Early Identification of Rice Leaf Blast Based on Hyperspectral Imaging

Jian Qing Yuan 1, *, Lei Li 1 and Wei Yan 1
1Department of Computer Science, Harbin Finance University, Harbin, 150030, China
*2017018@hrbfu.edu.cn

Abstract. Rice blast is one of the three major rice diseases recognized in the world, which greatly harms the quality and the yield of rice. In order to distinguish rice leaf blast disease from nutrient deficiency and diagnose early the leaf blast disease, this study was based on the natural incidence of rice and field experiments, hyperspectral imagers were used to obtain the imaging spectrum of health, nitrogen deficiency, mild disease and severe disease. Spectra of 4 types of leaves were extracted, and three kinds of different data pretreatment methods were used, and the SPA feature extraction method was combined with the support vector machine(SVM) and the linear discriminant analysis(LDA) to construct the rice leaf blast identification model. The experimental results show that, after preprocessing by the Savitzky-Golay method, 9 characteristic wavelengths were extracted by SPA for modeling, and the models had the best recognition effect. The prediction accuracy of the SG-SPA-SVM model and the SG-SPA-LDA model were both 98.7%.

1. Introduction
Rice blast is one of the three major rice diseases recognized in the world, which greatly harms the quality and the yield of rice. Therefore, the monitoring and control of rice blast has always been an important issue in the field of rice cultivation. Hyperspectral imaging technology combines image and spectrum technology, and can accurately provide continuous spectral information of each pixel of the detection target [1]. Because of its non-destructive and high efficiency, Hyperspectral imaging technology is widely used in vegetables, fruits and meat, quality testing, crop nutrient testing and various pests and diseases testing, etc [2-3]. The following researchers used hyperspectral imaging technology to study different rice diseases. Li Zhiwei et al. [4] applied hyperspectral technology to identify rice sheath blight; Zhu Mengyuan et al. [5] studied the early recognition of sheath blight based on hyperspectral data and chlorophyll content; Kobayashi T et al. [6] used aerial hyperspectral images to identify Rice blast region identification was studied; Huang Shuangping et al. [7] studied the hyperspectral detection method of panicle blast. So far, the research on rice blast leaf detection has mainly focused on identifying healthy and diseased leaves, as well as grading detection of the degree of infection.

Spectroscopy technology is used in early detection of vegetation diseases and insect pests, because when vegetation is infested by diseases and insect pests, it will cause changes in its internal components (such as chlorophyll, moisture, etc.) and structure, and its external features are the chlorosis and yellowing of vegetation leaves [8]. While the lack of nutrients in rice, such as nitrogen, potassium, etc., will also have similar phenomena, which may be confused with disease characteristics to a certain extent. Therefore, the distinction between them is of great significance for field
management, and there are few reports in this area. In this study, rice leaves infected with rice blast and nitrogen stress were selected as the research object. Different spectral pretreatment methods were selected, and the SPA feature extraction method combined with support vector machine and discriminant analysis were used to construct rice leaf blast hyperspectral recognition models. It is expected to provide a new method for the early identification of rice leaf blast.

2. Materials and methods

2.1. Experimental samples
The experimental samples were collected in Datian, Fangzheng County Rice Research Institute, Harbin City, Heilongjiang Province. The two test areas occupied an area of 380m² respectively. The normal nitrogen fertilizer treatment (150kg/hm²) each occupied 280m², and the low nitrogen treatment (50kg/hm²) each occupied 100m². The district did not carry out pest control treatment, other field management was normal, and rice leaf blast occurred naturally. The samples were collected at the jointing stage of rice, and 60 samples each of finally healthy leaves, nitrogen-deficient leaves, mildly susceptible leaves (small brown spots, affected area within 15%) and Severe susceptible leaves (more brown spots or large gray spots appeared, affected area was more than 25%) were obtained, totaling 240 samples.

2.2. Hyperspectral imaging acquisition system and data acquisition
The experiment used the hyperspectral imaging system produced by the American HeadWall company to collect sample hyperspectral data. The system consists of a hyperspectral camera, light source (150W adjustable halogen lamp), mobile platform, light box, collector and computer. The imaging method of the sensor is linear array push scan, the spectral range is 400~1000nm, the spectral resolution is 2.4nm, and the sampling interval is 3nm.

When capturing images, set the exposure time to 30ms, and the moving speed of the moving platform to 3.0mm/s. Place three rice leaves in a group on the black cardboard on the moving platform, with the lens vertically downward and 45cm away from the moving platform. The light current and dark current correction were performed before measurement. The correction formula is:

\[ R = \frac{I_{\text{raw}} - I_{\text{black}}}{I_{\text{white}} - I_{\text{black}}} \]  \hspace{1cm} (1)

Using ENVI5.1 software to extract the spectral reflectance values of various leaves, three rectangular areas (including diseased spots) were selected in the upper middle part of each leaf as the area of interest, and the average reflectance of all pixels in the area of interest were calculated to use as the spectral reflectance value of a leaf sample.

