

Unmanned aerial vehicle hierarchical detection of leaf blast in rice crops based on a specific spectral vegetation index

Guangming LI¹, Dongxue ZHAO¹, Jinpeng LI¹, Shuai FENG (✉)^{1,2}, Chunling CHEN (✉)^{1,2}

1 College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China.

2 Liaoning Key Laboratory for Intelligent Agricultural Technology, Shenyang 110866, China.

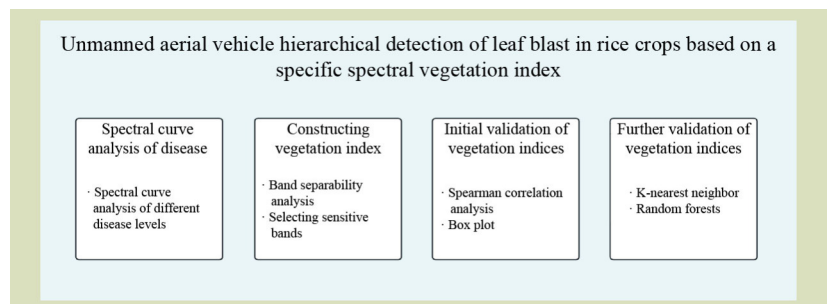
KEYWORDS

Drone remote sensing technology, hyperspectral technology, leaf blast disease, rice, vegetation index

HIGHLIGHTS

- A new vegetation index, rice blast index (RBI), was constructed to detect rice leaf blast.
- The disease detection performance of RBI, TVI, DDI and MTVI1 vegetation indices were compared.
- The level of leaf blast disease in the field was evaluated using the new RBI.

GRAPHICAL ABSTRACT



ABSTRACT

Leaf blast is a significant global problem, severely affecting rice quality and yield, making swift, non-invasive detection crucial for effective field management. This study used hyperspectral remote sensing technology via an unmanned aerial vehicle to gather spectral data from rice crops. ANOVA and the Relief-F algorithm were used to identify spectral bands sensitive to the disease and developed a new vegetation index, the rice blast index (RBI). This RBI was compared with 30 established vegetation indexes, using correlation analysis and visual comparison to further shortlist six superior indexes, including RBI. These were evaluated using the K-nearest neighbor (KNN) and random forests (RF) classification models. RBI demonstrated superior detection accuracy for leaf blast in both the KNN model (95.0% overall accuracy and 93.8% kappa coefficient) and the RF model (95.1% overall accuracy and 92.5% kappa coefficient). This study highlights the significant potential of RBI as an effective tool for precise leaf blast detection, offering a powerful new mechanism and theoretical basis for enhanced disease management in rice cultivation.

Received November 29, 2023;

Accepted June 21, 2024.

Correspondences: fengshuai@syau.edu.cn,
chenchunling@syau.edu.cn

1 Introduction

Rice accounts for a third of the global cereal production^[1] and

provides the primary food source for two-thirds of the global population^[2]. Particularly in China, where the rice cultivation area accounts for a fifth of the global cultivation area, it is

consumed as a staple food by over 65% of the population. Therefore, a high yield of rice is critical for food security and human health. However, leaf blast disease, a type of rice blast disease, that has become prevalent in over 85 countries worldwide^[3], poses a significant threat to global rice production, causing severe yield losses each year. Statistical data indicates that leaf blast disease causes yield losses as high as 10%–30% each year, and in severe cases, no grain can be harvested^[4]. Given these circumstances, the efficient and accurate detection of rice leaf blast disease becomes of paramount importance.

Established methods for detecting rice blast disease require experienced and knowledgeable professionals to sample rice crops and estimate the disease status of rice plants, resulting in significant labor costs. In contrast, polymerase chain reaction (PCR) is a commonly used biochemical detection method that can also be applied to detect rice blast disease. While PCR offers rapid and accurate detection, its complexity, destructive nature and time-consuming process make it unsuitable for meeting the efficient and non-destructive detection requirements of modern agriculture. Faced with the limitations of established rice blast disease detection methods and PCR technology, scientists have turned to more efficient hyperspectral imaging techniques. Hyperspectral technology has shown potential for application in various fields such as crop growth parameters^[5], crop classification^[6], biomass quantitative analysis^[7], and crop disease detection^[8], providing detailed information on crop conditions using a wide range of electromagnetic spectra^[9]. Gao et al.^[10] used hyperspectral imaging systems to collect data on grapevine leafroll-associated virus-infected grapevine leaves, and through spectral normalization and Monte Carlo method, preprocessing and removing outliers. Subsequently, they used the minimum noise fraction and selection operator to locate sensitive bands, tested wavelength sensitivity through variance analysis and linear regression, and finally validated wavelength effectiveness using a least squares support vector machine, confirming the potential application of hyperspectral image in non-destructive detection. Li et al.^[11] studied the hyperspectral reflectance of pine needles infected with pine wilt disease. They combined competitive adaptive reweighted sampling with the continuous projection algorithm to select sensitive bands and classified samples at different infection stages using linear discriminant analysis. Zhao et al.^[12] used hyperspectral imaging technology to collect spectral data of tea leaves subjected to tea green leafhopper, anthracnose and sunburn stress. They performed feature extraction using continuous wavelet analysis and constructed three tea tree stress identification and

discrimination models using the random forests algorithm. Although the aforementioned studies demonstrated the technical potential through hyperspectral data collection on leaves in well-controlled laboratory environments, their complex operations and slow detection make them unsuitable for the efficient detection demands of modern precision agriculture. Randive et al.^[13] successfully distinguished healthy and diseased cotton leaves using ASD FieldSpec 4 spectral radiometers, demonstrating the application of hyperspectral technology in disease diagnosis. Shukla et al.^[14] proposed a new method to improve the efficiency of detecting *Alternaria* blight in mustard using hyperspectral data in the range of 350–2500 nm, revealing spectral differences at different disease levels through correlation and continuum removal analysis. Zhang et al.^[15] classified naturally infected cotton leaves with *Fusarium* wilt using hyperspectral technology combined with various preprocessing techniques and algorithms, including multiplicative scatter correction, continuous wavelet analysis, the combination of support vector machine and genetic algorithm, grid search, particle swarm optimization, and gray wolf optimizer, to build accurate cotton *Fusarium* wilt grading models. The aforementioned studies were conducted in the field, with data directly collected from the canopy at close range. While accurate, this method is inefficient and not suitable for large-scale disease monitoring. Remote sensing technology^[16] enables the efficient acquisition of extensive data through non-destructive, non-contact means, providing detailed information on soil and crop conditions^[17]. Ma et al.^[18] used drones with hyperspectral cameras to detect wheat scab by analyzing spectral and texture features and optimizing neural networks with simulated annealing. In the experiment, the backpropagation neural network (BPNN) was improved through the integration of simulated annealing algorithms, and the improved BPNN was used to establish a wheat scab detection model. Moriya et al.^[19] demonstrated that hyperspectral imaging outperforms multispectral in identifying citrus gummosis. Zheng et al.^[20] found wavelet features effective in early detection of wheat yellow rust via support vector machines. Although drone-based data collection is efficient and preserves the crop, its accuracy can be compromised by environmental factors. Vegetation indices, calculated from specific spectral bands, have shown to effectively minimize background interference in remote sensing for reliable disease monitoring. Although efficient and non-destructive, unmanned aerial vehicle data collection may be affected by weather and environmental interference, whereas vegetation indices, by computing specific spectral bands, reduce background interference in remote sensing analysis due to their stability and anti-interference properties,

