A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments

Yu-Hui F. Yeh*. Wei-Chang Chung*. Jui-Yu Liao.**
Chia-Lin Chung. Yan-Fu Kuo*. Ta-Te Lin***

** Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (Tel: 886-2-3366-5331; e-mail: m456@ ntu.edu.tw).*
*** Department of Plant Pathology and Microbiology, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: clchung@ntu.edu.tw)*

Abstract: As plant diseases could cause agricultural production and economic loses, there is a need of fast and effective plant disease detection and assessment methods. Non-destructive methods have gained popularity among these methods as they do not affect plant growth while examining plant health conditions. Not only plant diseases can be detected but also production can be improved with proper quality controls. Hyperspectral imaging is one of the non-destructive examination techniques which have been widely applied in agriculture. Hyperspectral image analysis has been applied to different problems including plant disease detection and assessments. It provides not only spatial image but also spectral information of the observed object. This research has aimed to compare two hyperspectral image analysis methods: stepwise discriminant analysis (SDA) and spectral angle mapper (SAM) and the proposed Simple Slope Measure (SSM) method in strawberry foliage Anthracnose assessment. Anthracnose is one of the most devastating diseases for strawberries. Anthracnose disease can affect the whole plant and may result in 100 percent fruit loss from crown and fruit rot. Hence, an early detection of the Anthracnose disease will be beneficial to ensure production and quality of strawberries. This research has shown that the three different Anthracnose infection status (healthy, incubation and symptomatic) could be separated by the methods examined. The performance of these disease assessment models were evaluated and compared. The examination outcomes prove the feasibility to assess strawberry foliage Anthracnose nondestructively and as early as the symptoms not visible to naked eyes. As soon as early detections of the Anthracnose disease are achievable in the strawberry field, the damage to strawberry production due to the spread of Anthracnose disease could be reduced.

Keywords: Hyperspectral Imaging, Plant Diseases, Machine Learning Algorithms.

1. INTRODUCTION

The agricultural production and economic loses across the world can be affected critically by both physiological and infectious plant diseases. In particular, plant diseases in grains and crops may cause food insecurity and famines. Common sources of infection include insects and microorganisms like bacteria, funguses and viruses. Hence, plant disease predetermination and prevention has raised great interests in research, especially in real-time and non-destructive techniques.

Hyperspectral image analysis has been widely applied to geology and mining applications and other problems such as food safety control and surveillance. Plant disease detections and assessments are also popular research domains in hyperspectral image application field (Coops *et al.*, 2003; Mahlein *et al.*, 2012; Qin *et al.*, 2012). Hyperspectral images provide both spatial image and spectral information of the observed object. These information is much resourceful than other imaging techniques or spectral examination methods.

Anthracnose is one of the most devastating diseases for strawberries. Anthracnose is an all-year-round fungal disease spreading through air and rain-splash. This disease may result 100 percent fruit loss from crown and fruit rot (Curry *et al.*, 2002; Ellis and Erincik, 2008). The strawberry diseases were diagnosed through destructive procedures such as DNA extraction and fluorescence imaging (Sreenivasaprasad *et al.*, 1996; Vargas *et al.*, 2004). If non-destructive strawberry Anthracnose detection methods are available, strawberry production and quality can be ensured.

This research aims to assess strawberry foliage Anthracnose disease by hyperspectral imaging. Two hyperspectral image analysis methods, spectral angle mapper (SAM), stepwise discriminant analysis (SDA) and a simple slope measure (SSM) method were applied to evaluate the hyperspectral imaging assessment accuracy and efficiency. The results show the feasibility of hyperspectral imaging not only in strawberry foliage Anthracnose diagnosis but also early detection.

2. MATERIALS AND METHODS

2.1 Strawberry Foliage and Anthracnose Pathogen

In this research, the most widely cultivated strawberry in Taiwan, Taoyuan No. 3, was selected to be inoculated with Colletotrichum gloeosporioides (Gc001) to simulate the strawberry foliage Anthracnose. Taoyuan No. 3 is featured with high resistance to Anthracnose, Powdery mildew, Phytophthora fruit rot and two-spotted spider mites.

