DIAGNOSIS OF SCLEROTINIA INFECTED OILSEED RAPE (BRASSICA NAPUS L) USING HYPERSPECTRAL IMAGING AND CHEMOMTRICS

Na Chen[1], Fei Liu[1], Lulu Jiang[2], Lei Feng[1], Yong He[1], Yidan Bao[1]∗

1.College of Biosystems Engineering and Food Science
Zhejiang University
Hangzhou, Zhejiang Province, P.R. China

2. Zhejiang Technology Institute of Economy
Hangzhou, Zhejiang Province, P.R. China

Abstract: Brassica napus L leaf diseases could cause seriously reduction in crop yield and quality. Early diagnosis of Brassica napus L leaf diseases plays a vital role in Brassica napus L growth. To explore an effective methodology for diagnosis of Sclerotinia infected Brassica napus L plants, healthy Brassica napus L leaves and Brassica napus L leaves infected by Sclerotinia were prepared in a controlled circumstance. A visible/short-wave near infrared hyperspectral imaging system covering the spectral range 380-1030 nm was set up to identify healthy and infected Brassica napus L leaves. The clear and non-deformable hyperspectral images were captured and the spectral information was exacted from the hyperspectral images according to the predefined region of interest (ROI). The spectral ranges of 380-439 nm and 951-1030 nm which contained obvious noises were removed. Moving average was used as the pretreatment method for spectra to remove noises. Chemometric methods were applied to build classification models for healthy and infected Brassica napus L leaves identification, including Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Support Vector Machine (SVM) models. Weighted regression coefficient (*Bw*) was applied to select sensitive wavelengths, and PLS-DA model and SVM model were also built on the basis of the sensitive wavelengths for healthy and infected leaves identification. For the infected leaves, the healthy part, the infected part and the joint part of the healthy part and the disease part were analyzed separately to investigate which part was most efficient for disease identification by PLS-DA, and the results showed that the infected part obtained better identification results than the other parts. For both full spectra and the sensitive wavelengths, PLS models and SVM models showed good performances with identification rate over 85%, and PLS models using the spectra extracted from the infected part obtained the identification rate over 90%. The classification models using sensitive wavelengths showed similar or better performances compared with the classification models using full spectra, indicating that selected sensitive wavelengths could be used for Brassica napus L disease diagnosis with fewer

input variables. The overall results indicated that Brassica napus L leaf diseases could be early diagnosed by hyperspectral imaging and multivariate techniques effectively, which would be helpful for the prevention and treatment of Brassica napus L diseases. The results were gained in laboratory, and to obtain more accurate and practical results, the hyperspectral imaging system for field application should be developed, and more experiments should be conducted to select more accurate and stable sensitive wavelengths and build more robust identification models.

Key Words: Brassica napus L leaves, Sclerotinia Sclerotiorum, Hyperspectral Imaging, PLS-DA, SVM

INTRODUCTION

Brassica napus L is easily affected by various external factors during growth, such as nutrition, plant diseases and insect pests, etc. Disease is one of the biggest threats to Brassica napus L. Sclerotinia sclerotiorum is a soil spread fungus disease which is difficultly prevented and controlled in the open air. It happens in seeding stage and maturation stage. It breaks out under the condition of proper temperature, humid and rainy climate. It has an adverse impact on stems, leaves, flowers and fruits, This disease serious influences the quality and yield of Brassica napus L. In general, the yield can reduce by 10% to 70%, even 80% or more in serious situation. Unfortunately, there is no treatment or prevention that could completely eradicate sclerotinia sclerotiorum in field. The current emphasis is on minimizing disease influence and avoiding further infections. External and internal characteristics of oilseed rape leaves are changing after infected by diseases, and leaf spectral response characteristics also change. The plant disease diagnosis methods include human eye recognition, computer vision technique, polymerase chain reaction (PCR) technique, gene chip technology, serological technique and microscope identification technology, etc. (Pertot I et al., 2012; Park J et al., 2013; Yang W J et al., 2013; Bananej K et al., 2013; Ghosh D et al., 2013). Eye recognition method is the most commonly used methods, but this method has a strong subjectivity. Other methods could diagnosis plant diseases accurately, but these technologies require plenty of time and cost. Also, scientific instruments need professional operations. Clearly, a rapid and convenient recognition technology would be most useful for the real-time and on-line farmland disease detection. Spectroscopy and imaging techniques were applied to detect plant disease. Many studies have been reported for the detection of pant diseases using spectroscopy and spectral imaging techniques, which confirmed the feasibility of using spectral responses of infected plants to detect the diseases (Borràs E et al. 2014; Lucio-Gutiérrez J R et al., 2011; Sankaran S and Ehsani R, 2011; Sankaran S et al., 2011).