proving effective in disease monitoring^[21]. Liu et al.^[22] developed specific spectral indices using unmanned aerial vehicle hyperspectral sensors to detect rice leaf rollers. Ashourloo et al. ^[23] developed adaptive vegetation indices for detecting fire blight in pear orchards. In previous studies, established vegetation indices were also used to detect crop pest and disease conditions, but their accuracy and stability were lower. This may be because established vegetation indices have mainly focused on the overall health status of vegetation rather than the spectral response of specific diseases. For example, studies such as Gao et al.^[10] demonstrated the potential of hyperspectral image for non-destructive disease detection but highlighted the lower performance of established indices due to background interference and changes in light conditions. Similarly, Randive et al.^[13] showed that while established vegetation indices could distinguish healthy from diseased plants, their effectiveness varied significantly with different crop growth stages, leading to lower detection accuracy and stability. These studies indicate that customized vegetation indices can accurately reflect the spectral response of specific plant diseases, outperforming general indices, highlighting the importance of applying personalized hyperspectral indices in plant disease monitoring. In agricultural production, accurate and timely identification of crop diseases is crucial for ensuring food security and increasing crop yield. Established methods of disease identification, while effective under certain conditions, are often limited by factors such as high labor intensity, time consumption and potential damage to crops. In contrast, disease identification techniques using specific spectral indices, based on hyperspectral remote sensing technology, provide an efficient and non-invasive approach for early detection of crop diseases. This method not only allows for rapid identification of disease-affected vegetation without direct contact with the crops but also enables precise monitoring of diseases by analyzing subtle changes in crop reflectance spectra. Additionally, the application of specific spectral indices significantly expands the scope and efficiency of disease monitoring, making disease management in large-scale farmlands feasible^[24].

Consequently, the analysis of various existing research and methods in this study highlights the importance of establishing a vegetation index for the detection of rice leaf blast. This will facilitate large-scale, extensive, and more stable detection of the severity of rice leaf blast, meeting the needs of precision agriculture. Following key steps such as spectral preprocessing, data analysis, visual comparison and model verification, this study aims to establish a vegetation index for detecting rice leaf blast.

2 Materials and methods

2.1 Study area overview and experimental design

The study area was in Gengzhuang Town, Haicheng City, Anshan, Liaoning Province (40°58'58" N, 122°39'18" E), in the northwest of Haicheng City (Fig. 1). This region has a temperate continental monsoon climate, characterized by distinct seasons, concentrated precipitation, abundant sunshine and large temperature differences. The average annual precipitation is 652 mm and the annual average temperature is 8.4 °C. The soil is fertile and suitable for rice cultivation.

The experimental field had a planting area of 0.39 ha. The rice variety used for the study is Yanfeng 47, planted on 21 May 2022 with row spacing of 30 cm and plant spacing of 35 cm. The rice leaf blast fungus occurred naturally, with initial symptoms not being evident. From mid-July to early-August 2022, rice leaf blast of varying severity levels appeared. Therefore, experiments were conducted on 20 and 27 July, and 1 August on healthy rice plants and those with varying degrees of severity.

The experiment used a randomized complete block design to ensure the accuracy and representativeness of the research results. Prior to the experiment, we invited plant protection experts to conduct a preliminary assessment of the experimental fields. Based on the historical occurrence of rice blast disease and preliminary on-site survey results, the fields were divided into test blocks with different infection levels. Each block represented a certain degree of disease infection, ranging from no infection (level 0) to severe infection (level 4), with each level defined based on the disease index (DI), which comprehensively considers the proportion of leaf area covered by lesions and the severity of the disease. Additionally, detailed spectral data collection was conducted for each infection level of the rice crops using a drone-mounted hyperspectral sensor.



Fig. 1 Location and overview of the experimental site.

Also, to further validate the relationship between spectral data and actual disease infection levels, plant protection experts conducted on-site assessments and confirmed the disease grades of the plots on the first day of data collection (Fig. 2) with different colored markers represent different disease severity levels. Green areas indicate healthy plants, yellow, orange and red areas indicate mild, moderate-light and moderate disease severity, respectively, and dark red areas indicate severe disease. For example, we collected severe rice leaf blast disease data within the dark red areas. Within each marked area, the severity of rice blast disease was determined through on-site investigation and assessment by experienced rice disease experts and detailed records were made for each marked area.

2.2 Data collection

2.2.1 Grading of disease severity

The severity of leaf blast disease was quantified according to the national standard “Rice Blast Disease Survey and Report Standard (GB/T 15790-2009)” and the disease level was divided into five levels from 0 to 4 (Table 1) based on the disease index (in the experiment, no $DI > 50$ was found, so the highest level is 4).

Disease level	Severity of the disease	Disease index (DI)
0	Healthy	$DI = 0$
1	Mild	$0 < DI \leq 1$
2	Moderate-light	$1 < DI \leq 5$
3	Moderate	$5 < DI \leq 10$
4	Severe	$10 < DI \leq 50$

The disease level was determined by the disease index calculated as:

$$DI = \frac{\sum_i P_i \times D_i}{P \times D_m} \times 100 \tag{1}$$

where, P is the total number of plants, D_m is the representative value of the highest level, P_i is the value of each level, and D_i is the number of plants at each level.

2.2.2 Data collection and preprocessing

This study used a DJI Matrice 600 hexacopter drone equipped with a GaiaSky-mini hyperspectral imager to collect data. The spectral range of this setup is 400–1000 nm with a resolution of 3.5 nm encompassing 176 effective spectral bands. Specific equipment information is provided in Table 2. Data collection was conducted on clear, windless and cloudless days, between 11:00 and 14:00 during stable light intensity periods with the drone hovering at an altitude of 100 m.

To ensure the accuracy and consistency of the data, the collected hyperspectral images were preprocessed using Hyperscan Pro software, which included the following steps:

- (1) Radiometric correction: standardize the intensity of the spectral data,



Fig. 2 Hyperspectral image.

Device name	Gaiasky-mini-VN
Spectral range	400–1000 nm
Spectral resolution	3.5 nm
Spectral sampling rate	0.7 nm
Full-frame pixels	1932 × 1040
Pixel pitch	6.45 μm
Power	USB 2.0
Capture mode	45 W
Lens	23 mm

(2) Reflectance correction: account for varying surface reflectivity, and

(3) Area correction: adjust for specific regions of interest.

