Early detection of rice blast (*Pyricularia*) at seedling stage in Nipponbare rice variety using near-infrared hyper-spectral image

Yan Yang\(^1,2\), Rongyao Chai\(^3\) and Yong He\(^1\)*

\(^1\)College of Bio-systems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.
\(^2\)School of Mathematics Science, Guangxi Teacher Education University, Nanning 530023, China.
\(^3\)Institute of Plant Protection and Microbiology, Zhejiang Academy of Agricultural Science, Hangzhou 310021, China.

Accepted 26 January, 2012

Blast rice is the worst biological disaster in rice cultivation. It reduces the yield at least up to 40 to 50% (in the worst period of disease). In this study, the near-infrared hyper-spectral image was investigated to detect blast rice in Nipponbare at seedling stage. Two hundred rice seedlings were segregated into two classes: infected and healthy. All of rice seedlings were scanned with a hyper-spectral imaging system in the NIR (900 to 1700 nm) wavelength range. Principal component analysis (PCA) was performed on the images and the distribution of PCA scores within individual leaf were measured to develop linear discriminant analysis (LDA) models for predicting the infected leaves from healthy leaves. An LDA model classified all the leaves into infected and healthy categories, with an overall accuracy of 92% on validation set. Meanwhile, the classification model base on five selected wavelengths (1188, 1339, 1377, 1432 and 1614 nm) was comparable to that base on the full-spectrum image data.

**Key words:** Rice blast (*Pyricularia*), Nipponbare, near-infrared hyper-spectral image, principal component analysis, linear discriminant analysis.

**INTRODUCTION**

Rice Blast (*Pyricularia*) is a fungus disease which is favored by wet weather at rice growth period. According to the difference of infected position, rice blast is classified into seedling blast, leaf blast, rice node blast, neck blast and corn blast. As a result of rice blast, rice lost may be up to 40 to 50% at worst. With the climate becoming more severe, the rice blast has been a main biological disaster, especially at southern planting region in China. Statistics showed that the loss caused by rice blast is about 400 to 500 million tons each year in China. The grain production has been influenced seriously, so how to control the rice blast is a serious subject in China. Nowadays, one of the most widely used methods of controlling the rice blast is spraying fungicide. Usually fungicides (often a mixture of two) are applied uniformly despite numerous variations in the disease pressure. However, most disease infestations are not evenly distributed across the field but in patches (Moshou et al., 2005). Uniformly spraying requires an excessive amount of fungicide, which increases the cost, pollutes the environment and results in the development of multi-resistant fungal strains (LaMondia and Douglas, 1997).

The dispersal of pathogens, usually starting as a small, visible focus within a crop (Rapilly, 1979). In certain conditions, the lesion caused by a successful infection (Zadoks and Vandenbosch, 1994) and pathogens can spread rapidly over large distances (Brown and Hovmoller, 2002). Therefore, the early detection of the disease, which aims to determine when and where to target chemicals, could assist farmers to cut inputs in the culture and also significantly reduce residues of chemical in produce and environment. Disease detection can be conducted via micro-detection and macro-detection. The micro-detection is based on the molecular level (Putnam, 1995). The pathogens can be detected and quantified by immunu-

*Corresponding author. E-mail: y_yae@163.com. Tel: +86-571-86971143. Fax: +86-571-86971143.
logical and molecular methods such as enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) (Nutter and Esker, 2006; Jackson et al., 2007) and microscopy (Hilber and Schuepp, 1992). These methods are accurate but time-consuming and costly, most importantly requires professional operation, so it is impossible to be applied in real-time detection. On the other hand, macro-detection is based on remote sensing measurement. Remote sensing can obtain information about an object without having direct physical contact with it (De Jong and Van de Meer, 2006). For the detection of plant disease, general healthy leaves typically exhibit low reflectance at VIS wavelengths by photoactive pigments (chlorophylls, anthocyanins and carotenoids); high reflectance in the near-infrared (NIR) due to multiple scattering at the air-cell interfaces in the leaf internal tissue (Wiesler et al., 2002); and low reflectance in wide wavebands in the short wave infrared (SWIR) due to absorption by water, proteins, and other carbon constituents (Jacquemoud and Ustin, 2001; Woolley, 1971). Once the plant subject to disease stress, change in spectral characteristics would be the best discrimination basis.