Healthy mother plants were grown in a 1: 1: 1 (v/v/v) mixture of peat: perlite: vermiculite, at 60~70% relative humidity (RH) under a photoperiod of 16 hrs. These plants were watered/fertilized everyday with 0.3% HYPONeX No.2 (20-20-20, Hyponex, U.S.A.). Daughter plants from the stolons of the mother plants were initially cultivated in 9-cm diameter plastic pots. The plantlets were then separated from the mother plants and transferred to a chamber at 25◦C, 60% RH under a photoperiod of 8 hrs. After 12-13 more days, every plantlet (~25 day-old) would have 4-5 leaves. Detached strawberry leaves from the second or the third leaf positions of the plantlets were placed in the sterile 15-cm diameter petri dishes with 3 layers of sterilized paper towels and 10 ml of sterile distilled H_2O. Every detached strawberry leaf was immobilized using a specialized black mesh.

An isolate of Colletotrichum gloeosporioides (Gc001), originated from an infected strawberry leaf collected in Miao-Li in 2011, was used for in vitro inoculation test. For inoculum preparation, Gc001 was cultured on potato dextrose agar (PDA) plates under a 12 hr/12 hr normal light-dark cycle for 7-10 days at room temperature. Spores of C. gloeosporioides were dislodged from the colonies with sterile distilled water containing 0.02% (v/v) Tween 20, quantified with a haemocytometer, and diluted to a concentration of 106 conidia mL-1 for subsequent infection trials. About 0.5 ml of spore suspension (106 conidia mL-1) was evenly sprayed on each detached leaf (fixed on the plate) with an airbrush at 10-15 psi. The plates were then sealed with parafilm to maintain the RH at > 90%. The inoculated detached leaves were kept at 25°C under an 8 hr-light/16 hr-dark cycle until investigation.

In this research, five inoculated samples and one control sample were scanned every 12 hours for five days after the one full day incubation period.

2.2 Hyperspectral Imaging System

A hyperspectral imaging system set in a dark cabinet consists of a Headwall HyperspecTM VNIR A-series camera, two linear halogen light source and a moving platform to eliminate external light interference. The spectral range of the hyperspectral camera is from 400 nm to 1000 nm with the spectral resolution set to be 4.6 nm resulting 126 bands in each hyperspectral image obtained. The two linear halogen light source was aligned to the same position on the movable platform ensuring the linear light is strong and evenly distributed onto the investigated sample. The movable platform is powered by a stepping motor which move the sample in steps for it to be line-scanned by the hyperspectral camera.

Before the hyperspectral image acquisition, reference and dark current are required to be obtained to calibrate the hyperspectral camera from camera interior and environmental biases. The reference is learnt from the reflection from the standard white board provided by the Headwall company, while the dark current is obtained through covering the camera with a cap.

2.3 Spectral Angle Mapper

The Spectral Angle Mapper (SAM) was first proposed in Kruse *et al.*, 1993. It is a standard, fast and simple method to calculate spectrum similarity for hyperspectral imaging applications (Girouard *et al.*, 2004). The angle between the reflectance spectra of the investigated object and the reference object is calculated to evaluate the similarity between them. The smaller the angle calculated, the more similar these two spectra are. Hence, the object spectrum will be assigned to the specific class which has the smallest angle value with the specific reference spectrum. The reference spectrum is sometimes referred as the endmember spectrum. In general, the reference spectrum collected from laboratory measurements, field studies and available spectral libraries.

Each reflectance spectrum is a combination of reflectance values across the whole wavelength range, *b*. The reflectance spectrum of the investigated subject (*t*) is compared with the reference spectrum of the target subject (*r*) to evaluate the angle similarity value:

$$\mathrm{S} = cos^{-1}\left(\frac{\sum_{i=1}^{b} t_i r_i}{\sqrt{\sum_{i=1}^{b} t_i^2}\sqrt{\sum_{i=1}^{b} r_i^2}}\right) \quad (1)$$

There are different extensions of SAM have been proposed such as the Spectral Correlation Mapper (Carvalho and Meneses, 2000), the Multiple-endmember Spectral Angle Mapper (Cho *et al.*, 2010) and the Unsupervised Spectral Angle Mapper (Sohn and Rebello, 2002).