Hyperspectral imaging which combines spectroscopy and imaging

techniques have been applied in various fields (Wu D and Sun D W, 2013; Barbin D et al., 2012; Taghizadeh M et al., 2011; Talens P et al., 2013). Hyperspectral imaging acquires spectral and image data simultaneously, providing more detail information of the samples. Also, the spectrum of each pixel could be obtained by hyperspectral imaging technique. Some studies have been reported for using hyperspectral imaging to detect plant diseases (Qin J et al. 2009; Kong W et al. 2014).

In this study, our objective was to investigate the potential of hyperepstral imaging as a rapid, non-destructive, and accurate technique to diagnose Brassica napus L diseases (Müller K et al., 2008). Specific objectives of this study were as follows: 1) to analyze the spectral features of healthy and infected Brassica napus L leaves; 2) to build identification models for healthy and infected Brassica napus L leaves; 3) to select sensitive wavelengths for detection of sclerotinia sclerotiorum infected Brassica napus L; 4) to explore the influence of the healthy part, the infected part and the part between the healthy part and the disease part of the infected leaves on the diagnosis of sclerotinia sclerotiorum infected Brassica napus L.

MATERIAL AND METHODS

Sample Preparation

Sclerotinia sclerotiorum pathogen was provided by the Institute of Biotechnology, Zhejiang University. Brassica napus L were inoculated by mycelium inoculation method. Totally 60 plants were used, including 30 healthy plants and 30 plants inoculated by sclerotinia sclerotiorum disease pathogen. Plants were placed in a controlled environment for 48 hours after inoculation, with humidity of 85% and the temperature from 20 to 25℃. After the slight symptom emerging, 42 healthy leaves and 42 infected leaves were collected to an incubator for hyperspectral image acquisition. Fig. 1 shows the pseudocolor images(R: 662nm, G: 554nm, B: 450nm) of healthy and infected leaves.

Healthy leaf Infected leaf

Fig. 1 Figures of healthy and infected oilseed rape leaves.

Hyperspectral imaging system

A laboratory NIR/Vis hyperspectral imaging system was used to acquire spectral images of the samples in the reflectance mode. The system included a line-scan spectrograph (InSpectorV10E-QE, Spectral Imaging Ltd., Finland), a CCD camera with C-mount lens (OLES 23, Spectral Imaging Ltd., Oulu, Finland), the illumination unit which consisted of two 150 W tungsten halogen

lamps (Fiber-Lite DC950 Illuminator, Dolan Jenner Industries Inc., USA), a conveyer belt (Isuzu Optics Corp., Taiwan, China) and a data acquisition software (Spectral Image-V10E, Isuzu Optics Corp., Taiwan, China). The camera has 672×512 (spatial×spectral) pixels. The vertical distance between sample and lens was set to 360 mm and the speed of the conveyer belt was adjusted at 2.05 mm/s to avoid image size distortion. Also, the exposure time of camera was set to 0.05 s to achieve a square pixel. The spectral range obtained from this system was 380–1030 nm with 512 spectral bands. The schematic diagram of main components of the hyperspectral imaging system is illustrated in Fig. 2.