The common optical sensors include spectrometer and digital imaging system. Whether spectrometer or digital imaging system, they can quantify disease on plant alone. As previously reported image or spectral analyses have demonstrated their abilities to characterize disease stress of plants (Nilsson and Johnsson, 1996; Riedell and Blackmer, 1999; Richardson et al., 2001; Diaz-Lago et al., 2003; Karcher and Richardson, 2003; Yang et al., 2005). The detection of rice disease is usually based on spectrometer or multi-spectral imaging (Minekawa et al., 2005; Yang et al., 2007; Liu et al., 2010; Zhang et al., 2011). However, the previously related studies indicated that spectrometer cannot realize the visualization of disease stress. On the other hand, while the digital imaging systems can provide the visualization, it cannot provide enough spectral information. Therefore, disease discrimination based on digital image is not very accurate. Meanwhile, integrating the spectrum analysis and image process to detect the disease stress has seldom been documented. The hyper-spectral imaging system provided this possibility to integrate two types of data, including spectral information and image information. The hyper-spectral imaging system can acquire not only the spectral information, but also spatial information of object. Hyper-spectral imagery is new to plant disease severity measurement, and although it is still in the early phase of developmental use in plant disease detection and quantification, it offers interesting opportunities for application (Coops et al., 2003; Larsolle and Muhammed, 2007; Huang et al., 2007; Qin et al., 2008).

The objectives of this study were: (a) to investigate the use of hyper-spectral imaging in the NIR spectral range (900 to 1700 nm) for the detection of blast rice in seedling stage at early infection stage, (b) try to propose a more rapid and accurate classification method to detect the disease stress and (c) to identify a reduced set of wavelengths/wavebands (feature wavelengths) to be used in a future development of a low cost imaging system.

MATERIALS AND METHODS

Samples

Nipponbare was selected as the experimental variety in this study. All the samples were provided by 150 rice seedling plants. The rice blast (Pyricularia) was inoculated by spraying spore suspension after four true leaves have grown. One hundred rice plants inoculated with Pyricularia but not treated with any fungicide, provided infected rice seedlings leaves, while the other not inoculated with Pyricularia provided healthy rice seedlings leaves for comparison. One hundred and fifty leaves collected from rice plants were termed as the calibration set consisting of 75 healthy leaves and 75 infected leaves. Another 50 leaves were termed as the validation set, consisting of 25 healthy leaves and 25 infected leaves.

Hyper-spectral imaging systems

A push-broom type hyper-spectral imaging system in the NIR wavelength range (900 to 1700 nm) was used for spectral measurements of rice seedlings. The imaging system consisted of a prism-grating-prism spectrograph (ImSpector N17E, Specim, Finland), a 12-bit near infrared camera (XLNIR, Xenics, Belgium), a C-mount focusing lens, a motorized carriage and a personal computer. The sample was placed on the mobile carriage that moved along a horizontal direction under the stationary camera. The illumination systems (Illumination, USA) consisting of dual linear light covering the full spectrum from 400 to 1800 nm were used for imaging illumination. Intensity of linear light was regulated through DC regulated fiber optic light source. The software kit (Isuzu Optics, Taiwan, China) was used to acquire the spectral image data. The schematic diagram of hyper-spectral imaging is shown in Figure 1.