To achieve finer spectral data, the spectral resolution was interpolated to 1 nm. Hyperspectral data were then extracted using ENVI 5.3 software. A total of 250 regions of interest (ROIs) were established on plants marked with different disease levels. Each ROI was a regular rectangle of 3×3 pixels, with each disease level having 50 ROIs. The average spectral reflectance within each ROI was taken as the spectral reflectance for that disease level. The data set included spectral reflectance values in the range of 400–1000 nm, capturing the critical impact of the disease on the spectral properties of the plants. The data set was divided into training and testing sets with 200 samples used for training and 50 samples used for testing.

2.3 Feature selection methods and construction of vegetation indices

2.3.1 Feature selection method

The Relief-F algorithm, initially proposed by Kira and Rendell in 1992^[25], is a classic feature selection algorithm used for selecting important features from a given feature set. It has a certain robustness against noise and redundant features. The Relief-F algorithm was used to select sensitive bands, with the weight serving as the standard for band selection. The Relief-F algorithm is instance-based reasoning that evaluates the importance of each feature by comparing instances within a data set. The core idea of the algorithm is to determine the weight of features based on their contribution to the classification of instances. Specifically, the Relief-F algorithm estimates the importance of features by calculating the difference in features between each instance and its nearest neighbors. This difference value is used to update the weight of the feature, which is then sorted by weight to select the most important features.

The steps of the algorithm are as follows:

(1) Initialization: assign a weight $W[F] = 0$ to each feature F.

(2) For each instance x , iterate:

Find K-nearest neighbors (KNN) H from the same class.

Find KNN M from different classes.

(3) Update the weight of each feature:

$$W[F] = W[F] - \frac{\sum_{h \in H} \text{diff}(F, x, h)}{k \times h} + \frac{\sum_{m \in M} \text{diff}(F, x, m)}{k \times n} \quad (2)$$

where, n is the total number of samples, and $\text{diff}(F, x, y)$ calculates the difference in feature F between samples x and y .

(4) Feature selection: based on the final calculated feature weights $W[F]$, select the highest-weighted features for subsequent analysis.

2.3.2 Vegetation index construction

The vegetation index (VI) is an important indicator used to describe the state of vegetation growth. It can highlight vegetation information, extract features such as the growth state and coverage of vegetation, and is more suitable for vegetation monitoring and analysis. Additionally, the construction of the vegetation index considers the relationships between different bands, which can resist the interference of factors such as light and atmosphere on vegetation monitoring, thereby improving the accuracy and reliability of monitoring. The newly constructed vegetation index for rice leaf blight used in this study, designated as the rice blast index (RBI), was calculated as^[26].

$$\text{RBI} = a \times R_{\lambda_1} + b \times \frac{R_{\lambda_2} - R_{\lambda_3}}{R_{\lambda_2} + R_{\lambda_3}} \quad (3)$$

where, λ_1 is the band with the highest weight, and λ_2 and λ_3 are the bands with the highest weight in the normalized structure. The coefficients a and b are obtained by fitting the training set data with a linear discriminant analysis function.

Infection of rice with leaf blast^[27] will cause changes in the spectral reflectance of bands such as the visible light band and the near-infrared band. These changes can enhance the differences between the spectra of rice canopies at different disease severities by combining these bands, thereby achieving the detection and diagnosis of rice leaf blast^[23]. For this study, 31 established VIs (including RBI) were selected to detect the severity of rice leaf blight and compared with the classification effect of RBI (Table 3).

2.4 Classification algorithm

2.4.1 K-nearest neighbor (KNN)

The KNN algorithm is a supervised learning technique, considered one of the easiest to use in machine learning. It is characterized by simple operation and high classification accuracy. The KNN algorithm can be used for both classification and regression. The core concept of the KNN

Table 3 Established vegetation indices used in the study

Vegetation index	Definition
Normalized difference vegetation index (NDVI)	$(R800 - R670) / (R800 + R670)$
Triangular vegetation index (TVI)	$0.5 \times [120 \times (R750 - R550) - 200 \times (R670 - R550)]$
Photochemical reflectance index (PRI)	$(R570 - R531) / (R570 + R531)$
Modified chlorophyll absorption reflectance index (MCARI)	$[(R700 - R670) - 0.2 \times (R700 - R550)] \times R700 / R670$
Normalized difference 750/710 red edge NDVI (RENDVI)	$(R750 - R710) / (R750 + R710)$
Structural independent pigment index (SIPI)	$(R800 - R445) / (R800 + R680)$
Plant senescence reflectance index (PSRI)	$(R678 - R500) / R750$
Normalized pigment chlorophyll ratio index (NPCI)	$(R680 - R430) / (R680 + R430)$
Optimized soil adjusted vegetation index (OSAVI)	$1.16 \times (R800 - R670) / (R800 + R670 + 0.16)$
Aphid index (AI)	$(R740 - R887) / (R691 + R698)$
Healthy index (HI)	$(R534 - R698) / (R534 + R698) - 0.5 \times R704$
Ration triangular vegetation index (RTVI)	$[55 \times (R750 - R570) - 90 \times (R680 - R570)] / [90 \times (R750 + R570)]$
Renormalized difference vegetation index (RDVI)	$(R800 - R670) / (R800 + R670)$
Normalized difference 570/531 photochemical reflectance index 570/531 (PRI570/531)	$(R570 - R531) / (R570 + R531)$
Simple ratio 740/720	$R740 / R720$
Leaf chlorophyll index (LCI)	$(R850 - R710) / (R850 + R680)$
Green leaf index (GLI)	$(2 \times R500 - R700 - R450) / (2 \times R500 + R700 + R450)$
Anthocyanin reflectance index (ARI)	$1 / R550 - 1 / R700$
Blue-wide dynamic range vegetation index (BWDRVI)	$(0.1 \times R780 - R450) / (0.1 \times R780 + R450)$
Browning reflectance index (BRI)	$(1 / R550 - 1 / R700) / R780$
Chlorophyll index green (CIG)	$R780 / R550 - 1$
Difference 800/550 (D800/550)	$R800 - R550$
Difference 800/680 (D800/680)	$R800 - R680$
Double difference index (DDI)	$(R749 - R720) - (R701 - R672)$
Double peak index (DPI)	$(R688 + R710) / (R697)^2$
Green-blue NDVI (GBNDVI)	$[R780 - (R550 + R450)] / [R780 + (R550 + R450)]$
Green-red NDVI (GRNDVI)	$[R780 - (R550 + R653)] / [R780 + (R550 + R653)]$
Maccioni (Mac)	$(R780 - R710) / (R780 - R680)$
Modified triangular vegetation index 1 (MTVI1)	$1.2 \times [1.2 \times (R800 - R550) / 2.5 \times (R670 - R550)]$

algorithm is based on the proximity hypothesis. This hypothesis posits that categories similar to the sample tend to be in similar categories. The KNN algorithm measures the similarity between samples using the distance between sample data and classifies samples based on similarity. For a given data set, the KNN algorithm can predict the relationship between unseen data and existing data, and classify new data into the main category that best matches it.

