Talanta 83 (2010) 574–581

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Mid-infrared spectroscopy for detection of Huanglongbing (greening) in citrus leaves

Sindhuja Sankaran, Reza Ehsani*, Edgardo Etxeberria

Citrus Research and Education Center, IFAS, University of Florida, 700 Experiment Station Road, Lake Alfred, FL 33850, USA

ARTICLE INFO

Article history:
Received 14 August 2010
Received in revised form
30 September 2010
Accepted 1 October 2010
Available online 8 October 2010

Keywords:
Citrus disease
Huanglongbing
Disease detection
Mid-infrared spectroscopy
Pattern recognition

ABSTRACT

In recent years, Huanglongbing (HLB) also known as citrus greening has greatly affected citrus orchards in Florida. This disease has caused significant economic and production losses costing about $750/acre for HLB management. Early and accurate detection of HLB is a critical management step to control the spread of this disease. This work focuses on the application of mid-infrared spectroscopy for the detection of HLB in citrus leaves. Leaf samples of healthy, nutrient-deficient, and HLB-infected trees were processed in two ways (process-1 and process-2) and analyzed using a rugged, portable mid-infrared spectrometer. Spectral absorbance data from the range of 5.15–10.72 μm (1942–933 cm^{-1}) were preprocessed (baseline correction, negative offset correction, and removal of water absorbance band) and used for data analysis. The first and second derivatives were calculated using the Savitzky–Golay method. The preprocessed raw dataset, first derivatives dataset, and second derivatives dataset were first analyzed by principal component analysis. Then, the selected principal component scores were classified using two classification algorithms, quadratic discriminant analysis (QDA) and k-nearest neighbor (kNN). When the spectral data from leaf samples processed using process-1 were used for data analysis, the kNN-based algorithm yielded higher classification accuracies (especially nutrient-deficient leaf class) than that of the other spectral data (process-2). The performance of the kNN-based algorithm (higher than 95%) was better than the QDA-based algorithm. Moreover, among different types of datasets, preprocessed raw dataset resulted in higher classification accuracies than first and second derivatives datasets. The spectral peak in the region of 9.0–10.5 μm (952–1112 cm^{-1}) was found to be distinctly different between the healthy and HLB-infected leaf samples. This carbohydrate peak could be attributed to the starch accumulation in the HLB-infected citrus leaves. Thus, this study demonstrates the applicability of mid-infrared spectroscopy for HLB detection in citrus.

1. Introduction

U.S. is the second largest producer of citrus in the world [1], producing up to 9.8 million tons of citrus each year. In Florida, there are approximately 550,000 acres of citrus orchards producing several varieties of citrus and supporting 9.3 billion dollar citrus based-industries. According to the USDA Foreign Agricultural Service Report, orange production in 2009–2010 season declined by few hundred thousand metric tons compared to 2008–2009 season [1]. Among various factors, citrus diseases have been the prominent reasons for the precipitous decline in production. Huanglongbing (HLB) or citrus greening is a devastating vector-based disease caused by the phloem-limiting bacteria *Candidatus* Liberibacter spp. that threatens the economics of citrus production in Florida and other parts of world. Researchers, citrus industries, and other stakeholders are working together to control and eliminate the disease for sustainable citrus production. HLB spreads through the vector Asian citrus psyllids (*Diaphorina citri* Kuwayama) that transmits the bacteria from tree to tree upon feeding. Once infected, citrus trees develop typical symptoms such as yellowing and thickening of veins or the entire leaf, blotchy leaf appearance, and formation of deformed, asymmetric fruit. Progression of the disease leads to tree decline, dieback of twigs, and finally tree death. The time period between the HLB-infection and the appearance of symptoms, as well as the time period between HLB-infection and tree death, depends largely on age, cultivar, and physiological status of the tree. The time period between the HLB infection and symptom appearance can range between six months and two years, time during which the psyllids can acquire the bacterial pathogen from non-symptomatic trees and further spread the disease [2]. Similarly, a young tree can be killed within one to two years after HLB infection [3].

