Multi-scale monitoring for hazard level classification of brown planthopper damage in rice using hyperspectral technique

Juan Liao[1,2,3,4], Wanyan Tao[1,2], Yexiong Liang[1,2], Xinying He[1,2], Hui Wang[1,2], Haoqiu Zeng[1,2], Zaiman Wang[1,2,3,4,5], Xiwen Luo[1,2,3,4,5], Jun Sun[6], Pei Wang[1,2,3,4], Ying Zang[1,2,3,4,5*]

(1. *Guangdong Laboratory for Lingnan Modern Agriculture, College of Engineering, South China Agricultural University, Guangzhou 510642, China*;
2. *Key Laboratory of Key Technology on Agricultural Machine and Equipment (South China Agricultural University), Ministry of Education, Guangzhou 510642, China*;
3. *Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence (GDKL-AAI), Guangzhou 510642, China*;
4. *State Key Laboratory of Agricultural Equipment Technology*;
5. *Huangpu Innovation Research Institute of SCAU, Guangzhou 510715, China*;
6. *School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013, China*)

Abstract: The primary aim of this study was to classify the hazard level of brown planthopper (BPH) damage in rice. Three datasets, including spectral reflectance corresponding to the sensitive wavelengths from rice canopy spectral wavelengths, rice stem spectral wavelengths, and fusion information of rice canopy and stem spectral wavelengths were used for BPH hazard level classification by using different algorithms. Datasets and algorithms were optimized by the BPH hazard level classification effects (which was evaluated by indices of accuracy, precision, recall, F_1, and *k*-value). The optimized algorithm combination was used to build a hazard level classification model for spectral reflectance corresponding to the sensitive wavelength from the rice canopy spectral images. Results showed that: (1) The spectral reflectance corresponding to the sensitive wavelengths of fusion information dataset performed best in BPH hazard level classification, with the highest accuracy (99.08%), precision (99.31%), recall (98.83%), F_1 (0.99), and *k*-value (0.99). (2) The optimum algorithm combination was Savitzky-Golay (S-G) smoothing, principal component analysis (PCA) for sensitive wavelength selection, and broad-learning system (BLS) for modeling. (3) The spectral reflectance corresponding to the sensitive wavelengths dataset of rice canopy spectral images achieved accuracy (80.63%), precision (80.28%), recall (77.03%), F_1 (0.79), and *k*-value (0.74) in classifying BPH hazard level by using the optimum algorithm combination.

Keywords: brown planthopper (BPH), hazard level classification, hyperspectral technique, rice canopy, rice stem, fusion information

DOI: 10.25165/j.ijabe.20241706.9199

Citation: Liao J, Tao W Y, Liang Y X, He X Y, Wang H, Zeng H Q, et al. Multi-scale monitoring for hazard level classification of brown planthopper damage in rice using hyperspectral technique. Int J Agric & Biol Eng, 2024; 17(6): 202–211.

1 Introduction

Rice is the main staple food crop for more than half of the world's population. In China alone, around 30 million hm^2 are under rice cultivation, yielding an annual production of 208 Mt. Among the various pests affecting rice, the brown planthopper (BPH) is one of the most destructive, particularly in Asia[1]. The BPH infestation of rice annually results in about 20% of yield loss and significant economic damage across Asia[2]. Therefore, it is essential to monitor the incidence of BPH in the crop in a timely manner, and adopt appropriate methods to control and prevent the yield losses in rice.

However, the monitoring of BPH remains one of the world's most challenging technical tasks. BPH can significantly alter the chemical and physical properties of rice plants, including changes in chlorophyll characteristics, chemical concentrations, cell structure, nutrient and water uptake, and gas exchange. These alterations consequently lead to variations in spectral reflectance of the foliage. The changes in plant spectra can be distinguished by the reflectance "spectral signatures" of their surfaces produced over different wavebands of the electromagnetic spectrum[3].

The use of non-destructive methods for reflectance spectra detection of plants is widely adopted for monitoring vegetation conditions. This is primarily due to the advancement of hyperspectral remote sensing equipment[4]. Such technology offers additional bands within the visible, near-infrared (NIR), and

Received date: 2024-07-03 **Accepted date:** 2024-11-30

Biographies: Juan Liao, PhD, Assistant Research Fellow, research interest: agricultural information perception, Email: liaojuan0529@126.com; **Wanyan Tao**, PhD candidate, research interest: agricultural information perception, Email: 3466409273@qq.com; **Yexiong Liang**, MS Candidate, research interest: agricultural information perception, Email: lyxpululu@163.com; **Xinying He**, MS candidate, research interest: agricultural information perception, Email: 1041310305@qq.com; **Hui Wang**, MS candidate, research interest: agricultural information perception, Email: snowboarding7@163.com; **Haoqiu Zeng**, MS candidate, research interests: agricultural information perception, Email: 18075798632@163.com; **Zaiman Wang**, PhD, Professor, research interest: precision agriculture, Email: wangzaiman@scau.edu.cn; **Xiwen Luo**, Professor, research interest: precision agriculture, Email: xwluo@scau.edu.cn; **Jun Sun**, PhD, Professor, research interest: agricultural information perception, Email: sun2000jun@sina.com; **Pei Wang**, PhD, Associate Professor, research interest: agricultural information perception, Email:164930053@qq.com.

***Corresponding author: Ying Zang**, PhD, Professor, research interest: precision agriculture. College of Engineering, South China Agricultural University, Guangzhou 510642, China. Tel: +86-13724090346, Email: yingzang@scau.edu.cn.

shortwave-infrared (SWIR) regions, enhancing the ability to detect and analyze the biotic stresses in vegetation. Therefore, the use of hyperspectral remote sensing to detect the affected rice plants at an early stage of potential increase in BPH populations is promising for the development of better pest management tactics[5,6]. The reflectance pattern of plant foliage is determined by the chemical composition and physical properties of the plant tissues.

Prasannakumar et al.[7,8] identified four sensitive wavelengths, at 1986, 665, 1792, and 500 nm, in relation to BPH stress on rice plants. They developed and validated a multiple-linear regression model (R^2=0.71, RMSE=1.74, p<0.0001 in training set and R^2=0.73, RMSE=0.71, p<0.0001 in validation) that would help to monitor BPH stress on rice. Similarly, Huang et al.[9] investigated the sensitive bands affected by BPH damage across different growth stages of rice, utilizing SPAD and spectral data. Results showed that the spectral reflectance of the near-infrared wavelength range of the rice canopy significantly decreased with increasing BPH infestation. Furthermore, the modified chlorophyll absorption ratio index (MCARI710) proved to be related to the number of BPH under different nitrogen fertilizer rates and durations of BPH infestation. Tan et al.[10] established a stepwise regression model linking the spectral indices to the amount of BPH, and found that the ratio vegetation index $RVI_{746/670}$ was sensitive to BPH damage, suggesting a potential advance for monitoring BPH.

These studies demonstrated a significant correlation between the spectral index and BPH infestation, with most research to date focusing on the rice canopy spectra for BPH monitoring. However, the detection of stem spectra for BPH infestation monitoring remains unexplored. BPH generally feeds by inserting its stylet into the vascular tissue of a rice stem to extract the rice sap, which results in changes in leaf chlorophyll and relative water content of rice[11,12]. Hypothetically, alterations of the rice stem spectra would occur before changes in the rice canopy spectra when a rice plant is damaged by BPH. Hence, the detection of changes in the rice stem spectra could be crucial for the early monitoring of BPH.

Building an effective automatic system is essential for monitoring BPH. Advances in machine learning and deep learning have significantly enhanced crop pest monitoring capabilities[13,14]. These techniques have achieved breakthroughs in many applications[15,16]. However, despite the powerful capabilities, most deep learning methods suffer from the time-consuming training process, due to numerous hyperparameters and complex structures. To tackle this challenge, the Broad Learning System (BLS)[17] has been established in the form of a flat network, where the original inputs are transferred and placed as “mapped features” in the feature nodes, and the structure is expanded in a broad sense in the “enhancement nodes”. The incremental learning algorithms are developed for fast remodeling in broad expansion without a retraining process if the model deems to be expanded. It is especially suitable for systems with a small data scale and few features, but high requirements for real-time prediction, such as crop pests monitoring[18].