2.3. Spectral data preprocessing
In order to eliminate or reduce the interference of baseline drift, high-frequency noise, stray light, etc. caused by instrument and environmental noise on the original spectral information, the study used polynomial convolution smoothing Savitzky-Golay (S.G), standard normal variable (SNV), and multiplicative scatter correction (MSC) three preprocessing methods of spectral data to preprocess the original spectra. The S.G method can eliminate random noise well; the SNV method is mainly used to correct the spectral error caused by the scattering of the sample; the MSC method is mainly used to eliminate the scattering effect caused by the uneven distribution of the target particles. The preprocessed spectrum data was used to construct rice leaf blast prediction models and compared with the original spectrum to compare the recognition and prediction performance of each model.

2.4. Extraction of feature variables
The spectral curve in the hyperspectral image has extremely high spectral resolution, the number of bands is large, the amount of data is very large, and the information contained in different bands has multiple correlations. Therefore, it was necessary to compress hyperspectral data and extract effective
spectral information for modeling. The SPA continuous projection algorithm can find the least redundant information variable group in the spectral data through the projection analysis of the vector, which minimizes the correlation between the variables, greatly reduces the number of variables used for modeling, and improves the calculation speed and the effectiveness of the model.

The study used the SPA algorithm to extract the characteristic wavelengths for distinguishing nitrogen-deficiency and rice blast leaves from the original reflectance spectrum and the spectrum after three kinds of pretreatments, which were used to establish distinguishing recognition models.

2.5. Model building algorithm
Support vector machine (SVM) is a new type of modeling method, which improves generalization ability through the principle of structural risk minimization, and shows outstanding advantages in data processing problems of small samples, nonlinearity and high-dimensional feature spaces.

Linear discriminant analysis (LDA) first selects the variable with the most significant discriminative ability from all variables, and then selects the significant variable with the smallest correlation with the selected variable as the next introduced variable, at the same time, judges the selected variable significance, and deletes those variables that have lost significant discriminative ability due to the introduction of new variables to ensure that the selected variable combination has the most significant discriminant analysis ability for explanatory variables.

The research used SVM algorithm and LDA method to construct the classification model of rice leaves on the characteristic wavelengths extracted by SPA. 40 samples of four types of rice healthy, nitrogen deficiency, mild disease and severe disease were randomly selected, a total of 160 samples were used for model establishment, and the remaining 80 were used as prediction samples.

3. Results and analysis

3.1. Spectral preprocessing
The original reflectance spectra of all rice leaf samples, Savitzky-Golay smoothing, standard normal variable transformation, and multi-scatter correction processed spectral information were shown in Figure 1.

![Figure 1. The reflectance spectrum after preprocessing.](image-url)
3.2. The SPA feature extraction

The SPA algorithm was used to extract the characteristic wavelengths of healthy, nitrogen-deficient and rice blast leaves from the original reflectance spectrum and the pre-processed spectrum, which were used to establish a distinguishing recognition model. The selection results of SPA characteristic wavelengths of different preprocessing methods were shown in Figure 2.

The number of wavelength points and spectral positions of the spectral features extracted by SPA were shown in Table 1. The characteristic wavelengths extracted from the original reflectance spectrum included 4 blue bands, 1 green band, 1 red edge position, and 2 near red, totaling 8 features; the characteristic wavelengths extracted by the spectrum after S.G smoothing processing included 2 blue bands, 2 green light bands (including green peaks), 2 red light bands, 1 red edge position, and near red 1 total 9 features; the characteristic wavelengths extracted from the spectrum after SNV processing included 2 blue light bands and green light bands One, one at the red edge position, and the remaining 7 near-red, a total of 11 features; the characteristic wavelengths extracted from the spectrum after MSC processing were mainly concentrated in the blue band and near-red range, totaling 8 features. Judging from the band range of the characteristic wavelengths extracted from the spectrum after different pre-processing methods by SPA, the characteristic wavelengths of the S.G smoothed spectrum were the most comprehensive, and the characteristic wavelengths were distributed in the blue, green, red, near-red, and red edges, and there are only 9 spectral features.