Image acquisition and calibration

For imaging, rice seedlings in batches of 5 to 6 leaves per image were placed on the carriage and hyper-spectral image (also known as hyper-cubes) were collected in the diffuse reflectance mode. There are two types hypercube, one is healthy rice seedlings hyper-cube and the other is infected rice seedlings hyper-cube. Image size was 320 by 256 pixels by 256 wavebands within 900 to 1700 nm range at a spectral resolution of approximately 5 nm. The exposure time was set at 4000 us. Dark current and white reference images were collected before imaging each sample to calibrate spectra at each pixel as percent reflectance value. A polytetrefluoroethylene panel with 99% reflectance (Spectralon, Labsphere, USA) was used to collect white reference images. Dark current response images were collected with the lamp off or a cap covering the focusing lens. Calibrated reflectance images (R) were calculated using the following Equation:

$$R = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}}$$

Where, $I_{\text{raw}}$ is the non-calibrated original image of a sample; $I_{\text{white}}$ is the image of the white reference; and $I_{\text{dark}}$ is the dark current image. Calibrated hyper-cubes were subset to keep 208 bands between 950 and 1650 nm for further analyses. Data below 950 nm or above 1650 nm were removed.
Figure 1. The schematic of hyper-spectral imaging system.

Figure 2. The spectral characteristic of rice seedlings (healthy and infected) and background.

nm were excluded due to the presence of excessive noise in this range.

**Spectral characteristics and rice seedlings background separation**

In order to separate rice seedlings from the background, a threshold value was determined based on spectral differences between seedlings and image background. Representative spectra of rice seedlings and background were extracted and visually observed using the region of interest (ROI) tool in the ENVI software (Version 4.5; ITT Visual Information Solutions, Denver, CO, USA). The healthy leaf region and infected leaf region were manually selected at random from image to represent ‘healthy’ and ‘infected’. For each category, a representative spectrum was computed as the average of all pixel spectra in the respective category (Figure 2). Based on the spectral differences between seedlings and the image background as shown by the representative spectra, a threshold value of 2300 at 1466 nm was determined to separate rice seedlings
Figure 3. (a) False color representation of a hyper-spectral image of the healthy rice seedlings; (b) a binary mask image of (a); (c) applied mask image to (a); (d) false color representation of a hyper-spectral image of the infected rice seedlings; (e) a binary mask image of (d); (f) applied mask image to (d).

from the image background for subsequent analyses. A binary mask image was created for each hypercube by thresholding the image band at 1466 nm where pixel intensities less than 2300 were labeled as seedlings (white) and values greater than 2300 were labeled as background (dark) (Figure 3b and e).

Spectral characteristics of healthy rice seedlings and infected rice seedlings

The healthy rice seedlings were selected to represent ‘healthy’ categories (Figure 3a). Similarly, the infected rice seedlings were selected to represent ‘infected’ categories (Figure 3d). A binary mask image was created to separate the objects from background; the rice seedlings were marked (white) and background was marked as (dark) (Figure 3b and e). The image applied the mask image (Figure 3c and f) to eliminate the background information. The rice seedlings original spectra (Figure 4a) were mean-normalized by dividing each spectrum with its mean value computed along the wavelength direction to visualize spectral similarities or differences due to the extent of infection, thus minimizing the effect of lighting inconsistencies within the image plane (Figure 4b).

Principal component analysis (PCA)

Principal component analysis (PCA) is a widely used statistical technique for data reduction (He et al., 2007; Liu and He, 2008; Liu et al., 2008). Using inbuilt ENVI subroutine PC.Rotate, PCA was performed on the full-spectrum (950 to 1650 nm) hyper-spectral images. Image background was excluded from these calculations using a binary mask image created for each hypercube to separate seedlings from the background as earlier explained. For each rice seedling, mean values of the first 5 PCA scores (mpc1 to mpc5) and the corresponding standard deviation values (sdpc1 to sdpc5) were extracted for further analysis. The ENVI subroutine ENVI.STATS.DOIT was used to compute these image statistics. A macro was written in IDL software (Version 7.0.2; ITT Visual Information Solutions, Denver, CO, USA) to automate the process of PCA calculations and extraction of mean and standard deviation values of PCA scores for each rice seedling in all the images in a batch mode.