This study aimed to classify the hazard level of BPH damage in rice. The specific objectives included: (1) Detecting the spectral reflectance of spectral wavelengths of the rice canopy and stem using a FieldSpec3 ASD, and collecting the spectral images of the rice canopy using a Nano-Hyperspec® VNIR. (2) Acquiring datasets of spectral wavelengths that correlate with the spectral wavelengths of the rice canopy, rice stem, the fusion information from both, as well as the spectral images of rice. (3) Developing models to classify BPH hazard level based on the spectral reflectance corresponding to sensitive wavelengths datasets, and optimizing both the datasets and algorithms for effective BPH hazard level classification.

2 Materials and methods

2.1 Materials and devices

The experiments were conducted at the research test base of South China Agricultural University in Guangzhou, China (31.24°N, 121.29°E). Rice seedlings were planted in seven cement pit test plots, each measuring 7 m in length and 2 m in width. These cement pits were enclosed with mesh to protect the rice samples from external pest invasions and to contain any insects released during study. Prior to the experiments, preparations were made, including soil preparation, weeding, pest control, etc.

BPHs were provided by the Plant Protection Research Institute at the Guangdong Academy of Agricultural Science. Additionally, rice plants of the Hybrid rice variety No. 5 were provided by the College of Agriculture at South China Agricultural University (SCAU).

The spectral reflectance of rice plant canopies and stems infested with different BPH densities (and thus potential damage levels) was measured using a field-portable spectroradiometer (FieldSpec3, Analytical Spectral Devices® [ASD]) at 1 nm intervals with a waveband from 350 to 2500 nm. The instrument had a facility to communicate through wireless access with a laptop computer. The laptop was used to record and process the spectral data with the ASD software. The main specifications of the ASD FieldSpec3 are listed in Table 1.

Table 1 Main specifications of FieldSpec3, Analytical Spectral Devices® [ASD]

Specifications	Parameters
Waveband	350-2500 nm
Sampling interval	1.4 nm (350-1000 nm) 2 nm (1001-2500 nm)
Spectral resolution	3 nm@700 nm 10 nm@1400, 2100 nm
Number of detector array channels	512@350-1000 nm, 520*2@1001-1800 nm, 520*2@1801-2500 nm

The spectral images of the rice canopy were obtained using the Nano-Hyperspec® VNIR, a push broom imaging spectrometer manufactured by Headwall in the United states. The main specifications of Nano-Hyperspec® VNIR are listed in Table 2.

Table 2 Specifications of Nano-Hyperspec® VNIR

Specifications	Parameters
Waveband	400-1000 nm
Sampling interval	1.74 nm/pixel
Spectral resolution	6 nm@700 nm
Number of detector array channels	340

In this study, Intel(R) Core™ i7-10750H CPU @ 2.30GHz and NVIDIA GeForce 2080Ti GPU were employed for network modeling.

2.2 Spectral data measurements

2.2.1 Rice spectral reflectance measurements

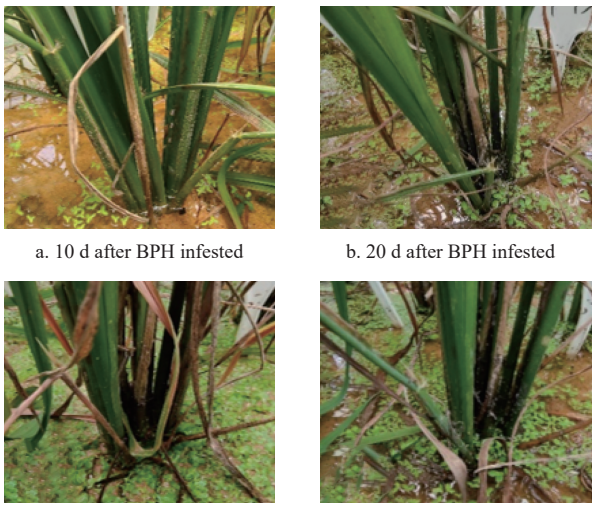
Four BPH hazard level classification strategies were tested in an experiment conducted from 2021 to 2023. Rice seedlings were transferred to test plots, and then inoculated with different BPH densities, corresponding to potential damage levels based on the BPH Hazard Level (2009) classification standard, during the tillering stage as listed in Table 3.

Table 3 Classification standard of brown planthopper hazard level (2009 Standards Press of China, National Standards Survey and Reporting Standards of the People's Republic of China)

	Level 0	Level 1	Level 2	Level 3
Hazard Levels	No damage	Light damage (Leaves partially yellowing but with no hopper burn)	Moderate damage (Leaves with pronounced yellowing and some wilting)	Serious damage (Most of the leaves wilting with hopper burn; all plants dead)
Density of BPH on one hundred rice plants	0 (uninfected)	<250	700-1200	>1600

Four treatments with different BPH densities were arranged, including a control treatment with hazard level 0 (marked as CK), a test treatment with BPH hazard level 1 (marked as T_1), a test treatment with BPH hazard level 2 (marked as T_2), and a test treatment with BPH hazard level 3 (marked as T_3). Each test plot used a block design comprising four treatments with different BPH densities, with each treatment replicated three times. The four treatments with different BPH densities were arranged in a cement pit to avoid the influence of water and nutrition difference on the rice spectra. In each treatment area, 25 rice seedlings were planted in a 5×5 grid, resulting in 75 rice seedlings per BPH hazard level. Rice spectral data were continuously monitored for 35 d after BPH inoculation.

Prior to measurement, the instrument was calibrated for solar radiation using a Spectralon® reference panel. The spectroradiometer was configured to provide an average of 50 spectral reflectance data for each target at a time. To ensure comprehensive coverage, each rice plant designated for testing was measured three times with a 25° field of view, and the sensor was kept at 80 cm height above the rice plant. The mean spectral reflectance from these measurements was calculated and recorded as the spectral reflectance for each rice plant. As the spectral reflectance data measurement is highly sensitive to solar radiation, the experiments were conducted on sunny days from 10:00 AM to 2:00 PM to ensure optimal lighting conditions. Additionally, to reduce the influence of BPH migration, reproduction, and mortality, the number of BPH on each rice seedling was recorded. Rice seedlings with abnormal BPH densities were excluded from the data records to maintain the integrity and accuracy of the results (Figure 1).



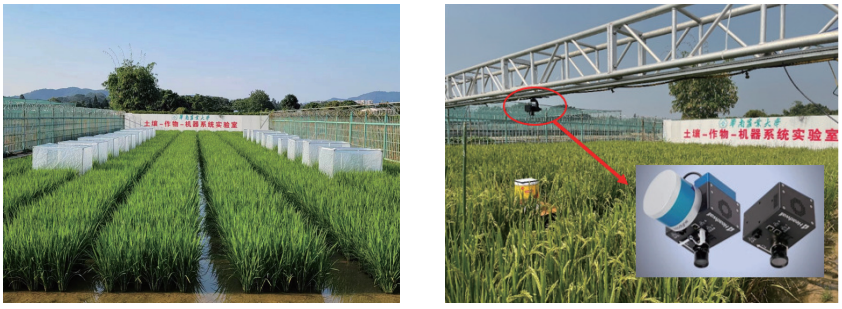
a. 10 d after BPH infested b. 20 d after BPH infested

c. 30 d after BPH infested d. 35 d after BPH infested

Figure 1 Test plot with BPH hazard level 3

2.2.2 Rice spectral images acquisition

Four BPH hazard levels with different BPH densities were established for the experiment. Each hazard level plot was enclosed with mesh, measuring 1 m in width and 1 m in length, and planted with rice plants in an arrangement of 6 rows by 6 columns. Each hazard level plot consisted of three replications (Figure 2).