2.4 Stepwise Discriminant Analysis

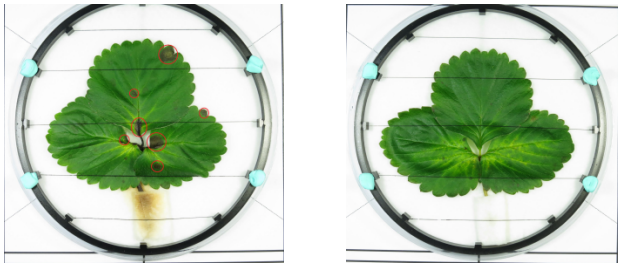
The Stepwise Discriminant Analysis (SDA) is a statistical method which has been widely applied to different fields for dimensionality reduction and model selection purposes. The SDA is based on the principle of multivariate F-test. In order to select significant variables among all variables available, the Wilk's lambda needs to be maximized. In other words, the sum of variances between categories need to be maximized while the within category variances are minimized.

In each variable selection iteration, variables are added or removed according to their contribution significance to the discriminant function given the variables included in previous steps. Any variable that is selected to be included but is fail to provide significant contribution to the discriminant function after the new variable addition will be removed from the

significant variable set. This selection process cycles until no more variable can provide contribution to the discrimination performance. The classification function for each category is formulated with the significant variables selected. The category value of each category will be evaluated for each investigated subject. The investigated subject will be assigned to the category whose category classification value is the biggest.

3. RESULTS AND DISCUSSIONS

The Anthracnose infection severities were quite different across five inoculated samples. At the end of the observation period, some samples had many disease spots developed into dark black areas, while few samples only showed tiny black spots (Fig. 1).



(a) (b)

Fig. 1. Anthracnose infection status is much critical in one sample (a) than in another sample (b).

The analysis in this research is broken down into two sections: two-stage diagnosis and three-stage assessment. In the two-stage diagnosis section, only the healthy and symptomatic areas were examined. The incubated areas were only discussed in the three-stage assessment section. Fifty 3x3 pixel blocks were selected to be examined for each class. All healthy areas were sampled randomly from the control sample across different days. As the infection severities across samples were not even, the symptomatic areas were chosen by the defined darkness level of the symptomatic spot. In other words, the number of symptomatic areas selected in each sample was not the same. This was the same case for the incubated areas. The incubated area was defined as the area which has no visible Anthracnose symptom but develops into black spot symptoms (visible symptoms) 24 hours later.

The reflectance spectra of selected areas were extracted from the hyperspectral images obtained via a self-developed program. As the camera signals at both end of the wavelength range are not stable, the wavelength range was limited to 460 nm to 930 nm for analysis to eliminate data noise. The reflectance of each 3x3 pixel block was the average reflectance of the 9 pixels in the block.

3.1 Anthracnose Assessment

Diagnosing strawberry leaves into two classes: healthy and symptomatic, is the first and simplest investigation of this research. The classification accuracy reached 95% via SAM in 5-fold cross validation as shown in Table 1. This performance is expected as the classification problem is quite simple. In addition, it is clear that the symptomatic areas were easier to be misclassified than the healthy areas.

Table 1. Strawberry Anthracnose Diagnosis with Spectral Angle Mapper

		Prediction	
		Healthy	Symptomatic
Infection Status	Healthy	100%	0%
	Symptomatic	10%	90%
Overall Accuracy		95%	

Although we see a promising result of SAM on VNIR reflectance, this research has another focus on early detection. In other words, plant leaves not only classified into healthy and symptomatic stages but also the stage before symptoms are visible to naked eyes for early detection. In the next section of analysis, three infection stages investigated: healthy, symptomatic and incubation (the early detection stage) stages.

