, X., Ma, H., Dong, Y., Shi, Y., Geng, Y., Huang, Y., Jiao, Q., 2021. Quantitative identification of yellow rust in winter wheat with a new spectral index: Development and validation using simulated and experimental data. International Journal of Applied Earth Observation and Geoinformation 102, 102384.

Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. Computers and Electronics in Agriculture 72, 1–13.

Shi, Y., Huang, W., González-Moreno, P., Luke, B., Dong, Y., Zheng, Q., Ma, H., Liu, L., 2018. 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 Sensing 10, 525.

Tian, L., Wang, Z., Xue, B., Li, D., Zheng, H., Yao, X., Zhu, Y., Cao, W., Cheng, T., 2023. A disease-specific spectral index tracks Magnaporthe oryzae infection in paddy rice from ground to space. Remote Sensing of Environment 285, 113384.

Wang, J., Jing, Y., Huang, W., Zhang, J., Zhao, J., Zhang, Q., Wang, L., 2015. Comparative research on estimating the severity of yellow rust in winter wheat. Spectroscopy and Spectral Analysis 35, 1649–1653.

Wang, H., Qin, F., Ruan, L., Wang, R., Liu, Q., Ma, Z., Li, X., Cheng, P., Wang, H., 2016. Identification and severity determination of wheat stripe rust and wheat leaf rust based on hyperspectral data acquired using a black-paper-based measuring method. PLoS One1 11, e0154648.

Xu, S., Atherton, J., Riikonen, A., Zhang, C., Oivukkamäki, J., MacArthur, A., Honkavaara, E., Hakala, T., Koivumäki, N., Liu, Z., 2021. Structural and photosynthetic dynamics mediate the response of SIF to water stress in a potato crop. Remote Sensing of Environment 263, 112555.

Xue, D., 2019. FOTF toolbox for fractional-order control systems. Applications in Control 6, 237–266.

Zarco-Tejada, P.J., Pushnik, J., Dobrowski, S., Ustin, S., 2003. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. Remote Sensing of Environment 84, 283–294.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J.M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., 2022. Optical vegetation indices for monitoring terrestrial ecosystems globally. Nature Reviews Earth & Environment 3, 477–493.

Zhang, J., 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. Computers and Electronics in Agriculture 100, 79–87.

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., Huang, W., 2019. Monitoring plant diseases and pests through remote sensing technology: A review. Computers and Electronics in Agriculture 165, 104943.

Zhang, J., Jing, X., Song, X., Zhang, T., Duan, W., Su, J., 2023. Hyperspectral estimation of wheat stripe rust using fractional order differential equations and Gaussian process methods. Computers and Electronics in Agriculture 206, 107671.

ZHANG, J.-c., Lin, Y., WANG, J.-h., HUANG, W.-j., CHEN, L.-p., ZHANG, D.-y., 2012. Spectroscopic leaf level detection of powdery mildew for winter wheat using continuous wavelet analysis. Journal of Integrative Agriculture 11, 1474-1484.

Zheng, Q., Huang, W., Cui, X., Dong, Y., Shi, Y., Ma, H., Liu, L., 2018. Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. Sensors 19, 35.