a. Test plots including BPH hazard level 0, level 1, level 2 and level 3

b. Nano-Hyperspec® VNIR spectral imaging system

Figure 2 Spectral images acquisition

2.3 Spectral data processing

2.3.1 Spectral reflectance extraction

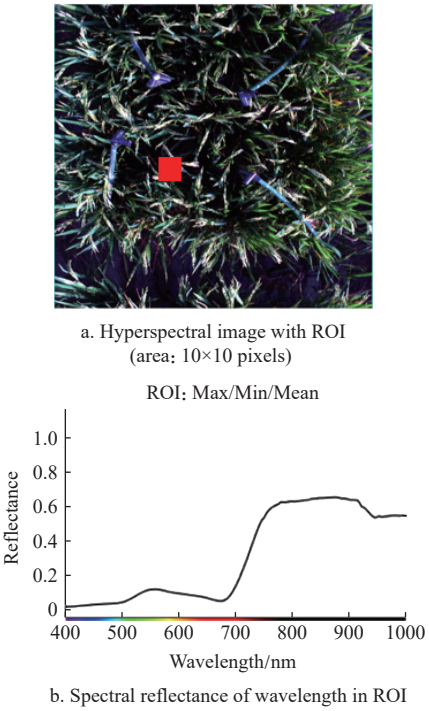
For the analysis of hyperspectral images, the spectral reflectance of wavelength should be extracted from the hyperspectral images to facilitate sensitive wavelength selection and modeling. The Environment for Visualizing Images (ENVI) was employed for Region of Interest (ROI) extraction, and the spectral reflectance of wavelength was extracted in each ROI (Figure 3).

2.3.2 Savitzky-Golay (S-G) smoothing

Due to the high resolution of the hyperspectral instruments, the large amount of redundant information in raw spectral data aggravates the problems related to the curse of dimensionality[19]. The information contained in each set of the correlation matrix principal diagonal (over the threshold) is contained in one of those bands. In this study, in order to select the most discriminative spectral band between classes, Savitzky-Golay (S-G)[20] was used to reduce dimensionality while preserving relevant information for posterior feature selection. S-G smoothing applies a polynomial to fit the wavelength data, thereby eliminating noise and preserving the overall trend and characteristics of the spectral signal. This method is a weighted mean algorithm, which is realized by Equation (1)[21].

$$X_{k,\text{smooth}} = \bar{X}_k = \frac{1}{H}\sum_{i=-w}^{+w} X_{k+1} h_i, \quad (1)$$

where, $X_{k,\text{smooth}}$ is the spectra data of k point after smoothing; H is a normalizing factor; and h_i is a smoothing factor.



a. Hyperspectral image with ROI (area: 10×10 pixels)

b. Spectral reflectance of wavelength in ROI

Figure 3 ROI extraction for hyperspectral images

2.3.3 Principal component analysis (PCA) sensitive wavelength selection

After eliminating the redundant information, a set of self-contained information bands were obtained. The next crucial step involves identifying those bands that most effectively discriminate between different BPH hazard levels for classification purposes. Principal component analysis (PCA) is a fundamental method for analyzing multivariate signals, including spectral reflectance, hyperspectral images, and other high-dimensional datasets which contain hidden information[22]. PCA uses principal components for linear composition of spectral variables[23]. Each principal component contains distinct information of the spectrum, ensuring that the principal components are mutually independent and with no overlapping information. This approach effectively resolves the collinearity problem among spectral variables.

In the context of PCA, given m number of spectral variables that have been normalized, such as x_1, x_2, … x_m, each variable is represented by a linear combination of n (where $n<m$) factors including f_1, f_2, … f_n. Then the mathematical model of principal component analysis can be expressed as[24]:

$$x_1 = a_{11}f_1 + a_{12}f_2 + \ldots a_{1n}f_n + \varepsilon_1 \quad (2)$$

$$x_1 = a_{21}f_1 + a_{22}f_2 + \ldots a_{2n}f_n + \varepsilon_2 \quad (3)$$

$$\ldots\ldots \quad (4)$$

$$x_m = a_{m1}f_1 + a_{m2}f_2 + \ldots a_{mn}f_n + \varepsilon_m \quad (5)$$

where, F is factor, and A is factor loading matrix. The PCA approach aims to extract the most important information from the original data by keeping the optimal number of Principal Components (PCs). The number of retained PCs can be determined according to the characteristic value, cumulative contribution rate, and actual needs of each PC. In actual calculation, PCs with a cumulative contribution rate greater than 85% (80%, 90%, or any other predefined value) of the total data variance are usually used for practical analysis to represent the original spectral variables for modeling calculations[25]. In this study, 85% was selected as the number of PCs that explains least of the total data variance.

2.4 Information fusion

In this study, the spectral wavelength datasets for both the rice stem and canopy were fused by concatenating with each other to create a comprehensive dataset. This fused dataset, represented in a matrix format, was subsequently utilized as the basis for selecting sensitive wavelengths and developing models aimed at classifying the BPH hazard levels.

2.5 Modeling

2.5.1 Broad Learning System (BLS) algorithm model

BLS (Broad Learning System) algorithm was employed to build models for BPH hazard level classification. BLS is a machine learning algorithm proposed by Chen et al. [17] The essence of BLS is a kind of random vector functional link neutral network (RVFLNN). Compared with deep learning methods, BLS significantly reduces the training time. The algorithm structure of BLS is shown in Figure 4.

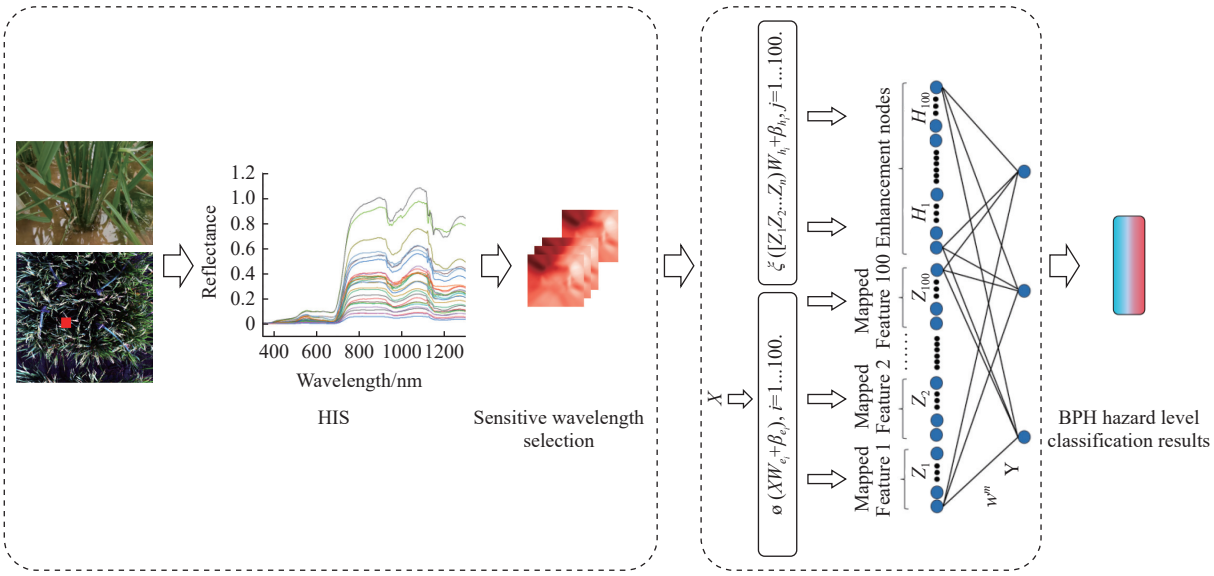


Figure 4 An illustration of the BLS algorithm

The processing steps of BLS are as follows:

(1) Input Layer. The input layer of BLS typically consists of raw data. These data are sent into the mapping nodes of BLS. In this study, the BLS input layer consists of spectral reflectance of sensitive wavelengths datasets.