The reflectance spectra of three infection stages were investigated and plotted as in Fig. 2. The figure shows the average spectrum of each stage and its one standard deviation span. It is clear in the figure that the incubation stage spectra were overlapping with the other two stages, especially the healthy ones. Hence, it is not possible to classify strawberry leaves into their correct infection stages just by comparing the spectra.

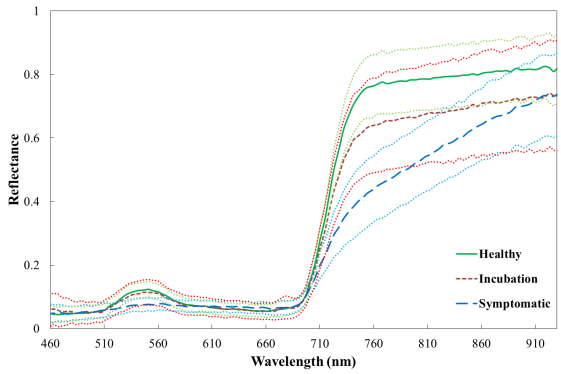


Fig. 2. Spectral plot of three strawberry Anthracnose infection stages from 460 nm to 930 nm.

3.2 Hyperspectral Imaging Analysis

Although the infection stages cannot be read directly from the reflectance spectra, we see different spectral patterns across infection stages as shown in Fig. 2. In particular, the three infection stages show different climbing slopes from 692 nm to 737 nm and dissimilar reflectance levels at 750 nm. The healthy spectra have much higher reflectance levels from 750 nm onward comparing with the other two infection stages. The slopes of healthy spectra are much steeper than the symptomatic spectra ones. Based on this observation, we

propose a measure called Simple Slope Measure (SSM) and formulated as:

$$\text{SSM} = (R_{737} - R_{692})/(737 - 692) \quad (2)$$

This measure is the first derivative (slope) of the spectrum from 692 nm to 737 nm. This simple measure only requires two reflectance values for calculation; hence, this measure is fast to compute.

In this research, we had applied SAM, SDA and SSM to the same spectral dataset and compared their performance in classifying strawberry foliage into correct corresponding infection stages. In order to prove the robustness of each hyperspectral imaging analysis method, 5-fold cross validation was applied. The F-to-enter and F-to-remove thresholds were set to corresponding to a tail probability of 0.0005 and 0.001 respectively.

Table 2 presents the classification accuracy for each analysis method in each cross validation round and its average. As the assessment was a three-class classification problem, with the same selected healthy and symptomatic areas the performance of SAM degraded to 82% from 95% in the two-class problem. Although in the second round of cross validation the classification accuracy reached 96.7%, which is even higher than the two-class classification average accuracy, the performances in different cross validation rounds showed an accuracy difference over 25%. This huge accuracy difference can be observed in SDA as well.

Table 2. Strawberry Anthracnose Assessment Results with Different Analysis Methods

Accuracy	SAM	SDA	SSM
1	90.0%	90.0%	70.0%
2	96.7%	86.7%	80.0%
3	76.7%	80.0%	63.3%
4	70.0%	63.3%	70.0%
5	76.7%	83.3%	80.0%
Average	82.0%	80.7%	72.7%

Studying SDA classification results more closely, the leave-one-out-cross-validation (LOOCV) accuracy achieved 83.3% using only three characteristic wavelengths. This result is actually better than the SAM average result and less wavelength information required. We see consistency in characteristic wavelength selection across different cross validation rounds: 825 nm was selected in all rounds. Only three characteristic wavelengths were picked in most rounds, apart from 825 nm, the other two wavelengths were around the 719 nm and 750 nm range. Fig. 3 illustrates clearly that where the selected characteristic wavelengths positioned are at which the spectra of three different stages are most parted or started to be separated in the neighbourhood wavelength range. The SDA classification accuracies across cross validation rounds were more stable than the ones in SAM classification rounds. This can be explained by the SAM using all wavelength property. As the SAM utilizes all reflectance across the whole wavelength range, the analysis would be more sensitive to minor differences accumulated from insignificant wavelengths than the SDA method. In other words, SAM is sensitive to the selection of the reference spectrum in each class (infection stage in this research).