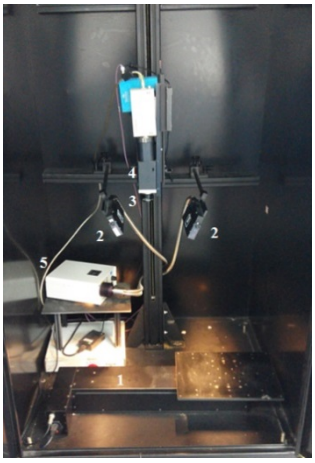

Fig. 2 The schematic diagram of main components of the hyperspectral imaging system included: (1) conveyer belt; (2) illumination unit; (3) lens; (4) spectral camera; (5) motor.

Image acquisition and calibration

To acquire three-dimensional hyperspectral image information, each sample was placed on a black background which had very low reflectance. The conveyer belt transferred the sample to the field of view of camera to be scanned one by one. Each hyperspectral image which was consisted of 512 contiguous spectral bands was regarded as a 3-D hypercube data. The first two dimensions represent spatial coordinates, and the third one is for spectral intensities. The acquired images were stored as raw format.

To calibrate the raw image obtained from the hyperspectral imaging system, dark reference (D) and white reference (W) images were acquired (Zhu F et al., 2014). A dark reflectance (D) image near to 0% reflectance was obtained by turning off the light source and covering the camera lens to measure the dark current of camera, and a white reference (W) image approach to 99.9% reflectance. To calculate a relative reflectance image (I) of the sample, the acquired images (R) from the hyperspectral imaging system were corrected by using the following equation:

$$I = \frac{R - D}{W - D} \tag{1}$$

The Environment for Visualizing Images (ENVI) 4.7 software(ITT Visual Information Solutions, Boulder, USA)was used for image correction and following processing. All the corrected images were then used as the basis for subsequent analysis to extract spectral information, important wavelength selection and calibration.

Spectral data extraction

After image calibration, the region of interest (ROI) of each hyspectral image was defined and the spectral data was extracted from the ROI. The average spectrum of all pixels within the ROI was used as the spectrum of the sample. In this study, the entire leaf region was defined as ROI image for the discrimination of healthy and infected samples.

In order to evaluate the best parts within the infected leaves for diagnosis of disease, the healthy part, the symptomatic part and the joint part between the healthy part and the infected part on the same infected leaf sample were selected as ROIs, and the average spectrum of each part was extracted.

Multivariate Techniques

Principal component analysis (PCA) is a statistical procedure that uses liner transformation to transform the original data into a set of new variables called principal components (PCs) (Gao J et al., 2013; Sankaran S et al., 2012). The PCs are linear combinations of the original variables. And the first few PCs which contain the most information of the original data have the largest possible variances. Usually the scores scatter plot of the first two PCs was implemented to reveal the features of the samples.

Partial Least Squares Discriminant Analysis (PLS-DA) is a supervised pattern recognition method based on the PLS regression (PLSR) (Wang S et al., 2012). A set of dummy integers which represent sample categories are used as Y instead of chemical constituent. The outputs of PLS-DA are real numbers but not dummy integers. Thus, a threshold value needs to be set to determine which class the sample belongs to. In this study, the cutoff value of 0.5 was used. If the absolute difference value between the predicted value and actual value was less than 0.5, the discrimination result was considered to be correct, or the discrimination result was considered to be error.

Support Vector Machine (SVM) is a supervised pattern recognition technique with good generalization ability (Rumpf T et al., 2010; Zang H et al., 2012). SVM shows advantages in the situation of small samples, non-linear and high-dimensional data. SVM could also be used for regression (Devos O et al., 2014). Kernel function was quite important in SVM analysis and radial basis function (RBF) was used as kernel function of SVM in this study.

Effective wavelengths selection

In this study, weighted regression coefficients (Bw) resulting from the PLS-DA model were used to select effective wavelengths (Kamruzzaman M et al., 2013). The wavelengths that corresponded to the highest absolute values of weighted regression coefficients were selected as effective wavelengths. In general, the threshold value of absolute values of weighted regression coefficients was defined to select effective wavelengths.