(2) Mapping Nodes. The mapping nodes are the first hidden layer in BLS, composed of numerous nodes. Each mapping node contains a simple linear function that performs a linear transformation on the input data using weights and biases, and passes the result as input to the next layer of enhance nodes.

(3) Enhance Nodes. The enhance nodes are the second hidden layer in BLS, also composed of numerous nodes. Each enhance node contains a non-linear function that performs a non-linear transformation on the output of the mapping nodes, and passes the result to the next layer of the output layer.

(4) Output Layer. The output layer consists of a group of nodes, with each node corresponding to an output of the BLS model, such as classification, regression, etc. The computation in the output layer follows the same way as traditional neural networks. In this study, the output layer consists of BPH hazard level classification results.

The core of BLS is to calculate the pseudo-inverse of the enhance nodes to the target value. The result of the computation corresponds to the weight of traditional neural network. The process of BLS mainly contains three steps: (1) Generating the Mapping Nodes, the input spectra are mapped to create a feature node matrix. (2) The enhanced node matrix is formed by the enhanced transformation of the feature node matrix, and the pseudo-inverse matrices are calculated. (3) Served the feature mapping node and the enhancement node together as the input of the hidden layer. (4) The pseudo-inverse is used to solve the weight matrix between the hidden layer and the output layer.

Suppose the input data: $X=[x_1^{\mathrm{T}}, x_2^{\mathrm{T}}, ..., x_n^{\mathrm{T}}]^{\mathrm{T}} \epsilon R^{N*M}$, the output matrix $Y=[y_1^{\mathrm{T}}, y_2^{\mathrm{T}}, ..., y_n^{\mathrm{T}}]^{\mathrm{T}} \epsilon R^{N*C}$, $x_i=(x_{i1}, x_{i2}, ..., x_{im}) \epsilon R^M$, $y_i=[y_{i1}, y_{i2}, ..., y_{ic}] \epsilon R^c$. Where, n is the total number of data samples; M is the sample dimension, and C is the number of categories of data samples; $Z_1, Z_2, \ldots, Z_n$ are feature nodes; $H_1, H_2, \ldots, H_m$ are enhance nodes, which are calculated by feature nodes; and W is output weight matrix, which is calculated by the ridge regression algorithm. In addition, due to the network structure of BLS, when adding new enhanced nodes, new weights can be obtained quickly by solving the pseudo-inverse of the block matrix, which makes BLS greatly efficient in incremental learning.

Compared with the traditional deep learning method, BLS has a simpler structure and higher computation velocity. This also improves the robustness of the classification model.

In this study, twenty-six spectral sensitive wavelengths were selected from canopy spectral wavelengths, and fourteen spectral sensitive wavelengths were selected from stem spectral wavelengths. These spectral sensitive wavelengths were unified dimension and concatenated with tabs into a 41-dimensional dataset as input $\boldsymbol{H}_1$, and $\boldsymbol{H}_1$ was incremented (Table 4). Then 100 mapping nodes were generated for each feature window. The details are described as follows: (1) Firstly, a random weight matrix with a size of 40×100 was generated and denoted as ***we***, and then a new matrix A_1 was calculated by $A_1=H_1 \times we$, after A_1 was normalized, mapping matrix $\boldsymbol{W}$ from $\boldsymbol{H}_1$ to $\boldsymbol{A}_1$ was calculated by the equation $W=argmin\left||ZW-H_1||_2^2+\lambda\right||W||_1$. Then the mapping node of this feature map was calculated by the equation $T_1=normal(H_1 \times W)$, and finally, all mapping nodes were obtained by this equation. (2) Secondly, the enhancement node for each feature window was obtained. The details are as follows: a) Random size matrix with a size of 16×100×100 was orthogonally normalized and denoted as ***wh***. b) After the feature nodes matrices were concatenated, they were normalized and incremented to be a new matrix which was denoted as $\boldsymbol{H}_2$; $T_2=tansig\left(\frac{H_2 \times wh \times s}{\max(H_2 \times wh)}\right)$, where ***s*** is scaling of enhancement node. Then the $T=\begin{bmatrix} y \\ T_2 \end{bmatrix}$ was calculated and was used in network. (3) Finally, the pseudo-inverse of $\boldsymbol{T}$ was obtained in order to obtain the final weight.

Table 4 Parameters for BLS modeling, including the main parameters of input dimension, mapping node, enhancement node, number of layers, feature window size, and enhanced node scaling

Hyper-parameters	Network structure parameters				Variable parameters	
	Input dimension	Mapping node	Enhanced serve node	Number of layers	Feature window size	Enhanced node scaling
Numbers	41	100	100	2	16	1

Scikit-learn was employed for data processing. The proportion of training sets to test sets was 4:1, and the network model proposed in this study was tested by training set. Confusion matrix was plotted, and the indices of accuracy, precision, recall, *k*-value, and F_1-value were calculated for model effects evaluation[26]. Each test was repeated ten times, and the arithmetic mean values were served as results.

2.5.2 Confusion matrix for models effects evaluation

The issue of learning the reliability of predictors is of great importance in multiclass classification[27]. The effectiveness of the BPH hazard classification model was evaluated by confusion matrix[28], kappa coefficient (*k*-value), accuracy, precision, recall, and F_1-score[29]. The confusion matrix is also known as the error matrix, and it is widely used in determination of the behavior of accuracy evaluation in multiclass classification models[30]. In the confusion matrix, four measures, namely, "true positive" (TP), "true negative" (TN), "false positive" (FP), and "false negative" (FN), respectively, have been reported as very important indices for model accuracy analysis[31], as listed in Table 5.

Table 5 Specifications of confusion matrix

Confusion matrix		True value	
		Positive	Negative
Predicted value	Positive	TP	FP
	Negative	FN	TN

Note: Four measures, namely, "true positive" (TP), "true negative" (TN), "false positive" (FP), and "false negative" (FN), respectively, have been reported as very important indices for model accuracy analysis.

The *k*-value was used to assess the classification ability of the classification models. Generally, between $0 \le k \le 1$, the greater the *k*-value is, the better the classification ability of the models. The *k*-value is calculated by Equations (6)-(8):

$$k=\frac{P_0-P_e}{1-P_e} \tag{6}$$

$$P_0=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN}} \tag{7}$$

$$P_e=\frac{(\mathrm{TP}\times\mathrm{FP})+(\mathrm{TN}\times\mathrm{FN})+(\mathrm{TP}\times\mathrm{FN})+(\mathrm{TP}\times\mathrm{FP})}{(\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN})^2} \tag{8}$$

where, P_0 is the observed proportional agreement between true values and predicted values, and P_e is the expected agreement by chance.

The accuracy, precision, recall, and F_1-value can be obtained by Equations (9)-(12):

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FN}+\text{FP}} \times 100\% \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \quad (11)$$

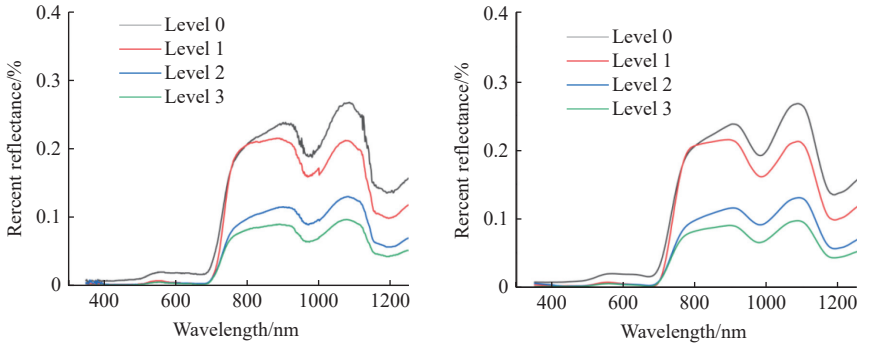
$$F_1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

The square structure of a confusion matrix is represented through rows and columns, where rows are the true classes of the instances, and columns are the predicted classes[32]. The confusion matrix shows the numbers of similarities and differences between true values and predicted values. For multi-class problems, a concern fusion matrix with the n class has an $n \times n$ confusion matrix[33], where the size of $n \times n$ associated with a classifier shows the true and predicted classification, and n represents the number of different classes. Because four different classes of BPH hazard level were set in this study, n=4 was set in the present study.