Table 1. Statistical results of SPA characteristic band with different preprocessing methods.

| Preprocessing method | Number of characteristic wavelengths | Characteristic spectral positions |
|----------------------|--------------------------------------|----------------------------------|
|                      |                                      |                                  |
Discriminant model based on SPA

The characteristic wavelengths extracted by SPA were used as the input vector of the support vector machine algorithm and the stepwise discriminant analysis method to establish the distinguishing model of health, nitrogen deficiency, mild disease and severe disease. The 180 samples in the modeling set were used for model training and verification, and the remaining 80 samples were used as the prediction set to test the predictive discrimination ability of the model.

3.3.1. SPA-SVM discriminant model. The SVM algorithm used Linear, PLOY, RBF and Sigmoid 4 classical kernel functions to establish the classification model. The search range of the support vector classifier’s penalty parameters and other kernel function parameters were $10^{-3}$ ~ $10^5$, and the initial parameters were in the search Randomly generated numbers within the value range, using grid search method to determine the optimal combination of the parameters of the support vector classifier. The prediction results of the classification model corresponding to different preprocessing methods and the best vector machine base parameters were shown in Table 2.

| Preprocessing method | Kernel function | Parameters | Prediction set accuracy rate% |
|----------------------|-----------------|------------|------------------------------|
|                      |                 |            | Health | Nitrogen deficiency | Mild infection | Severe infection | Overall |
| the original         | Linear          | $\sigma = 100$ | 95.5  | 100   | 93.3        | 93.3   | 95.0      |
|                      | Polynomial      | $\sigma = 100, \gamma = 10$ | 95.0  | 100   | 60   | 100 | 89.2 |
|                      | RBF             | $\sigma = 1000, \gamma = 1$ | 100   | 100   | 86.7 | 86.7 | 93.4 |
|                      | Sigmoid         | $\sigma = 100, \gamma = 0.1$ | 100   | 100   | 66.7 | 93.3 | 89.2 |
| S. G                 | Linear          | $\sigma = 100$ | 95.0  | 100   | 93.3 | 100 | 96.3 |
|                      | Polynomial      | $\sigma = 100, \gamma = 1$ | 100   | 100   | 93.3 | 93.3 | 96.3 |
|                      | RBF             | $\sigma = 100, \gamma = 10$ | 95.0  | 100   | 100 | 100 | 98.7 |
|                      | Sigmoid         | $\sigma = 100, \gamma = 1$ | 95.0  | 100   | 73.3 | 100 | 92.3 |
| SNV                  | Linear          | $\sigma = 100$ | 100   | 100   | 86.7 | 93.3 | 95.0 |
|                      | Polynomial      | $\sigma = 100, \gamma = 0.09$ | 95.0  | 100   | 86.7 | 93.3 | 93.4 |
|                      | RBF             | $\sigma = 100, \gamma = 0.1$ | 95.0  | 100   | 86.7 | 93.3 | 93.4 |
|                      | Sigmoid         | $\sigma = 1000, \gamma = 0.01$ | 100   | 100   | 86.7 | 93.3 | 95.0 |
| MSC                  | Linear          | $\sigma = 100$ | 95.0  | 100   | 93.3 | 100 | 96.3 |
|                      | Polynomial      | $\sigma = 0.01, \gamma = 100$ | 100   | 86.7 | 93.3 | 93.3 | 93.4 |
|                      | RBF             | $\sigma = 1000, \gamma = 0.125$ | 100   | 100   | 93.3 | 93.3 | 96.3 |
|                      | Sigmoid         | $\sigma = 1000, \gamma = 0.25$ | 95.0  | 100   | 93.3 | 93.3 | 95.0 |

From the prediction results, the prediction accuracy of the SVM discriminant model established with the SPA characteristic wavelengths of the four types of spectra was all above 89.2%, and the overall effect was very good. It can also be seen that the overall recognition accuracy of the spectrum processed by different preprocessing methods was improved to different degrees than the original reflectance spectrum. From the analysis of the misjudgment situation, the most serious misjudgment was the misjudgment between mild and severe samples. There were also cases where healthy samples were misjudged as nitrogen-deficient samples and mild samples were misjudged as nitrogen-deficiency samples. Only a few models misjudged mild samples as nitrogen-deficient and healthy samples.
samples. The model with the highest classification accuracy rate was the SPA-SVM model, which was preprocessed by S.G smoothing processing and RBF as the kernel function. In the prediction, only one healthy sample was misjudged as a nitrogen deficiency sample, and the overall recognition accuracy rate was 98.7%.