Table 3. Strawberry Anthracnose Assessment by Stepwise Discriminant Analysis

	SDA	Characteristic Wavelength (nm)
1	90.0%	719, 750, 825
2	86.7%	728, 750, 825
3	80.0%	723, 750, 825
4	63.3%	715, 746, 825
5	83.3%	546, 551, 613, 666, 701, 719, 750, 825
LOOCV	83.3%	719, 750, 825

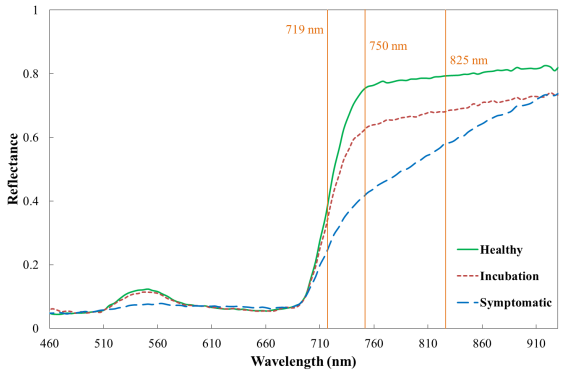


Fig. 3. Characteristic wavelengths selected by Stepwise Discriminant Analysis method: 719 nm, 750 nm and 825 nm.

From Table 2, it is clear that applying the slope from 692 nm to 737 nm solely to assess strawberry leaves into three infection stages was not efficient. The classification accuracy of SSM only reached 72.7% in average. Comparing the SSM performance with the SDA analysis, both methods utilized only a small number of wavelength information. However, the SSM was much ineffective than the SDA method, except in the cross validation round 4. This indicates that the training spectra in cross validation round 4 were quite noisy at most wavelengths; hence, the performance of SAM and SDA were as good as the SSM. However, the spectra in the range of 692 nm to 737 nm were representative and distinguishable for each infection stage. In other words, the SSM method can work much successfully than the other methods if the spectra data are noisy at most wavelengths but not in the range of 692 nm to 737 nm.

Examining the confusion matrix of each analysis method, each method's features are highlighted in Table 4. Misclassification error of healthy areas into other two stages was much lower in the SAM than in other two methods. This finding signifies that the healthy spectra at other wavelengths, apart from the range of 692 nm to 750 nm, have significant properties in the three-stage assessment. For example, the higher reflectance level from 750 nm onward is one characteristic to distinguish incubated areas from healthy areas. In addition, healthy areas had been misclassified into incubated area but not symptomatic areas in SAM and SDA, reinforcing the importance of other wavelength information to healthy areas.

On the other hand, the classification accuracy of symptomatic areas was higher for SSM than the other two methods. This

indicates that the slope from 692 nm to 737 nm is sufficient to represent symptomatic spectra and no other wavelength information required. The incubated areas were the most challenging spectra as they overlapped with the other two infection stage spectra. However, the SDA classification accuracy for incubated area was much better than the other two methods, especially than the SSM method. In other words, the slope in the 692 nm to 737 nm range is irrelevant to separate out the incubated area. Also, the reflectance information from other wavelengths other than the selected characteristic wavelength range can mislead incubated areas into healthy areas.

Table 4. Confusion Matrices of Hyperspectral Imaging Analysis on Strawberry Anthracnose Assessment

		Prediction			
	SAM	H	I	S	Total
Actual Infection	Healthy (H)	94%	6%	0%	50
	Incubation (I)	38%	62%	0%	50
	Symptomatic (S)	4%	6%	90%	50
					150
	SDA	H	I	S	Total
Actual Infection	Healthy (H)	80%	20%	0	50
	Incubation (I)	22%	74%	4%	50
	Symptomatic (S)	2%	10%	88%	50
					150
	SSM	H	I	S	Total
Actual Infection	Healthy (H)	78%	20%	2%	50
	Incubation (I)	38%	46%	16%	50
	Symptomatic (S)	4%	2%	94%	50
					150