RESULTS AND ANALYSIS

Spectral features of healthy and infected leaves

Raw visible/near infrared spectra were extracted from the healthy and infected oilseed rape leaves by ENVI software and were preprocessed by moving average with 7 points. The spectral range of 380-439 nm and 951-1030 nm were removed due to absolute noises, and 440-950 nm spectral data were chosen to analyze. The averaged preprocessed spectra of healthy sample and infected sample were shown in Fig. 3.

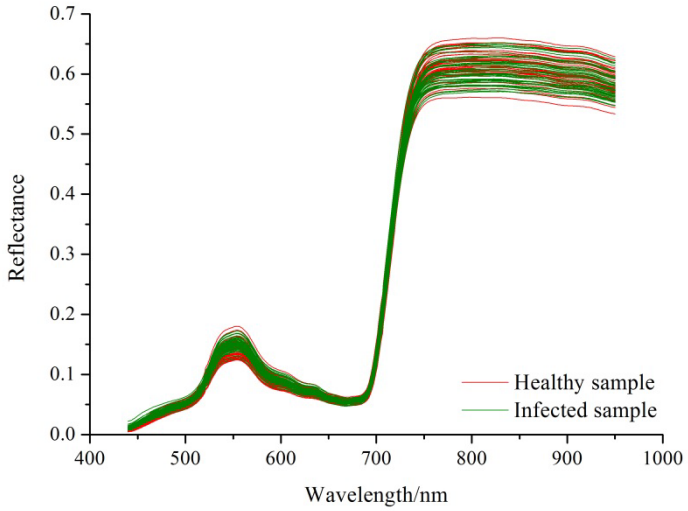

Fig.3. The reflectance spectra curves of healthy samples and infected samples.

Preliminary investigation by PCA

According to spectral features of the tested samples, PCA was used to interpret and visualize the spectral data to highlight their properties, similarities and differences by identifying the most important directions. In essence, when the scores of the principal components were plotted, samples with similar spectral signatures tended to aggregate together. Generally, the first few principal components resulting from PCA were used to examine the natural pattern among samples and their clustering. The scores plot of the first two PCs (explained 98% variation) were shown in Fig. 4 and revealed the clear clustering between healthy leaf and infected leaf. As shown in Fig.4, two different clusters were clearly observed that some overlaps were among them. These results implied the possibility of analytical method which was based on hyperspectral imaging for discrimination of infected leaves and healthy leaves, and the chemometric methods should be applied to more accurately identify the healthy and the infected leaves.

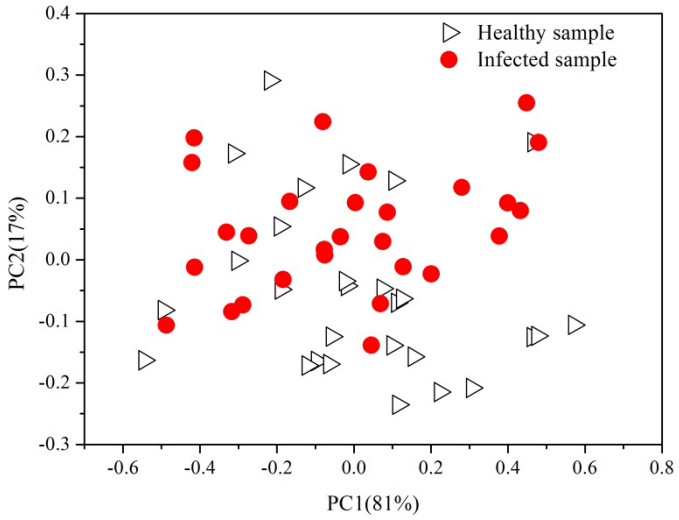

Fig.4. Scores plot of the first and second principal components of PCA at spectral range of 440–950 nm for healthy sample and infected sample.