3 Results and discussion

3.1 Spectral wavebands smoothing

As shown in Figure 5, after S-G smoothing processing, the noise information, including environmental information, illumination influence, etc., was eliminated from original spectral wavebands, which was beneficial for spectral sensitive wavelengths selection.



a. The original spectral wavebands of BPH level 0, level 1, level 2 and level 3

b. The spectral wavebands of BPH level 0, level 1, level 2 and level 3 after S-G smoothing.

Figure 5 Spectral wavebands smoothing processing

3.2 Spectral sensitive wavelength selection

The PCA algorithm was used to remove the collinearity between multiple bands and reduce the data dimension, thus effectively improving the effects of the models. The sensitive wavelengths from spectral wavelengths of rice canopy wavelengths, rice stem wavelengths, fusion information, and rice canopy spectral images were selected based on the contribution rate and variance percentage (Table 6).

Table 6 Variance percentage and contribution rate of each PCA variable for spectral wavelengths of rice canopy, rice stem, fusion information, and rice canopy spectral images

PCA variables	Rice canopy spectral wavelengths		Rice stem spectral wavelengths		Fusion information		Rice canopy spectral images	
	Variance percentage/%	Contribution rate/%	Variance percentage/%	Contribution rate/%	Variance percentage/%	Contribution rate/%	Variance percentage/%	Contribution rate/%
PC1	46.11	46.11	85.09	85.09	41.18	41.18	90.68	90.68
PC2	26.06	72.18	9.23	94.32	19.99	61.17	8.43	99.11
PC3	7.13	79.31	1.72	96.04	12.94	74.11	0.43	99.54
PC4	5.74	85.04	0.77	96.81	10.4	84.51	0.25	99.78
PC5	3.53	88.57	0.69	97.49	3.03	87.54	-	-
PC6	2.89	91.46	0.57	98.06	2.49	90.03	-	-

Based on the contribution rate in Table 6, PC1-PC4, PC1-PC3, PC1- PC5, and PC1 were utilized as the principal components for sensitive wavelength selection from spectral wavelengths of rice canopy, rice stem, fusion information, and rice canopy spectral images, respectively. The principal component load denotes the correlation coefficient between each principal component and the wavelength position. The peak and valley point denotes the maximum weight coefficient of a local area; the wavelengths of these positions were determined as sensitive wavelengths. The load curves of the principal component for different datasets are shown in Figure 6.

Twenty-six, fourteen, forty, and five spectral sensitive wavelengths on positions of the principal component load curves were selected from spectral wavelengths of rice canopy, rice stem, fusion information, and rice canopy spectral images, respectively. These sensitive wavelengths are listed in Table 7.

These selected sensitive wavelengths were used for BPH hazard level modeling. It is important to note that, in the sensitive wavelength selection for stem spectral wavelengths, although PC1 explains about 85.09% of the data variation and contribution, PC2 and PC3 were also selected to improve the accuracy and precision of BPH hazard level classification.

3.3 Effect evaluation of models

Spectral reflectance of sensitive wavelengths from rice canopy spectral wavelength, rice stem spectral wavelength, and fusion information were used for BPH hazard level classification modeling by Back Propagation Neural Network (BPNN), Random Forest (RF), and BLS. The optimum modeling algorithm was used to build the BPH hazard level classification model for sensitive wavelength from rice canopy spectral images. The BPH hazard level classification effects of different modeling algorithms and datasets are listed in Table 8.

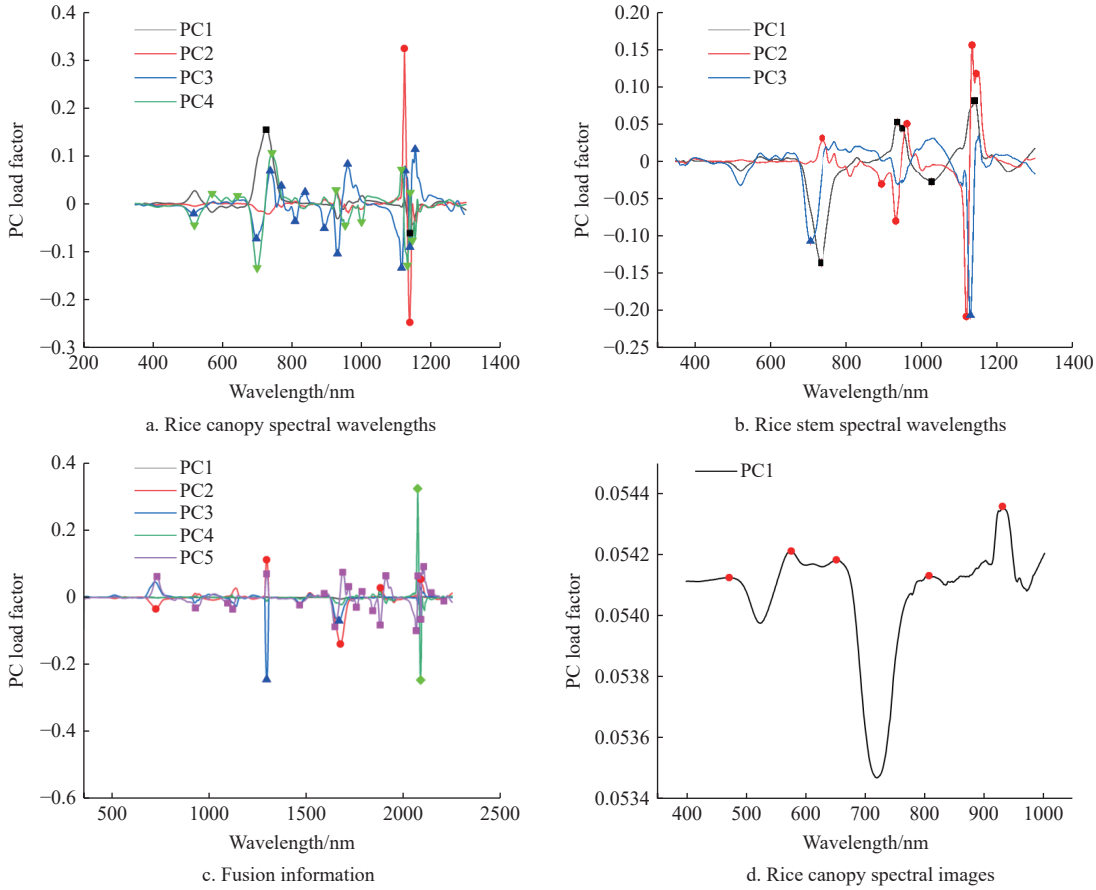


Figure 6 Load curves of the principal component

Table 7 Sensitive wavelengths selected from different spectral wavelength datasets