3.3.2 SPA-LDA discriminant model. Fisher linear discriminant was used for discriminant analysis, and the "leave one out" cross-validation method was used when building the model. The prediction results of the SPA-LDA modeling corresponding to the four types of spectra were shown in Table 3. The modeling cross-validation and prediction results of each model were very good. The model classification accuracy of the original spectrum was the lowest, the modeling verification set also reached 92.8%, and the prediction accuracy was also above 93.8%. The SPA-LDA prediction accuracy of SG, SNV and MSC preprocessing methods were improved to different degrees compared with the original spectrum. Among them, the best effect of comprehensive cross-validation classification results and prediction classification results was the SPA-LDA construction corresponding to the SG preprocessing method. The modeling cross-validation accuracy was 96.9%, and the prediction accuracy rate was 98.7%, as shown in Figure 3. It can also be seen from the figure that healthy leaves, nitrogen-deficient leaves and diseased leaves could be distinguished well, there are misjudgments between mild infection and Severe infection.

Table 3. Prediction results of SPA-LDA models with different preprocessing methods.

| Preprocessing method | Model set | Prediction set |
|----------------------|-----------|----------------|
|                      | Number of samples | Correct number | Accuracy rate% | Number of samples | Correct number | Accuracy rate% |
| 无                   | 160        | 148            | 92.5            | 80              | 75             | 93.4            |
| S.G                  | 160        | 155            | 96.9            | 80              | 79             | 98.7            |
| SNV                  | 160        | 156            | 97.5            | 80              | 76             | 95.0            |
| MSC                  | 160        | 155            | 96.9            | 80              | 77             | 96.3            |

Figure 3. Classification plot of discriminant analysis based on SPA characteristic band.

4. Conclusion
The research took field healthy leaves, naturally-occurring rice leaves, and nitrogen stress experiment nitrogen-deficient leaves as the research objects. Hyperspectral imaging technology was used to obtain
spectral information of 4 types of leaves, and different spectral reflectance preprocessing methods, SPA algorithms and classification algorithms were used to build rice leaf blast classification models. The main conclusions were:

(1) The study used three methods of SG, SNV and MSC to preprocess the original spectrum. After the characteristic wavelengths were extracted by the SPA algorithm, whether the SVM algorithm or the discriminant analysis method was used for classification, the recognition effect of the original spectrum was improved.

(2) Using the SPA algorithm to extract characteristic spectra from the original spectrum and the three pre-processed spectra, the characteristic wavelengths of the SG smoothed spectrum were the most comprehensive, and the characteristic wavelengths were distributed in blue, green, red, near red and Red edge, and there were only 9 spectral features, the accuracy of the classification model constructed using these 9 features was also the best.

(3) The model with the best recognition effect was the SPA-SVM model preprocessed by S.G and RBF as the kernel function and the SPA-LDA model preprocessed by S.G. The overall recognition accuracy rate was 98.7%.

Experimental results show that the study used the SPA method to extract characteristic spectra from various pre-processed spectra and built classification models. The models could effectively distinguish healthy, nitrogen-deficient, mildly diseased and Severe diseased leaves. It provides a scientific basis for rice leaf blast monitoring, early control, and precise application of pesticides, and it also provides a basis for the development of portable equipment for rapid detection of rice diseases.

Acknowledgment
The authors would like to thank for the support of the Fundamental Scientific Research Business Expenses project of Heilongjiang Provincial Undergraduate University Projects (No: 2020-KYYWF-E009).

References
[1] Brown J K M, Hovmoller M S. Epidemiology-Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease[J]. Science, 2002, 297(5581): 537—541.
[2] Wei Xinhua, Wu Shu, Fan Xiaodong, et al. Identification of Slight Bruises on Winter Jujube Based on Hyperspectral Imaging Technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(3): 242—246.
[3] Tian Youwen, Cheng Yi, Wang Xiaoji, et al. Recognition method of insect damage and stem/calyx on apple based on hyperspectral imaging[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 23(4): 325-331.
[4] LI Zhiwei,YUAN Jing, DING Weimin, et al. Identification of rice sheath blight based on hyperspectral imaging technique [J]. Journal of South China Agricultural University, 2018, 39(06):103-109.
[5] ZHU Meng-yuan,YANG Hong-bing,LI Zhi-wei. Early Detection and Identification of Rice Sheath Blight Disease Based on Hyperspectral Image and Chlorophyll Content [J]. Spectroscopy and Spectral Analysis, 2019, 39(6).
[6] Kobayashi T, Sasahara M, Kanda E, et al. Assessment of Rice Panicle Blast Disease Using Airborne Hyperspectral Imagery[J]. open agriculture journal, 2016, 10(1):28-34.
[7] Huang Shuangping,Sun Chao,Qi Long, et al. Rice panicle blast identification method based on deep convolution neural network [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017(20):177-184.
[8] Wu Shuwen, Wang Renchao, Chen Xiaobin, et al. Effects of rice leaf blast on spectrum reflectance of rice[J]. Journal of Shanghai Jiaotong University: Agricultural Science, 2002, 20(1): 73—80.