Sample spilt

28 infected leave samples and 28 healthy leave samples were randomly chosen for calibration set, and the rest were used for prediction set. The assignment of different varieties of oilseed rape leaves and the division of calibration and prediction sets samples were shown in Table 1.

Table1. Sample split and assignment value of healthy and infected samples

Sample types	Total number	Assignment value	Calibration set	Prediction set
[a]H	42	1	28	14
[b]I	42	2	28	14

a: H refers to healthy leaves; b: refers to infected leaves, and the same in the entire paper

Discrimination models based on full spectra

In order to discriminate the healthy sample leaves and the infected sample leaves, PLS-DA model and SVM model were established based on the full spectra data. The latent variables of PLS-DA model was 13, the parameters of SVM model was c = 256 and g = 0.0359(Zheng H and Lu H, 2012).

The results of the calibration set recognition accuracy and the prediction set recognition accuracy based on different discriminant models were shown in Table 2.

Table 2.Results of different classification models using full spectra

Methods	Calibration set			Prediction set		
	H	I	Total	H	I	Total
PLS-DA	100%	100%	100%	85.71%	100%	92.86%
SVM	100%	100%	100%	71.43%	100%	85.72%

It could be seen from Table 2 that the recognition accuracies of calibration sets in PLS-DA model and SVM model were both reached to 100%

for healthy and infected leaves. The total recognition accuracies were over 80% for the prediction sets in PLS-DA model and SVM model. However, it could be noticed that the recognition results of healthy samples in the prediction sets were much worse than the infected samples. On the other hand, the performance of the PLS-DA model was much better than SVM model. The results showed that discrimination models could accurately distinguish the healthy leaves and the infected leaves, indicating that hyperspectral imaging combined with discrimination models could be used for detection of Sclerotinia Sclerotiorum infected oilseed rape leaves.

Effective wavelengths selection

It is essential to select effective wavelengths to eliminate the redundant information in hyperspectral images and develop an optimized discrimination model for diagnosis infected leaves and healthy leaves. In this study, effective wavelengths selection was to pick out the wavelengths that contained the most information for diagnosis of Sclerotinia Sclerotiorum infected oilseed rape leaves. According to the large absolute values of weighted regression coefficients (*Bw*) resulting from PLS-DA model, eight effective wavelengths were selected (458, 490, 528, 666, 743, 787, 837 and 901 nm) as shown in Fig.5. PLS-DA model and SVM model were then established using these important wavelengths to predict plant disease infection.

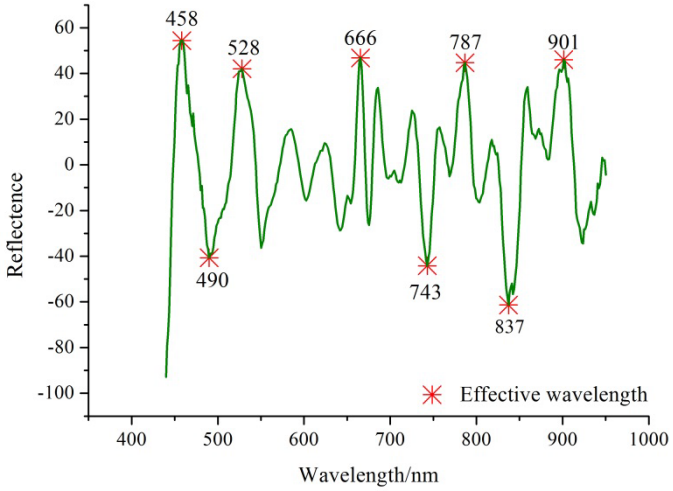

Fig. 5 Effective wavelengths selection by *Bw*

Discrimination models based on sensitive wavelengths

After effective wavelengths selection, the effective wavelengths were used as inputs of PLS-DA and SVM. The results were shown in Table 3.