Dataset	Number of sensitive wavelengths	Sensitive wavelengths/nm
Canopy spectral wavelengths	26	518, 520, 571, 644, 698, 700, 726, 738, 743, 770, 809, 838, 893, 927, 931, 953, 960, 1000, 1115, 1123, 1125, 1132, 1139, 1140, 1146, 1154
Stem spectral wavelengths	14	704, 732, 737, 813, 894, 931, 935, 962, 1027, 1119, 1130, 1134, 1141, 1129
Fusion information	40	518, 520, 571, 644, 698, 700, 726, 738, 743, 770, 809, 838, 893, 927, 931, 953, 960, 1000, 1115, 1123, 1125, 1132, 1139, 1140, 1146, 1154, 704, 732, 737, 813, 894, 931, 935, 962, 1027, 1119, 1130, 1134, 1141, 1129
Rice canopy spectral wavelengths	5	472.117, 576.101, 651.992, 807.178, 930.622

Table 8 BPH hazard level classification effects evaluation indices of different datasets and different modeling algorithms

Dataset	Modeling algorithm	Accuracy/%	Precision/%	Recall/%	F_1-score	k-value
Rice canopy spectral wavelengths	BP	94.89	94.02	95.93	0.95	0.93
	RF	95.09	93.59	96.78	0.95	0.93
	BLS	98.05	98.91	97.17	0.98	0.97
Rice stem spectral wavelengths	BP	89.20	87.69	90.77	0.89	0.85
	RF	91.37	91.13	92.08	0.92	0.88
	BLS	95.16	96.32	93.46	0.95	0.93
Fusion information	BP	97.96	97.97	97.97	0.98	0.97
	RF	98.30	97.81	98.84	0.98	0.98
	BLS	99.08	99.31	98.83	0.99	0.99
Rice canopy spectral images	BLS	80.63	80.28	77.03	0.79	0.74

Note: The model effects evaluation indices include accuracy, precision, recall, F_1-score, and kappa-value.

3.3.1 Algorithm combination optimization

In this study, PCA and SPA were used for sensitive wavelength selection, and BPNN, RS, and BLS were used for BPH hazard level classification modeling. Table 8 displays that: (1) PCA performed better in sensitive wavelength selection compared with SPA, as the spectral reflectance corresponding to the sensitive wavelengths selected by PCA obtained greater BPH hazard level classification effects (greater accuracy and precision, etc.). (2) BLS performed better in BPH hazard level modeling compared with BPNN and RF, as the BLS model obtained greater accuracy and precision in BPH hazard level classification.

Based on the above results, the algorithm combination of S-G, PCA, and BLS was optimized for wavebands smoothing, sensitive wavelength selection, and modeling, respectively, in the present study. However, it is important to note that these algorithms are just some typical algorithms for sensitive wavelength selection and modeling, and they cannot represent all of the algorithms for sensitive wavelength selection and modeling. There are many other optional algorithms, i.e., first-order differentiation (FD)[34] and sequential forward selection[35] for sensitive wavelength selection, and decision tree (DT)[36] and support vector machine (SVM)[37] for modeling.

3.3.2 Dataset optimization

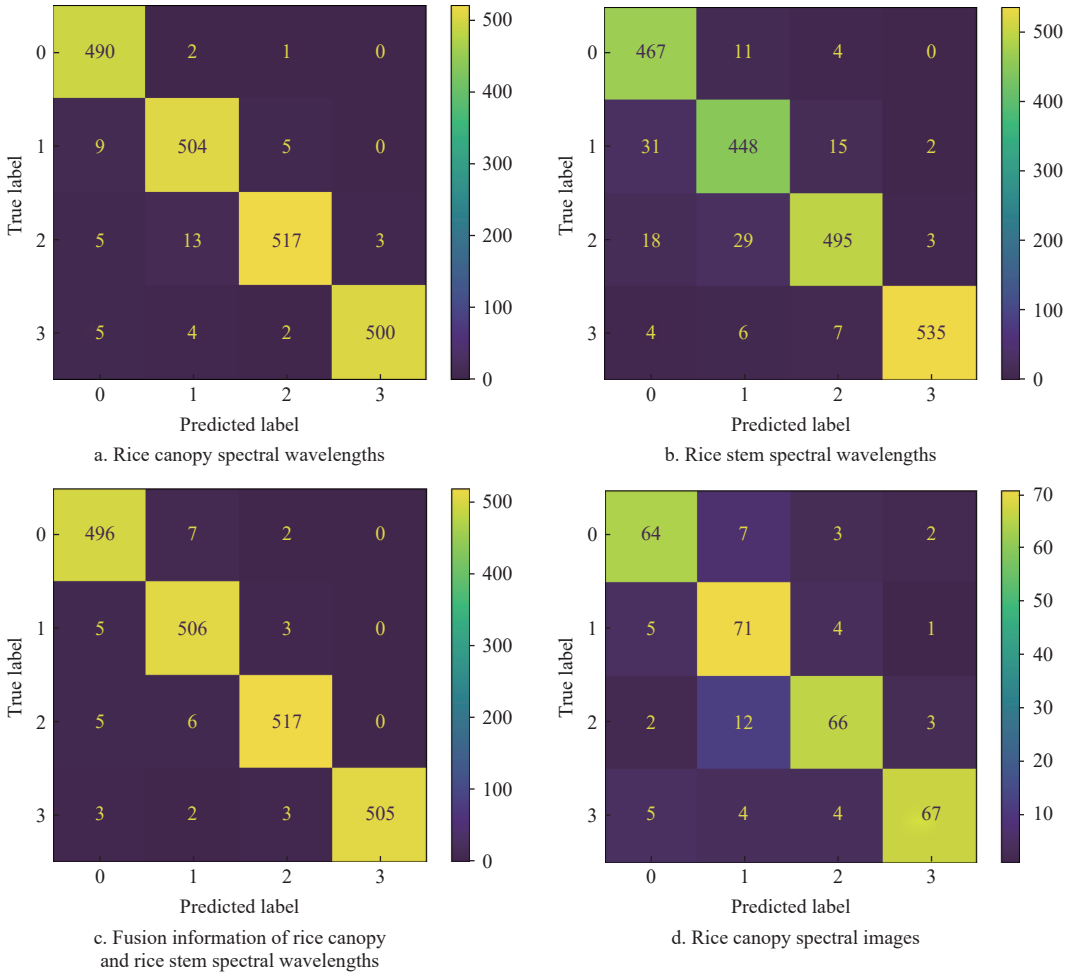
1) Accuracy and precision analysis

Table 8 displays that: (1) The stem spectral reflectance can contribute to improving the accuracy of BPH hazard classification by fusing with the rice canopy spectral reflectance, as the rice canopy and rice stem spectral wavelengths fusion information with BLS model performed best in BPH hazard classification, with the best accuracy (99.08%), precision (99.31%), recall (98.83%), F_1-score (0.99), and kappa-value (0.99). (2) However, the rice stem spectral wavelengths (with accuracy of 95.16%, precision of

96.32%, recall of 93.46%, F_1-score of 0.95, and kappa-value of 0.93) was not a better option compared with rice canopy spectral wavelengths (with accuracy of 98.05%, precision of 98.91%, recall of 97.17%, F_1-score of 0.98, and kappa-value of 0.97) for BPH hazard level classification. (3) Rice canopy spectral images (with accuracy of 80.63%, precision of 80.28%, recall of 77.03%, F_1-score of 0.79, and kappa-value of 0.74) showed potential in large scale measurement of BPH hazard level classification, due to its high efficiency in spectral data collection.

2) Confusion matrix analysis

The confusion matrices for BPH classification model analysis based on different datasets are shown in Figures 7a-7d, where along the *x*-axis are listed the true classes, and along the *y*-axis are the predicted classes.



a. Rice canopy spectral wavelengths

b. Rice stem spectral wavelengths

c. Fusion information of rice canopy and rice stem spectral wavelengths

d. Rice canopy spectral images

Note: BPH hazard level classification based on spectral reflectance corresponding to sensitive wavelengths from the (a), (b), (c), and (d).