2.4.2 Random forests (RF)

RF is an ensemble learning algorithm that was improved by

Breiman based on the bagging method. It performs classification and regression tasks by constructing multiple decision trees. RF has the advantages of handling high-dimensional data, parallel computing, strong noise resistance, and generalization ability, and has been widely applied in the field of remote sensing. RF constructs decision trees by randomly selecting feature subsets, focusing on local features and addressing the issues of high-dimensional data. Additionally, RF uses a parallel construction of decision trees, significantly enhancing data processing speed. For samples to be classified, RF determines the final classification result through a voting mechanism of multiple decision trees.

2.5 Evaluation metrics

To assess the accuracy of the vegetation index in distinguishing different disease severity levels, overall accuracy (OA) and Cohen's kappa coefficient were used as evaluation metrics. OA and the kappa coefficient can be obtained based on the confusion matrix. The confusion matrix is shown in Table 4, and the specific formulas for OA and the kappa coefficient are shown in Eqs. (4) and (5).

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where, TP (true positive) is the number of instances classified as positive that are positive, TN (true negative) is the number of instances classified as negative that are negative, FP (false positive) is the number of instances classified as positive but are negative, and FN (false negative) is the number of instances classified as negative but are positive.

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (5)$$

where, p_o is the sum of the number of correctly classified samples in each category divided by the total number of samples, which is also known as OA. p_e represents the expected accuracy of a classifier or evaluator under random conditions, where each category is classified according to its probability.

$$p_e = \sum_{i=1}^m \frac{(TP_i + FN_i) + (TP_i + FP_i)}{n^2} \quad (6)$$

where, n is the total number of samples.

3 Results

3.1 Average spectral reflectance of different disease severity levels

The average spectral reflectance of different disease severity levels can provide a general trend of these levels. In this study, the spectral data were processed to obtain the hyperspectral reflectance of rice in the 400–1000 nm range for different disease severity levels (Fig. 3). These observations revealed that the average spectral reflectance of different disease severity levels follows a similar trend. In the 400–550 nm range, the spectral reflectance of each level shows a gradual increase, forming a local peak at 550 nm. Following that, the spectral

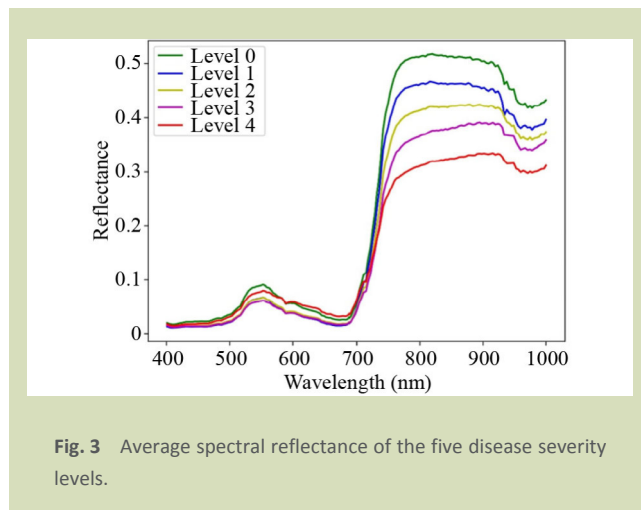


Fig. 3 Average spectral reflectance of the five disease severity levels.

reflectance of each level begins to decline. Between 680 and 770 nm, the spectral reflectance of each level abruptly rises, and between 770 and 900 nm, it tends to flatten. In the 900–1000 nm range, the reflectance curves of each level first slowly decrease, then gradually increase. In the 750–1000 nm range, the reflectance values show a gradient decline with the rise of disease severity levels.

3.2 Discriminability of different disease severity levels

3.2.1 Fatty acid metabolism

The discriminability of different bands was quantitatively evaluated using one-way analysis of variance (ANOVA). The computed variance F-values were normalized to demonstrate the discriminability effect of different disease severity levels across various bands, where 0 indicates the worst discriminability and 1 indicates the best discriminability. As evident in Fig. 4, the discriminability effect of a single band in the 400–700 nm range is not good. In the 740–880 nm range, the band discriminability is the best, whereas in the 880–1000 nm range, the band discriminability gradually decreases from high.

3.2.2 Band combination discriminability of different disease severity levels

The band intervals in hyperspectral data are relatively small, which often leads to information redundancy. To identify this redundant information, the correlation between each band was first calculated and the results are presented in the form of a heatmap (Fig. 5). It is clear that the correlation within the 400–720 nm and 720–1000 nm ranges exceeded 0.7, indicating a high degree of correlation and substantial redundancy within

Table 4 Confusion matrix

	Positive	Negative
Positive	True positive	False positive
Negative	False positive	True negative

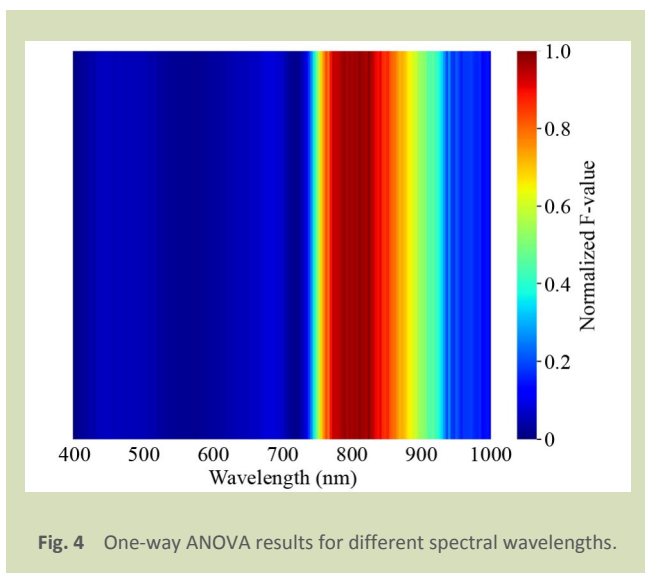


Fig. 4 One-way ANOVA results for different spectral wavelengths.