Table 3.Results of different classification models using effective wavelengths

Methods	Calibration set			Prediction set		
	H	I	Total	H	I	Total
PLS-DA	100%	89.29%	94.65%	92.86%	100%	96.43%
SVM	100%	100%	100%	78.57%	100%	89.29%

Seen from Table 3, PLS-DA model obtained satisfactory results in both

calibration set and the prediction set, the recognition accuracies were over 90%. And for healthy and infected leaves in calibration set and prediction set, the recognition accuracies were all over 85%.The recognition accuracy for calibration set in SVM model (100%) was much higher than prediction set, and the recognition accuracies of calibration set and prediction set were over 85%. However, the recognition accuracy of healthy leaves in prediction set was 78.57%, which needed to be improved. The performances of PLS-DA model and SVM model were similar and satisfactory, indicating that effective wavelengths could be used for Sclerotinia Sclerotiorum infected oilseed rape disease detection.

From Table 2 and Table 3, it could be noticed that the performances of the different discrimination models based on full spectra and effective wavelengths were similar, and PLS-DA models performed a bit better for the reason that the recognition accuracies of healthy and infected leaves were all over 85%. The overall results indicated that effective wavelengths could be used to build discrimination models for Sclerotinia Sclerotiorum infected oilseed rape detection instead of full spectra.

Evaluation of the best efficient part in the infected leaves for disease detection

The above analysis has proved the feasibility of using hyperspectral imaging combined with discrimination methods to detect Sclerotinia infected oilseed rape, and the ROI was defined as the sample region of the entire leaf. It was noted that the infected leaves contained 3 parts: the healthy part, the infected part with symptoms, and the joint part of the healthy part and the infected part. To evaluate the influences of different parts in the infected leaves on the recognition of Sclerotinia infected oilseed rape leaves, the ROIs were defined in these three parts and the spectral data were extracted, and 42 samples were obtained for each part. PLS-DA was applied to build discrimination modes using these three parts and the previous healthy leaves, respectively.

For each part, the healthy samples and the samples of special part were randomly selected into calibration and prediction parts at the ratio of 2:1. The results of PLS-DA models were shown in Table 4.

Table 4.The results of PLS-DA models for different parts of infected leaves and healthy leaves

Methods	Calibration set			Prediction set		
	H	I	Total	H	I	Total
H-[d]A	100%	100%	100%	100%	82.14%	91.07%
H-[e]B	100%	100%	100%	100%	100%	100%
H-[f]C	100%	100%	100%	100%	100%	100%

d refers to the healthy part in the infected leaves; e refers to the infected part of the infected leaves; f refers to the joint part of the healthy part and the infected part

As shown in Table 4, the discriminant accuracies of healthy leaves and the healthy part of infected leaves could reach to 90%. The reason might be

that although the healthy part in infected leaves did not present obvious symptoms, it was influenced by the infection and spectral reflectance changed. Moreover, the PLS-DA models for the remaining two parts and the healthy leaves obtained recognition accuracies for calibration set and prediction set were 100%. The overall results also indicated that Sclerotinia infected oilseed rape could be detected by hyeprspectral imaging. Therefore, the infected part in the ROI could be better for discrimination, and the healthy part in infected leaves could not be used as healthy samples for analysis.

CONCLUSION

The diagnosis of sclerotinia infected oilseed rape is attracting an increasing amount of attention for gardeners, farmers and scientists during the plant growing process. It is an important issue in the scope of traceability, yield and quality control of oilseed rape. This study applied hyperspectral imaging technique for the identification of infected leaf and healthy leaf sample. The present study evaluated the applicability of using features extracted from 440 to 950 nm spectral reflectance data for detecting oilseed rape in leaves. PLS-DA models using full spectra and effective spectra selected by *Bw* got better results than SVM models. PLS-DA model using effective spectra selected by *Bw* obtained similar results to SVM model using effective wavelengths. The recognition of infected leaf samples was better than the healthy leaf samples. All discrimination models based on full spectra and effective wavelengths obtained recognition accuracies over 80% for calibration set and prediction set. To evaluate which part of the infected leaves could be more effective for disease detection, the healthy part, the infected part with symptoms, and the joint part of the healthy part and the infected part of the infected leaves were used to build PLS-DA models with the healthy leaves. The parts containing the infected region showed recognition accuracies of 100% and performance of the PLS-DA model for the healthy part was worse. The overall results indicated that hyperspectral imaging could be used to detect Sclerotinia infected oilseed rape combined with chemometric methods, and the ROI which contained the infected region of the leaf would help to obtain better results.