Figure 7 Confusion matrix of classification model based on different datasets

Figure 7 and Table 9 show that: (1) The spectral wavelengths fusion information of rice canopy spectral reflectance and rice stem obtained the best agreement between true classes and predicted classes in class 1 (98.44%) and class 2 (97.91%), and the lowest misclassified rate overall. (2) The rice canopy spectral sensitive wavelengths obtained the best agreement between true classes and predicted classes in class 0 (99.39%) and class 3 (98.83%). (3) Rice stem spectral sensitive wavelengths obtained the lowest agreement between true classes and predicted classes in class 0 (96.89%), class 1 (90.32%), class 2 (90.83%), and class 3 (96.92%), and the highest misclassified rate. (4) The classification results of rice canopy spectral images showed great potential in BPH hazard level classification, with an agreement of 84.21% in class 0, 87.65% in class 1, 79.51% in class 2, and 83.75% in class 3.

The confusion matrix analysis results show that a very small ratio of other class samples was misclassified to class 3. In this study, the spectral reflectance of sensitive wavelengths was used for BPH hazard level classification, and the spectral reflectance of sensitive wavelengths decreased as the BPH density increased, as shown in Figure 8.

Table 9 Classification results of prediction models based on different datasets

Dataset	True class	Predicted class			
		0	1	2	3
Canopy spectral reflectance	0	99.39%	9.23%	0.20%	0
	1	1.79%	97.30%	0.99%	0
	2	0.92%	2.42%	96.10%	0.56%
	3	0.59%	0.78%	0.39%	98.83%
Stem spectral reflectance	0	96.89%	2.28%	0.83%	0
	1	6.30%	90.32%	3.02%	0.40%
	2	3.30%	5.32%	90.83%	0.55%
	3	0.72%	1.09%	1.27%	96.92%
Fusion information	0	98.22%	1.39%	0.40%	0
	1	0.97%	98.44%	0.58%	0
	2	0.95%	1.13%	97.91%	0
	3	0.58%	0.39%	0.58%	98.44%
Rice canopy spectral images	0	84.21%	10.93%	3.95%	2.63%
	1	6.17%	87.65%	4.94%	1.23%
	2	2.41%	14.46%	79.51%	3.61%
	3	6.25%	5.00%	5.00%	83.75%

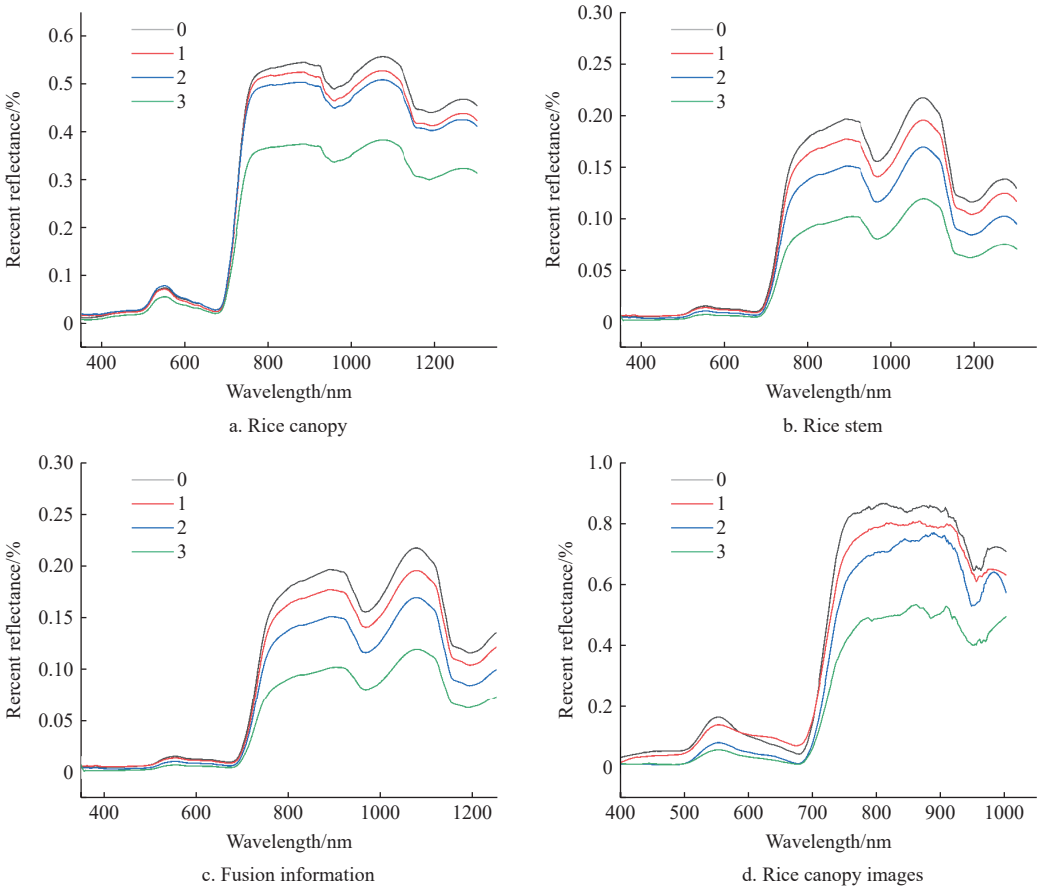


Figure 8 Spectral reflectance corresponding to the spectral wavelength

It is illustrated in Figure 8 that, when the BPH density increased to level 3, the spectral reflectance of sensitive wavelengths decreased significantly compared with BPH density of class 2, which resulted in the significant difference of spectral reflectance of sensitive wavelengths (which was used for BPH hazard level classification) between class 3 and other classes. This might explain the low misclassified rate of other classes as compared to class 3.

4 Conclusions

The present study used three datasets including spectral reflectance of sensitive wavelengths from rice canopy spectral wavelengths, rice stem spectral wavelengths, and spectral wavelengths fusion information of rice canopy and rice stem to develop a BPH hazard level classification model with different algorithms. Datasets and algorithm combinations were optimized by the BPH hazard level classification effects. The optimum algorithm combination was used to build the BPH hazard level classification model for spectral reflectance corresponding to sensitive wavelengths from rice canopy spectral images. The specific research conclusions are as follows:

(1) The most effective algorithm combination for BPH hazard level classification was S-G for spectral bands smoothing, PCA for sensitive wavelength selection, and BLS for modeling, as the best accuracy and precision were obtained by this algorithm combination.

(2) The fusion information was the most effective dataset for BPH hazard level classification among the three datasets, with an accuracy of 99.08% and precision of 99.31%. The dataset spectral reflectance corresponding to sensitive wavelengths from rice canopy spectral wavelength performed better in BPH hazard level classification (with an accuracy of 98.05% and precision of 98.91%) compared with that of rice stem (with an accuracy of 95.16% and precision of 96.32%). It is observed that the rice stem spectral reflectance can improve the accuracy and precision of BPH hazard level classification when it is fused with canopy spectral reflectance.

(3) The optimized algorithm combination was used for BPH hazard level classification by using rice canopy spectral images, with an accuracy of 80.63% and precision of 80.28%. This shows the potential of rice canopy spectral images in BPH hazard level classification. However, in the present study, only a small volume of rice canopy spectral images was collected, while in the future more rice canopy spectral images should be collected for BPH hazard level classification.

The results in this study provide a theoretical basis for exploring BPH hazard level classification based on spectral reflectance fusion information of rice stem and canopy spectral reflectance, and rice canopy spectral images. Further research should be undertaken to develop a robot which can detect the spectral reflectance of rice canopy and rice stem, and then provide fusion information for BPH hazard level classification, thereby enhancing the effects of BPH hazard level classification by rice canopy spectral images.

Acknowledgements

This research was financially supported by Key Research and Development Project of China (Grant No. 2022YFD2002400), the Key Scientific and Technological Projects in Key Areas of Crops (Grant No. 2023AB014), the National Natural Science Foundation of China (Grant No. 31901401), the Key Area Research and Development Program of Guangdong (Grant No.