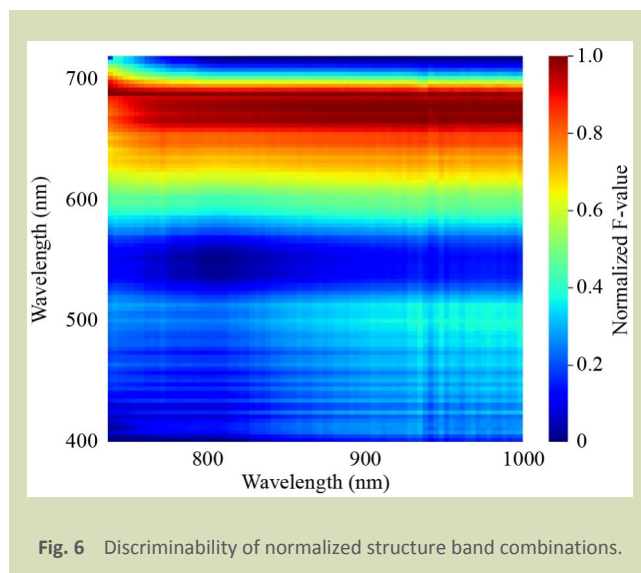


Fig. 6 Discriminability of normalized structure band combinations.

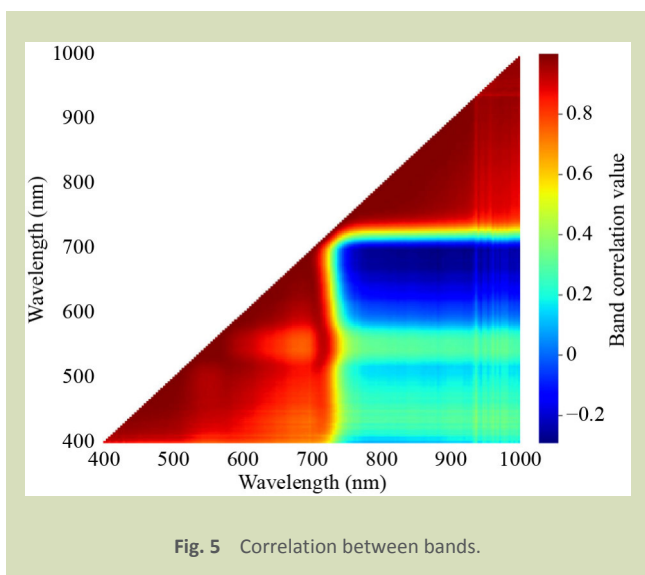


Fig. 5 Correlation between bands.

these two ranges. Therefore, one band from each of these two ranges was selected to represent the normalized vegetation structure parameters, to reduce the impact of information redundancy.

To select the best-normalized structure band combination, the study adopted an exhaustive method and performed normalization on the variance F-values. The 720–1000 nm range was used as the first band of the normalized structure, and the 400–720 nm range was used as the second band. Subsequently, one-way ANOVA was used to assess the discriminability of each band combination (Fig. 6). The results indicated that the normalized structure comprising the band combination of 720–1000 nm and 400–580 nm exhibited poor discriminability, with a value below 0.4. In contrast, the

normalized structure formed by the band combination of 720–1000 nm and 580–700 nm gradually demonstrated improved discriminability. Among all combinations, the normalized structure constructed from the band combination of 720–1000 nm and 650–680 nm showcased the best discriminability performance. Consequently, the normalized structure composed of the band combination of 720–1000 nm and 650–680 nm exhibits optimal discriminability.

3.3 Band selection results

The Relief-F algorithm was used to select single bands and band combinations (Fig. 7). The weight curve (Fig. 7(a)) rapidly rose around 400–700 nm, with a peak around 550 nm, and then gradually decreases to 700 nm; the curve sharply rose to the global peak from 700 to 780 nm and had a declining trend from 780 to 1000 nm. Based on the principle of maximum weight, the single band obtained was at 778 nm (Fig. 7(a)). Each point represents a band combination (Fig. 7(b)) and following the principle of maximum weight, the band combination obtained was 722 and 664 nm.

3.4 Comparison of constructed vegetation index and established vegetation index

Based on the Relief-F algorithm, this study has selected three sensitive bands: $\lambda_1 = 778$ nm, $\lambda_2 = 722$ nm, $\lambda_3 = 664$ nm, which are used to construct the RBI and compare it with the established vegetation index. To evaluate the correlation between the vegetation index and the severity level of rice blast disease, each vegetation index was analyzed with the disease severity level using Spearman correlation, with specific results

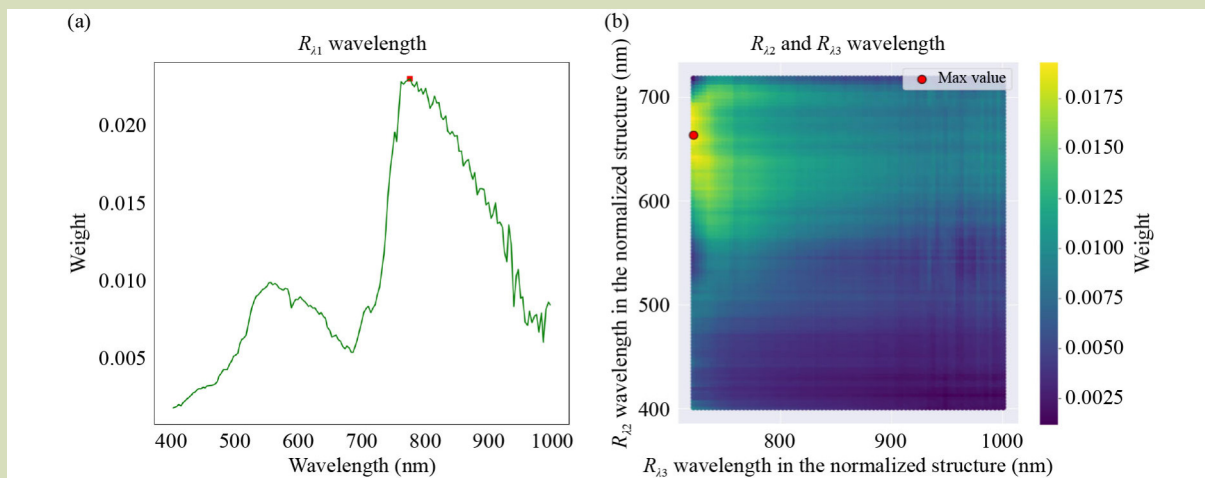


Fig. 7 Results and weights of band selection using Relief-F algorithm.

shown in Table 5.

According to the Spearman correlation analysis results, this study found that the established vegetation indices TVI, D800/500, D800/680, DDI, MTVI1 and the custom-built vegetation index RBI have a high correlation with the severity level of rice blast disease, with correlation coefficients all reaching 0.95 or above ($p = 0.01$).

To further visualize the classification effects of these six vegetation indices in differentiating disease severity levels, boxplots were used (Fig. 8), with RBI slightly overlapping between disease grades 0 (healthy) and 1 (mild disease), but no overlapping between the other disease grades, indicating its good discriminative ability across different disease severities. In contrast, TVI overlapped between different disease grades,

indicating poorer performance in disease detection and severity differentiation. D800/550, D800/680 and DDI indices demonstrate good discrimination between moderate to high disease grades (grades 3 and 4) but weaker discrimination between low disease grades (grades 0–2). MTVI1 exhibits some overlap between the lowest and lower disease grades (grades 0–3), while differentiation between other disease grades is clearer.