ACKNOWLEDGMENTS

This research was supported by 863 National High-Tech Research and Development Plan (2013AA102405), Natural Science Foundation of China (31201137) and Fundamental Research Funds for the Central Universities (2014FZA6005).

REFERENCE

[1] Bananej K, Kianfar N, Vahdat A, et al. 2013. J PHYTOPATHOL. 162(3):p. 205-208.
[2] Barbin D, Elmasry G, Sun D W, et al. 2012. MEAT SCI. 90(1):p. 259-268.
[3] Borràs E, Amigo J M, van den Berg F, et al. 2014. FOOD CHEM. 153:p. 15-19.
[4] Devos O, Downey G, Duponchel L. 2014. FOOD CHEM. 148:p. 124-130.

[5] Gao J, Li X, Zhu F, et al. 2013. COMPUT ELECTRON AGR.99:p. 186-193.
[6] Ghosh D, Bhose S, Manimekalai R, et al. 2013. J PLANT BIOCHEM BIOT. 22(3): p. 343-347.
[7] Kamruzzaman M, Sun D W, ElMasry G, et al. 2013. TALANTA. 103: p. 130-136.
[8] Kong W, Liu F, Zhang C, et al. 2014. SPECTROCHIM ACTA A. 118:p. 498-502.
[9] Lucio-Gutiérrez J R, Coello J, Maspoch S. 2011. FOOD RES INT. 44(2):p. 557-565.
[10] Müller K, Böttcher U, Meyer-Schatz F, et al. 2008. BIOSYST ENG. 101(2): p. 172-182.
[11] Park J, Kil E J, Kim J, et al. 2013. J PHYTOPATHOL. 162(4):p. 209-217.
[12] Pertot I, Kuflik T, Gordon I, et al. 2012. COMPUT ELECTRON AGR. 84:p. 144-154.
[13] Qin J, Burks T F, Ritenour M A, et al. 2009. J FOOD ENG. 93(2):p. 183-191.
[14] Rumpf T, Mahlein A K, Steiner U, et al. 2010. COMPUT ELECTRON AGR. 74(1): p. 91-99.
[15] Sankaran S, Ehsani R, Inch S A, et al. 2012. PLANT DIS. 96(11):p. 1683-1689.
[16] Sankaran S, Ehsani R. 2011. CROP PROT. 30(11):p. 1508-1513.
[17] Sankaran S, Mishra A, Maja J M, et al. 2011. COMPUT ELECTRON AGR. 77(2):p. 127-134.
[18] Taghizadeh M, Gowen A A, O'Donnell C P. 2011. BIOSYST ENG. 108(2):p. 191-194.
[19] Talens P, Mora L, Morsy N, et al. 2013. J FOOD ENG. 117(3): p. 272-280.
[20] Wang S, Huang M, Zhu Q. 2012. COMPUT ELECTRON AGR. 80: p.1–7.
[21] Wu D, Sun D W. 2013. INNOV FOOD SCI EMERG. 19:p. 15-28.
[22] Yang W J, Chen J H, Chen G N, et al. 2013. BIOSENS BIOELECTRON. 41:p. 820-826.
[23] Zang H, Li L, Wang F, et al. 2012. J PHARMACEUT BIOMED. 61:p. 224-229.
[24] Zheng H, Lu H. 2012. COMPUT ELECTRON AGR. 83: p. 47-51.
[25] Zhu F, Zhang H, Shao Y, et al. 2014. FOOD BIOPROCESS TECH. 7:p. 1208-1214.