2023B0202130001), and Guangdong Provincial Basic and Applied Basic Research Fund Project (Grant No. 2022A1515011528).

[References]

[1] Kang K, Yue L, Xia X, Liu K, Zhang W Q. Regulatory function of the trehalose‐6‐phosphate synthase gene TPS3 on chitin metabolism in brown planthopper, Nilaparvata lugens. Insect Molecular Biology, 2022; 31(2): 241–250.

[2] Min S, Lee S W, Choi B R, Lee S H, Kwon D H. Insecticide resistance monitoring and correlation analysis to select appropriate insecticides against Nilaparvata lugens (Stl), a migratory pest in Korea. Journal of Asia-Pacific Entomology, 2014; 17(4): 711–716.

[3] Gomez-Chova L, Calpe J, Camps-Valls G, Martín J D, Soria E, Vila J, et al. Feature selection of hyperspectral data through local correlation and SFFS for crop classification. IEEE International Geoscience & Remote Sensing Symposium. IEEE, 2003; pp.555–557.

[4] Liang K, Ren Z Z, Song J P, Yuan R, Zhang Q. Wheat FHB resistance assessment using hyperspectral feature bandimage fusion and deep learning. Int J Agric & Biol Eng, 2024; 17(2): 240–249.

[5] Carter G A, Miller R L. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. Remote Sensing of Environment, 1994; 50: 295–302.

[6] Delalieux S, Somers B, Verstraeten W W, Van Aardt J A N, Keulemans W, Coppin P. Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. International Journal of Remote Sensing, 2009; 30(8): 1887–1912.

[7] Prasannakumar N R, Chander S, Sahoo R N, Gupta V K. Assessment of Brown Planthopper, (Nilaparvata lugens) [Stl], damage in rice using hyperspectral remote sensing. Pans Pest Articles & News Summaries, 2013; 59(3): 180–188.

[8] Prasannakumar N R, Chander S, Sahoo R N. Characterization of brown planthopper damage on rice crops through hyperspectral remote sensing under field conditions. Phytoparasitica, 2014; 42(3): 387–395.

[9] Huang J R, Sun J Y, Liao H J, Liu X D. Detection of brown planthopper infestation based on SPAD and spectral data from rice under different rates of nitrogen fertilizer. Precision Agriculture, 2015; 16(2): 148–163.

[10] Tan Y, Sun J Y, Zhang B, Chen M, Liu Y, Liu X D. Sensitivity of a ratio vegetation index derived from hyperspectral remote sensing to the brown planthopper stress on rice plants. Sensors, 2019; 19(375). DOI: 10.3390/s19020375

[11] Sōgawa K. The rice brown planthopper: Feeding physiology and host plant interactions. Annual Review of Entomology, 1982; 27: 49–73.

[12] Nault B A, Taylor A G, Urwiler M, Rabaey T, Hutchison W D. Neonicotinoid seed treatments for managing potato leafhopper infestations in snap bean. Crop Protection, 2004; 23(2): 147–154.

[13] Liu Y, Liu S, Xu J, Kong X, Xie L, Chen Y, et al. Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy. Computers and Electronics in Agriculture, 2002; 192: 1–14.

[14] Wang K, Chen K, Du H, Liu S, Xu J, Zhao J, Chen H, Liu Y. New image dataset and new negative sample judgment method for crop pest recognition based on deep learning models. Ecological informatics: an International Journal on Ecoinformatics and Computational Ecology, 2002; 69. DOI:10.1016/j.ecoinf.2022.101620

[15] Erfani S M, Rajasegarar S, Karunasekera S, Leckie C. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition, 2016; 58(C): 121–134.

[16] Guo W J, Feng S, Feng Q, Li X Z, Gao X Z. Cotton leaf disease detection method based on improved SSD. Int J Agric & Biol Eng, 2024; 17(2): 211–220.

[17] Chen C L P, Liu Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE Transactions on Neural Networks & Learning Systems, 2018; 29(99): 10–24.

[18] Yasuda T, Bateni M H, Chen L, Fahrbach M, Fu G, Mirrokni V. Sequential attention for feature selection. arXiv e-prints. ICLR, Kigali Rwanda, 2023. DOI: 10.48550/arXiv.2209.14881

[19] Borzov S M, Potaturkin O I. Increasing the classification efficiency of hyperspectral images due to multi-scale spatial processing. Computer Optics, 2020; 44(6): 937–943.

[20] Sun J, Yang W, Zhang M, Feng M, Ding G. Estimation of water content in corn leaves using hyperspectral data based on fractional order Savitzky-Golay derivation coupled with wavelength selection. Computers and Electronics in Agriculture, 2021; 182(1). DOI: 10.1016/j.compag.2021.105989

[21] Zimmermann B, Kohler A. Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. App.Spectrosc, 2013; 67(8): 892–902.

[22] Wold S, Esbensen K, Geladi P. Principal component analysis, Chemometr. Intell. Lab. Syst, 1987; 2(1987): 37–52.

[23] Hervé A, Williams L J. Principal component analysis. Wiley Interdisciplinary Reviews Computational Statistics, 2010; 2(4): 433–459.

[24] Beattie J R, Esmonde-White F W L. Exploration of principal component analysis: deriving principal component analysis visually using spectra. Applied Spectroscopy, 2021; 75(4): 361–375.

[25] Rodionova O, Kucheryavskiy S, Pomerantsev A. Efficient tools for principal component analysis of complex data—A tutorial. Chemometrics and Intelligent Laboratory Systems, 2021; 213: 1–11.

[26] Saloni K, Deepika K, Mamta M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier - ScienceDirect. International Journal of Cognitive Computing in Engineering, 2021; 2: 40–46.

[27] Machart P, Ralaivola L. Confusion matrix stability bounds for multiclass classification. Computer Science, 2012; 1-11. DOI:10.48550/arXiv.1202.6221

[28] Simon D, Simon D L. Analytic confusion matrix bounds for fault detection and isolation using a sum-of-squared-residuals approach. IEEE Transactions on Reliability, 2021; 59(2): 287–296.

[29] Paulino J L C, Almirol L C A, Favila J M C, Aquino K A G L, Cruz A H D L, Roxas R. Multilingual sentiment analysis on short text document using semi-supervised machine learning. 5th International Conference on E-Society, E-Education and E-Technology, 2021. DOI: 10.1145/3485768.3485775

[30] Hasnain M, Pasha M F, Ghani I, Imran M, Alzahrani M Y, Budiarto R. Evaluating trust prediction and confusion matrix measures for web services ranking. IEEE Access, 2020; 8: 90847–90861.

[31] Lydia A, Meena K, Sekar R R, Swaminathan J N. Parkinson's disease prediction through machine learning techniques. In: Chen J I Z, Wang H, Du K L, Suma V (eds.) Machine Learning and Autonomous Systems. Smart Innovation, Systems and Technologies, 269. Springer, Singapore, 2022.

[32] Caelen O. A Bayesian interpretation of the confusion matrix. Ann. Math. Artif. Intell, 2017; 81(3-4): 429–450.

[33] Rajalakshmi R, Aravindan C. A Naive Bayes approach for URL classification with supervised feature selection and rejection framework. Computational Intelligence, 2018; 34(1): 363–396.

[34] Popescu S A, Marilena J N. First-order differential equations. in: advanced mathematics for engineers and physicists. Springer, Cham, 2022.

[35] Cotter S F, Adler R, Kreutz-Delgado K. Forward sequential algorithms for best basis selection. Vision Image & Signal Processing IEE Proceedings, 1999.

[36] Ramirez S, Lizarazo I. Decision tree classification model for detecting and tracking precipitating objects from series of meteorological images. Semantic Scholar, 2016. DOI: 10.3990/2.384

[37] Bruzzone L, Chi M, Marconcini M. A novel transductive SVM for semisupervised classification of remote-sensing images. IEEE Transactions on Geoscience & Remote Sensing, 2006; 44: 3363–3373.