Based on the loadings (Eigenvectors) of the first five principal components (PC), a set of five significant wavelengths were selected maintaining the overall behaviors of all first five principal components. PCA scores of the images were recalculated with the
five selected wavelengths. Similarly, mean and standard deviation values of PCA scores were extracted using an IDL macro. The purpose of the second PC_ROTATE was not for data reduction, but was done to generate orthogonal components and to compute within seedlings variations in the rotated planes so that the same set of features as for the full-spectrum could be used for seedlings classification. The PCA score measurements from the full-spectrum and selected wavelengths were used to develop and compare classification models.

**Classification**

The PCA scores data along with the target values were imported into the SAS software (version 9.2, SAS Inc., USA) to develop classification models. The variables with significant contribution towards desired classification were selected from the PCA score measurements of the calibration data set using stepwise discriminant analysis.

Linear discriminant analysis (LDA) classifier models were developed based on the selected variables using discriminate analysis. Data from the calibration sample set was used to develop the LDA classifier model using the full cross-validation (leave one out) option, while data from the validation sample set was used for performance evaluation of the models. The following describes the analysis process:

(a) PCA was performed on the full-spectrum (950 to 1650 nm) hyper-spectral images using inbuilt ENVI subroutine PC ROTATE for all of samples.
(b) Mean values of the first 5 PCA scores images (mpc1 to mpc5) and the corresponding standard deviation values (sdpc1 to sdpc5) were extracted using inbuilt ENVI subroutine ENVI_STATS_DOIT.
(c) Stepwise discriminant analysis was used to select variables with significant contribution towards desired classification.
(d) Linear discriminant analysis (LDA) classifier models were developed based on the above selected variables.
(e) Five significant wavelengths were selected based on the loadings of the first 5 PCA scores images and PCA was performed again. The statistics of PCA score images were also calculated again.
(f) Similarly, the stepwise discriminant analysis was performed again and the linear discriminant analysis (LDA) classifier models were developed based on the selected variables again.
(g) The classifier models based on full-spectrum and selected wavelengths were compared.

**RESULTS**

**Visual and spectral characteristics**

False color representation of hyper-spectral images of healthy seedlings and infected seedlings exhibited observable visual differences (Figure 3c...
The infected region appeared red than the healthy region. Representative spectra of healthy seedlings and infected seedlings as well as the image background extracted from one of the hyper-spectral images are shown in Figure 2. Each spectrum shown is the average of all pixel spectra in the respective category. In contrast to the spectral response of the seedlings, the image background exhibited higher response in the 1400 to 1500 nm range of the spectrum. Based on these spectral differences, the image background could be separated from seedlings using a simple threshold value in a single image band in the 1400 to 1500 nm range as shown in Figure 2, which is produced from the image band at 1466nm using a threshold value of 2300.

A mask produced could be used to extract and analyze seedlings spectra excluding image background from analyses. Figure 4b shows the normalized spectra of the seedlings eliminating intensity differences caused by illumination inconsistencies within the image plane. Healthy and infected seedlings exhibited differently shaped spectral characteristics. The spectral characteristics are consistent with literature report (Curran, 1989). Especially in the 1200 and 1400 nm band, there are obviously two absorption regions due to the O-H bending and extension. Healthy seedlings exhibited deep absorption in the 1450 nm band, and this is due to the strong absorption by water in the healthy organization.

While for the disease leaves, the collapse of the cells result in water loss, and infected region produce browning. Therefore, the low absorption will appear in the corresponding bands. Moreover, in the 1200 nm band, the different of the absorption performance is not obvious. Probably the reason is that absorb waveform stack by many biochemical substances (cellulose, starch, lignin and water) made absorption characteristics not obvious.