3.5 Classification performance

KNN and RF algorithms were used to establish disease severity detection models for rice blast disease using six vegetation indices, and the classification results are summarized in Table 6. Looking at the classification performance of the KNN model, RBI has the highest classification accuracy, with an OA

Table 5 Correlation between 31 vegetation indices (as given in) and disease severity level

Vegetation index	Correlation	Vegetation index	Correlation	Vegetation index	Correlation	Vegetation index	Correlation
RDVI	0.89	PSRI	-0.32	GRNDVI	-0.45	OSAVI	-0.90
PRI	0.64	Mac	-0.33	RTVI	-0.51	TVI	-0.97**
PRI570/531	0.64	LCI	-0.34	RENDVI	-0.52	D800/550	-0.96**
ARI	0.16	BWDVI	-0.40	NPCI	-0.52	D800/680	-0.97**
AI	-0.10	SIPI	-0.42	SR740/720	-0.52	DDI	-0.95**
DPI	-0.12	CIG	-0.43	GI	-0.55	MTVI1	-0.97**
GLI	-0.20	GBNDVI	-0.43	HI	-0.55	RBI	-0.98**
BRI	-0.20	NDVI	-0.45	MCARI	-0.68		

Note: ** $p = 0.01$.

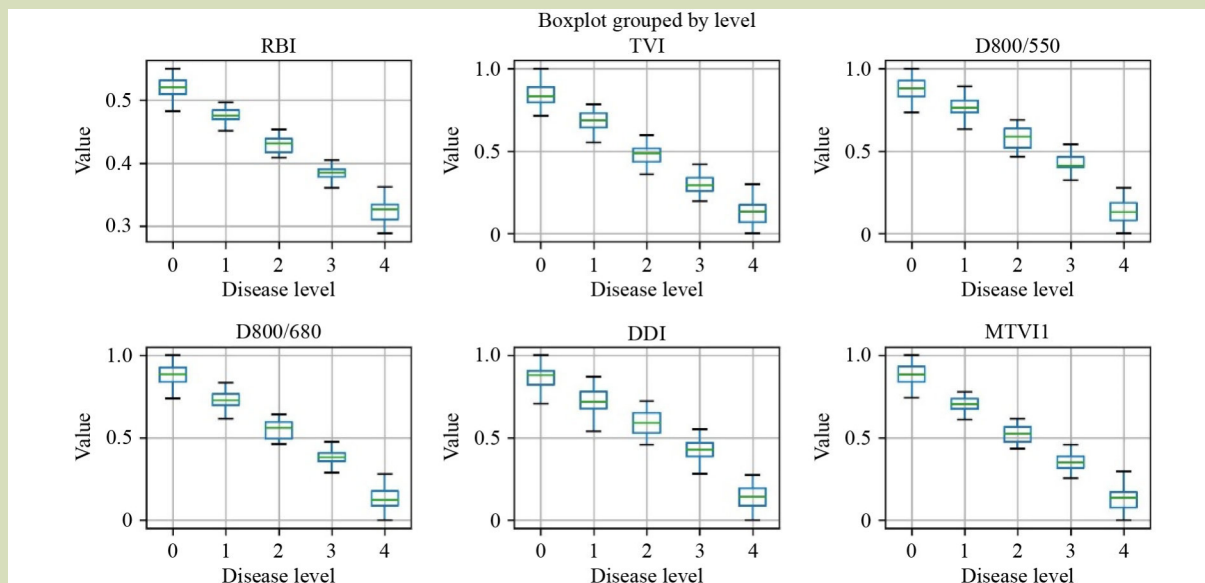


Fig. 8 Classification effects of different vegetation indices in distinguishing different disease severity levels.

Table 6 Classification results of vegetation indices

Model	Vegetation index	Overall accuracy (OA) (%)	Kappa coefficient (%)
K-nearest neighbor (KNN)	TVI	87.5	84.4
	D800/550	75.0	68.8
	D800/680	90.0	87.5
	DDI	67.5	59.4
	MTVI1	92.5	90.6
	RBI	95.0	93.8
Random forests (RF)	TVI	77.5	71.9
	D800/550	70.0	62.5
	D800/680	87.5	84.4
	DDI	67.5	62.5
	MTVI1	94.0	91.9
	RBI	95.1	92.5

of 95.0% and a kappa coefficient of 93.8%; followed by MTVI1, with an OA of 92.5% and a kappa coefficient of 90.6%. In terms of the RF model classification performance, RBI achieved the highest classification accuracy, with an OA reaching 95.1% and a kappa coefficient reaching 92.5%. MTVI1 classification accuracy is second to RBI, with an OA of 94.0% and a kappa coefficient of 91.9%. From the classification accuracy of KNN and RF, RBI provided the best classification performance, and the classification results are superior to the other five vegetation indices.

4 Discussion

To achieve efficient and non-destructive detection of rice blast disease, this study established the vegetation index, RBI, specifically for rice blast disease. Spectral data of rice blast canopy were collected using drone and qualitative analysis of spectral curve trends for different disease grades was conducted based on the average spectral reflectance curves of different disease grades. This qualitative analysis revealed specific patterns of spectral curve changes with grade variations.

Subsequently, the separability of spectral data under different disease grades was quantitatively evaluated using ANOVA, establishing the discriminative power of spectral data under different disease grades. Then, the Relief-F algorithm was used to obtain the weight curve of bands, and based on these weights, feature bands were selected and used to construct the RBI using Fisher discriminant analysis fitting coefficients. To evaluate the detection performance of RBI, 30 established vegetation indices were introduced for comparison. Spearman correlation analysis and visualization methods (boxplots) were used to further screen the vegetation indices and, finally, the classification accuracy of the vegetation indices was assessed through model classification accuracy.

Based on the average spectral reflectance curves of different levels, the average spectral curves of different disease severity levels all have a peak at 550 nm, and the curves of diseased rice are lower than those of healthy rice. This may be due to changes in the chlorophyll and carotenoid content in diseased rice^[28]. In the 680–740 nm range, the spectral curves of each disease severity level rose sharply, approximating a straight line. In the 740–770 nm range, the reflectance slope decreased with increasing disease severity, which may be related to the reduction of chlorophyll content caused by rice disease stress^[29]. In the 770–1000 nm range, the spectral reflectance curves decreased in a gradient with the disease severity level, which could be due to changes in the canopy structure after the rice is stressed by the blast disease^[30]. In the subsequent band separability, it was highest after 770 nm, which is consistent with the analysis of the average spectral curve in the near-infrared band.