4. CONCLUSIONS

The feasibility of hyperspectral imaging on strawberry foliage Anthracnose disease assessment has been shown in this research. The performances of Anthracnose assessment by two hyperspectral imaging analysis methods: Spectral Angle Mapper and Stepwise Discriminant Analysis, and one genuine Simple Slope Measure method were compared in great details. Each analysis method achieved an average accuracy of 82.0%, 80.7% and 72.7% respectively in classifying the examined leaves into healthy, incubation and symptomatic stages. With similar accuracy, Stepwise Discriminant Analysis uses less spectral information and save computational time than the Spectral Angle Mapper. In addition, the Simple Slope Measure performs well for filtering out symptomatic areas from other two infection stages. With Stepwise Discriminant Analysis, the incubated areas of strawberry leaves were classified correctly with an accuracy of 74.0% indicating early detection of Anthracnose is achievable by multispectral imaging. Although the Spectral Angle Mapper has simpler assessment processing and marginally better performance than the Stepwise Discriminant Analysis, it is not practical and expensive to have the hyperspectral imaging system scanning the strawberry leaves in the field. The characteristic wavelengths obtained from feature selection processes in Stepwise Discriminant Analysis can be reused in setting up the multispectral system, which requires less time in image acquisition. Hence, it is more cost efficient and feasible with acceptable assessment accuracy.

REFERENCES

Carvalho, O., and Meneses. P. (2000). Spectral Correlation Mapper (SCM): An Improving Spectral Angle Mapper. *The Ninth JPL Airborne Earth Science Workshop*, 65-74. Pasadena, CA, USA.

Cho, M.A., Debba, P., Mathieu, R., Naidoo, L., Aardt, J., and Asner, G.P. (2010). Improving Discrimination of Savanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 4133-4142.

Coops, N., Stanford, M., Old, K., Dudzinski, M., Culvenor, D., and Stone, C. (2003). Assessment of Dothistroma Needle Blight of Pinus radiate Using Airborne Hyperspectral Imagery. *Phytopathology*, 9(12), 1524-1532.

Curry, K.J., Abril, M., Avant, J.B., and Smith, B.J. (2002). Strawberry Anthracnose: Histopathology of Colletotrichum acutatum and C. fragariae. *Phytopathology*, 92(10), 1055-1063.

Ellis, M.A., and Erincik, O. (2008). Anthracnose of Strawberry. The Ohio State University Extension.

Girouard, G., Bannari, A., El Harti, A., and Desrochers, A. (2004). Validated Spectral Angle Mapper Algorithm for Geological Mapping: Comparative Study between Quickbird and Landsat-TM. *Proceedings of the International Society for Photogrammetry and Remote Sensing, The 10th ISPRS Congress*, 599-604. Istanbul, Turkey.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., and Goetz, A.F.H. (1993). The Spectral Image Processing System (SIPS) Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sensing of Environment*, 44, 145-163.

Mahlein, A., Steiner, U., Hillnhütter, C., Dehne, H., and Oerke, E. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods* 8, 3.

Qin, J., Burks, T.F., Zhao, X., Niphadkar, N., and Ritenour, M.A. (2012). Development of a two-band spectral imaging system for real-time citrus canker detection. *Journal of Food Engineering*, 108, 87-93.

Sohn, Y., and Rebello, N.S. (2002). Supervised and Unsupervised Spectral Angle Classifiers. *Photogrammetric Engineering and Remote Sensing*, 68(12), 1271-1280.

Sreenivasaprasad, S., Sharada, K., Brown, A.E., and Mills, P.R. (1996). PCR-based detection of Colletotrichum acutatum on strawberry. *Plant Pathology*, 45(4), 650-655.

Vargas, A.M., Kim, M.S., Tao, Y., Lefcourt, A., and Chen, Y. (2004). Safety inspection of cantaloupes and strawberries using multispectral fluorescence imaging techniques. *ASAE* Paper No. 043056, Ottawa, Canada.