**Principal component analysis (PCA)**

The first 10 principal components (PC1 to PC10) explained over 95.78% of the variance in hyper-spectral image data, with the first five components explaining more than 95.07% of the variance in the data. Contribution of PC6 to PC10 combined was approximately 0.71%. The first five principal component scores of the healthy rice seedlings (Figure 3a) and infected seedlings (Figure 3d) are shown in Figure 5a and b. These score images contain information related to classification; however, none of the principal components alone contained sufficient information to fully segregate healthy rice seedlings and infected rice seedlings. Accurate classification required a combination of PC score features. The stepwise discriminant analysis selected a set of seven significant features for the classification base on full spectrum (Table 1). Two mean score features (mpc1 and mpc2) and five standard deviation features (sdpc1 to sdpc5) were selec-
Table 1. Feature selection for classification into healthy rice seedlings and infected rice seedlings categories based on full-spectrum (950 to 1650 nm).

<table>
<thead>
<tr>
<th>Variable in the model</th>
<th>Partial R2</th>
<th>P-Value</th>
<th>Wilk’s lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdpc2</td>
<td>0.3286</td>
<td>&lt;.0001</td>
<td>0.6713</td>
</tr>
<tr>
<td>mpc1</td>
<td>0.4225</td>
<td>&lt;.0001</td>
<td>0.3877</td>
</tr>
<tr>
<td>sdpc1</td>
<td>0.2462</td>
<td>&lt;.0001</td>
<td>0.2922</td>
</tr>
<tr>
<td>sdpc5</td>
<td>0.0917</td>
<td>0.0002</td>
<td>0.2654</td>
</tr>
<tr>
<td>sdpc3</td>
<td>0.0399</td>
<td>0.0161</td>
<td>0.2548</td>
</tr>
<tr>
<td>sdpc4</td>
<td>0.0396</td>
<td>0.0168</td>
<td>0.2447</td>
</tr>
<tr>
<td>mpc2</td>
<td>0.0249</td>
<td>0.0600</td>
<td>0.2387</td>
</tr>
</tbody>
</table>

Table 2. Feature selections for classification into healthy rice seedlings and infected rice seedlings categories based on five selected wavebands (1188, 1339, 1377, 1432 and 1614 nm).

<table>
<thead>
<tr>
<th>Variable in the model</th>
<th>Partial R2</th>
<th>P-Value</th>
<th>Wilk’s lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>mpc1</td>
<td>0.3786</td>
<td>&lt;.0001</td>
<td>0.6214</td>
</tr>
<tr>
<td>sdpc5</td>
<td>0.1308</td>
<td>&lt;.0001</td>
<td>0.5401</td>
</tr>
<tr>
<td>mpc3</td>
<td>0.0449</td>
<td>0.0100</td>
<td>0.5159</td>
</tr>
<tr>
<td>sdpc4</td>
<td>0.0498</td>
<td>0.0068</td>
<td>0.4902</td>
</tr>
</tbody>
</table>

mpc1 to mpc5: Mean values of the 1st five principal component scores computed over a single rice seedling.
sdpc1 to sdpc5: Standard deviation values of the 1st 5 principal component scores computed over a single rice seedling.

Table 3. Infected and healthy classification results for an LDA model based on selected PCA score features extracted from full spectrum (950 to 1650 nm).

<table>
<thead>
<tr>
<th>Actual class</th>
<th>Number of rice seedlings classified into class</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healthy</td>
<td>Infected</td>
</tr>
<tr>
<td>Calibration set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>69</td>
<td>6</td>
</tr>
<tr>
<td>Infected</td>
<td>2</td>
<td>73</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>79</td>
</tr>
<tr>
<td>Validation set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>Infected</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>29</td>
</tr>
</tbody>
</table>

Mean score features (mpc1) contributed the most towards the target healthy and infected classification, followed by the standard deviation of PC2 (sdpc2) and standard deviation of PC1 (sdpc1). Meanwhile, the 4 significant features for the classification base on 5 selected wavebands (1188, 1339, 1377, 1432 and 1614 nm) were also selected (Table 2).