In the quantitative understanding of band separability, according to the band separability results of one-way ANOVA, the separability of 740–880 nm was the best, which may be due to the relatively large impact on the cell structure of diseased rice in rice blast disease. In-band combination, considering band information redundancy, a correlation analysis was first performed between the bands, resulting in two band ranges of 400–720 and 720–1000 nm. Subsequently, the band combinations of these two ranges were exhaustively enumerated and the same one-way ANOVA was used for separability analysis with disease severity. Further, the feature bands 778, 722, and 664 nm were obtained using Relief-F. Then, 31 vegetation indices were correlated with disease severity, and the vegetation indices (a total of 6) with a correlation greater than 0.95 were selected: RBI, TVI, D800/550, D800/680, DDI, and MTVI1. The effect of the vegetation index was then visualized effectively using boxplots. Finally, KNN and RF models were established, and the

classification results showed that the classification accuracy of the RBI index was the highest, demonstrating the feasibility and effectiveness of the RBI application. According to the classification model results, D800/680, MTVI1, and RBI have better classification accuracy than the other three vegetation indices. The bands of D800/680 is a near-infrared and red bands, and these two band ranges are easily affected by rice blast disease. The lower classification accuracy of D800/550 than D800/680 might be due to the significant change at 680 nm affected by chlorophyll in this study, which is in line with the RBI proposed in this study. MTVI1 is an improvement of the TVI vegetation index, mainly to solve the problem that TVI is easily affected by soil background noise under low vegetation coverage. This makes MTVI1 still accurately reflect the chlorophyll content of vegetation under low vegetation coverage and the chlorophyll content changes under the stress of rice blast disease. This explains why MTVI1 can be used to detect rice blast disease and also why the classification accuracy of TVI is lower than MTVI1. The low classification accuracy of DDI may be because the band is concentrated at 670–750 nm, only considering the impact of rice blast disease on chlorophyll, and ignoring other influencing factors such as cell structure and vegetation morphology. The highest classification accuracy of the RBI proposed in this study might be due to the RBI index considering the impact of rice blast disease on rice chlorophyll content, cell structure and vegetation morphology.

In this paper, we introduce a new spectral index, the RBI, aimed at improving the accuracy and efficiency of rice blast disease detection. Compared to existing spectral indices, RBI possesses several significant advantages. Firstly, RBI is the result of precise calibration based on the specific effects of blast disease on rice leaves. It combines multiple bands sensitive to the disease, allowing for a more accurate reflection of the impact of blast disease on rice vegetation spectral reflectance. This meticulous band selection process ensures high sensitivity and specificity of RBI in disease detection. Secondly, RBI effectively reduces errors caused by atmospheric conditions and differences in crop growth stages by optimizing band combinations, thereby maintaining high stability and reliability under different environmental conditions. Additionally, the applicability of RBI is not limited to laboratory conditions; its effectiveness in large-scale field environments has been widely validated, demonstrating its broad applicability and superiority in practical applications. Through comparative analysis with other commonly used indices such as NDVI and RVI, we demonstrate the significant advantage of RBI in distinguishing healthy vegetation from those affected by rice blast disease, further demonstrating the application potential and innovative

value of RBI in crop disease monitoring.

The reasons for choosing KNN and RF classification models for this study mainly lie in three aspects. Firstly, KNN and RF are different types of models, with KNN based on instance-based learning and RF based on ensemble learning of decision trees. Choosing diverse models can enhance the overall prediction robustness and generalization ability. Secondly, KNN is highly sensitive to the local structure of the data, while RF is more sensitive to the global structure of the data. These two models can help us better adapt to different data distributions. Lastly, KNN and RF have demonstrated useful classification results in many fields, thus KNN and RF were chosen as the classification models for this study based on these considerations^[31–33].

In this study, there was no clear distinction between the superiority and inferiority of KNN and RF from the classification model results. In the KNN classification model, the classification accuracies of TVI, D800/550 and D800/680 were higher than in the RF model; in the RF classification model, the classification accuracies of DDI, MTVI1 and RBI were higher than in KNN. This demonstrates that the classification accuracy of a certain model is not universally higher than other models and, when establishing a model, its characteristics, the distribution of data, and the selection of model parameters need to be considered^[34,35].

5 Conclusions

When plants are stressed by disease, their biochemical parameters such as pigment content and canopy structure will change, thereby leading to changes in spectral reflectance. Therefore, studying the spectral reflectance characteristics of

vegetation under disease stress has become one of the important means of remote sensing disease detection. This study aimed to establish a vegetation index detection method for rice blast disease using unmanned aerial vehicle hyperspectral data to meet the requirements of precise and efficient modern agricultural detection.

In this study, we first performed a one-way ANOVA, finding that the region between 760 and 840 nm had the best separability among single-band reflections and the combination of 720–1000 and 650–680 nm had the best separability among band combinations. The Relief-F algorithm was then used to select the bands at 778, 722, and 664 nm based on maximum weights, which were used to construct the RBI for detecting rice blast disease. To verify and ensure the effectiveness of the vegetation index, we compared it with 30 established vegetation indices. We performed a Spearman correlation analysis for the 31 vegetation indices and disease severity, selecting six vegetation indices with correlations greater than 0.95, including RBI. These indices were then further examined using boxplots to visually compare their performance. According to the visualization results, RBI had minimal overlap between severity levels 0 and 1, with no overlaps between other severity levels. This indicates that although there is minor overlap, RBI still has high discriminatory power. Finally, we established KNN and RF classification models, finding that RBI had the highest classification accuracy of all vegetation indices evaluated. In the KNN model, the OA and kappa coefficient were 95.0% and 93.8%, respectively; in the RF model, the OA and kappa coefficient were 95.1% and 92.5%, respectively. In conclusion, the vegetation index constructed in this study can effectively detect rice blast disease, providing technical support and scientific methods for precise detection in modern agriculture.

Compliance with ethics guidelines

Guangming Li, Dongxue Zhao, Jinpeng Li, Shuai Feng, and Chunling Chen declare that they have no conflicts of interest or financial conflicts to disclose. This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Zhuang Y H, Ruan S H, Zhang L, Chen J R, Li S S, Wen W J, Liu H B. Effects and potential of optimized fertilization practices for rice production in China. *Agronomy for Sustainable Development*, 2022, 42(2): 32
2. Sen S, Chakraborty R, Kalita P. Rice—not just a staple food: a comprehensive review on its phytochemicals and therapeutic potential. *Trends in Food Science & Technology*, 2020, 97: 265–285
3. Younas M U, Wang G D, Du H B, Zhang Y, Ahmad I, Rajput N, Li M Y, Feng Z M, Hu K M, Khan N U, Xie W Y, Qasim M,