**PCA based classification**

Classification results for healthy and infected categories based on PCA scores computed from the entire spectrum had an overall accuracy of 94.65 and 92% for the calibration and validation data sets, respectively (Table 3). Class-by-class classification accuracies for healthy and infected were 84 and 100%, respectively on the validation set. In addition, false positives and false negatives for the infected category were 16 and 0%, respectively on the validation set. As would be expected, classification accuracy was higher for the calibration data set than for the independent validation set. The behavior of the first five loading functions could be approximated by a set of five wavelengths or wavebands (Figure 6). Using the five selected wavebands (1188, 1339, 1377, 1432 and 1614 nm), the overall classification accuracy for healthy
Figure 6. Loadings (Eigenvectors) of the first five principal components (PC1 to PC5). Wavelengths marked with circles approximate the behaviors of the five loading functions.

Table 4. Infected and healthy classification results for an LDA model based on selected PCA score features extracted from 5 selected wavebands (1188, 1339, 1377, 1432 and 1614 nm).

<table>
<thead>
<tr>
<th>Actual class</th>
<th>Number of rice seedlings classified into class</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healthy</td>
<td>Infected</td>
</tr>
<tr>
<td>Calibration set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>healthy</td>
<td>66</td>
<td>9</td>
</tr>
<tr>
<td>infected</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>total</td>
<td>81</td>
<td>69</td>
</tr>
<tr>
<td>Validation set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>healthy</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>infected</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>total</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

and infected categories (Table 4) was 80% on the validation data set. Rice seedlings classification based on the five selected wavebands was comparable to that based on the full-spectrum (950 to 1650 nm). These results demonstrate that the infected rice seedlings can be detected with five selected wavebands. The fact that a few specific wavebands showed a potential for the detection of blast rice suggests that it may be possible to solve this problem using a low cost imaging system built around a monochrome digital camera and a set of optical filters in a motorized filter wheel. For industrial uptake, the lower cost of such a multi-spectral approach would be appealing. Such a system would initially have the potential to identify rice blast base on the canopy information imagery. Subsequent studies in this laboratory will continue to explore this possibility.

DISCUSSION

In this study, we try to capture the near-infrared hyperspectral image of rice blast disease on rice seedling, and built the classification model to indentify the rice blast disease. In order to compress the high-dimensional image information, the PCA analysis was employed. The statistics was calculated based on the first 5 PCA score
images and the stepwise discriminant analysis was developed. Based on the result of this research, the identification was realized only based on the statistics of first 5 PCA score images extracted from full-spectrum images. The changes of spectral characteristics caused by rice blast can be represented with the first 5 PCA score images. For classification model with image features calculated with first 5 PCA score images extracted from full spectrum (950 to 1650 nm), healthy and infected rice seedlings can be classified with an overall accuracy of 92% in the validation set.

Although, using the statistics of first 5 PCA score images can identify the rice blast disease effectively, but for prospects of practical applications, a few feature wavelengths contributed the most towards the target healthy and infected classification appears more appealing. Hence, we extracted a set of five wavebands and tried to use this image to build the simplified model. The result indicated that identification can be completed basically based on the information of five wavebands and with an overall accuracy of 80% in the validation set. This preliminary research results is encouraging and indicates that using the information of a few wavelengths to identify the rice blast disease is feasible.

ACKNOWLEDGEMENTS

This study was supported by Important Zhejiang Provincial Science and Technology Specific Projects (2009C12002), 863 National High-Tech Research and Development Plan (Project no: 2011AA100705), Natural Science Foundation of China (60802038), Zhejiang Provincial Natural Science Foundation of China (Project no: Z3090295) and National Agricultural Science and Technology Achievements Transformation Fund Programs (2011GB23600008).

REFERENCES