- Chen Z X, Zuo S M. Approaches to reduce rice blast disease using knowledge from host resistance and pathogen pathogenicity. *International Journal of Molecular Sciences*, 2023, **24**(5): 4985
4. Dean R, Van Kan J A L, Pretorius Z A, Hammond-Kosack K, Di Pietro A, Spanu P D, Rudd J J, Dickman M, Kahmann R, Ellis J, Foster G D. The top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 2012, **13**(4): 414–430
 5. Tao H L, Feng H K, Xu L J, Miao M K, Long H L, Yue J B, Li Z H, Yang G J, Yang X D, Fan L L. Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors*, 2020, **20**(5): 1296
 6. Agilandeswari L, Prabukumar M, Radhesyam V, Phaneendra K L N B, Farhan A. Crop classification for agricultural applications in hyperspectral remote sensing images. *Applied Sciences*, 2022, **12**(3): 1670
 7. Marang I J, Filippi P, Weaver T B, Evans B J, Whelan B M, Bishop T F A, Murad M O F, Al-Shammari D, Roth G. Machine learning optimised hyperspectral remote sensing retrieves cotton nitrogen status. *Remote Sensing*, 2021, **13**(8): 1428
 8. Deng X L, Zhu Z H, Yang J C, Zheng Z, Huang Z X, Yin X B, Wei S J, Lan Y B. Detection of Citrus Huanglongbing based on multi-input neural network model of UAV hyperspectral remote sensing. *Remote Sensing*, 2020, **12**(17): 2678
 9. Zhang N, Yang G J, Pan Y C, Yang X D, Chen L P, Zhao C J. A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sensing*, 2020, **12**(19): 3188
 10. Gao Z M, Khot L R, Naidu R A, Zhang Q. Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Computers and Electronics in Agriculture*, 2020, **179**: 105807
 11. Li N W, Huo L, Zhang X L. Classification of pine wilt disease at different infection stages by diagnostic hyperspectral bands. *Ecological Indicators*, 2022, **142**: 109198
 12. Zhao X H, Zhang J C, Huang Y B, Tian Y Y, Yuan L. Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis. *Computers and Electronics in Agriculture*, 2022, **193**: 106717
 13. Randive P U, Deshmukh R R, Janse P V, Gupta R S. Discrimination Between Healthy and Diseased Cotton Plant by Using Hyperspectral Reflectance Data. In: *Recent Trends in Image Processing and Pattern Recognition*. Singapore: Springer, 2019, 342–351
 14. Shukla K S, Birah A, Nigam R, Kanojia A K, Km Khokhar M, Bhattacharya B K, Chander S. Selection of sensitive bands for assessing Alternaria blight diseased severity grades in mustard crops using hyperspectral reflectance. *Journal of Agrometeorology*, 2023, **25**(2): 274–279
 15. Zhang N N, Zhang X, Shang P, Ma R, Yuan X T, Li L, Bai T C. Detection of cotton Verticillium wilt disease severity based on hyperspectrum and GWO-SVM. *Remote Sensing*, 2023, **15**(13): 3373
 16. Khanal S, Kc K, Fulton J P, Shearer S, Ozkan E. Remote sensing in agriculture—Accomplishments, limitations, and opportunities. *Remote Sensing*, 2020, **12**(22): 3783
 17. Raj R, Kar S, Nandan R, Jagarlapudi A. Precision agriculture and unmanned aerial vehicles (UAVs). In: Avtar R, Watanabe T, eds. *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Springer, 2020, 7–23
 18. Ma H Q, Huang W J, Dong Y Y, Liu L Y, Guo A T. Using UAV-based hyperspectral imagery to detect winter wheat fusarium head blight. *Remote Sensing*, 2021, **13**(15): 3024
 19. Moriya É A S, Imai N N, Tommaselli A M G, Berveglieri A, Santos G H, Soares M A, Marino M, Reis T T. Detection and mapping of trees infected with citrus gummosis using UAV hyperspectral data. *Computers and Electronics in Agriculture*, 2021, **188**: 106298
 20. Zheng Q, Huang W J, Ye H C, Dong Y Y, Shi Y, Chen S S. Using continuous wavelet analysis for monitoring wheat yellow rust in different infestation stages based on unmanned aerial vehicle hyperspectral images. *Applied Optics*, 2020, **59**(26): 8003–8013
 21. Abdulridha J, Min A, Rouse M N, Kianian S, Isler V, Yang C. Evaluation of stem rust disease in wheat fields by drone hyperspectral imaging. *Sensors*, 2023, **23**(8): 4154
 22. Liu T, Shi T Z, Zhang H, Wu C. Detection of rise damage by leaf folder (*Cnaphalocrocis medinalis*) using unmanned aerial vehicle based hyperspectral data. *Sustainability*, 2020, **12**(22): 9343
 23. Ashourloo D, Mobasheri M, Huete A. Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements. *Remote Sensing*, 2014, **6**(6): 5107–5123
 24. Heim R H J, Wright I J, Allen A P, Geedicke I, Oldeland J. Developing a spectral disease index for myrtle rust (*Austropuccinia psidii*). *Plant Pathology*, 2019, **68**(4): 738–745
 25. Kira K, Rendell L A. The Feature Selection Problem: Traditional Methods and a New Algorithm. In: *Proceedings of the Tenth National Conference on Artificial Intelligence—AAAI'92*. San Jose, California: AAAI Press, 1992, 129–134
 26. Huang W J, Guan Q S, Luo J H, Zhang J C, Zhao J J, Liang D, Huang L S, Zhang D Y. New optimized spectral indices for identifying and monitoring winter wheat diseases. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, **7**(6): 2516–2524
 27. Mandal N, Das D K, Sahoo R N, Adak S, Kumar A, Viswanathan C, Mukherjee J, Rajashekara H, Ranjan R, Das B. Assessing rice blast disease severity through hyperspectral remote sensing. *Journal of Agrometeorology*, 2022, **24**(3): 241–248
 28. Zheng Q, Huang W J, Xia Q, Dong Y Y, Ye H C, Jiang H, Chen S S, Huang S Y. Remote sensing monitoring of rice diseases and pests from different data sources: a review. *Agronomy*, 2023, **13**(7): 1851

29. Wan L, Li H, Li C S, Wang A C, Yang Y H, Wang P. Hyperspectral sensing of plant diseases: principle and methods. *Agronomy*, 2022, **12**(6): 1451
30. Zheng Q, Huang W J, Cui X M, Shi Y, Liu L Y. New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery. *Sensors*, 2018, **18**(3): 868
31. Kuncheva L I, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 2003, **51**(2): 181–207
32. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 1967, **13**(1): 21–27
33. Caruana R, Niculescu-Mizil A. An Empirical Comparison of Supervised Learning Algorithms. In: Proceedings of the 23rd International Conference on Machine Learning—ICML'06. Pittsburgh, Pennsylvania: *ACM Press*, 2006, 161–168
34. Altman N S. An Introduction to Kernel and nearest-neighbor nonparametric regression. *American Statistician*, 1992, **46**(3): 175–185
35. Biau G, Scornet E. A random forest guided tour. *Test*, 2016, **25**(2): 197–227