The current disease management practices to control HLB and maintain productivity include bacterial inoculum reduction

* Corresponding author. Tel.: +1 863 956 1151x1228; fax: +1 863 956 4631.
E-mail address: ehsani@ufl.edu (R. Ehsani).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.10.008

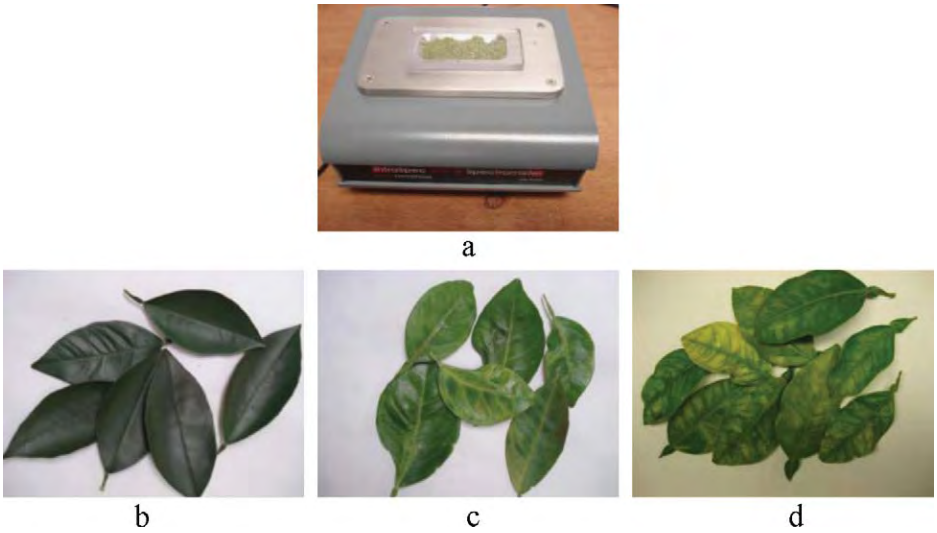

Fig. 1. (a) Mid-infrared spectrometer, (b) healthy, (c) HLB, and (d) nutrient-deficient leaves.

through disease detection and removal of HLB-infected trees, vector control through pesticide application, and replanting new citrus cultivars to maintain future productivity [2]. Presently, scouting for visible symptoms of HLB is the only available method for the selection of suspect plant material. Once identified by the scouting crew, the leaves are randomly sent to the laboratory for polymerase chain reaction (PCR) analysis for confirming the presence of bacteria.

HLB detection is a critical step in disease management. Once the HLB-infected trees are identified, the infected trees are immediately removed to prevent further spread of the disease. One of the major challenges in identifying leaf symptoms for HLB detection is the resemblance of the yellowing patterns between HLB-infected leaves and other nutrient-deficient conditions (such zinc, iron, and manganese) in leaves. In addition, with the scouting efficiency of visual symptoms varying from 50 to 60% in a single survey [2,4], there is a need for an efficient field-based sensing technique for HLB detection in citrus leaves. Spectroscopic techniques offer a real-time, rapid field-based detection of HLB. This study explores the applicability of mid-infrared (MIR) spectroscopy for HLB detection in citrus leaves. The MIR region of the electromagnetic spectra can be used for detecting biochemical compounds such as sugars and acids in leaves and in other materials such as corn, soil, jellies, food supplements and cotton trash [5–11].

In this study, a portable, mid-infrared spectrometer was used for detecting HLB in citrus leaves. This work evaluates the potential of mid-infrared spectroscopy in distinguishing the HLB-infected leaves from nutrient-deficient and healthy leaves.

2. Materials and methods

2.1. Sample preparation

Four to six leaves from 44 healthy, 17 nutrient-deficient, and 54 HLB-infected citrus trees, located at the Citrus Research and Education Center (CREC) groves, Lake Alfred, FL were collected. The HLB-infected samples consisted of symptomatic leaves with blotchy mottle or some yellowing. The HLB-infected leaves were collected from trees confirmed as HLB-infected by PCR analysis. The nutrient-deficient leaves, having typical nutrient-deficient symptoms, were identified by well-trained scouting team members of CREC groves. These leaves were collected from the grove blocks that were free from HLB infection. The leaves were acquired in batches to process the leaf samples within 24–48 h.

Orange trees sampled were of Hamlin, Valencia, and Midsweet varieties. In addition, some grapefruit samples were also analyzed. The leaves were processed in two different ways, termed process-1 and process-2, henceforth. In process-1, the leaves (four to six leaves for each sample) were processed by directly grinding them into fine powder under liquid nitrogen. In process-2, the leaves were dried in a forced draft oven for 48 h at 70 °C, followed by grinding using a Cyclotec^{TM} 1093 sample mill (Foss North America, Eden Prairie, MN). The ground leaf samples were stored in vials and used for spectral data collection using a portable mid-infrared spectrometer.

2.2. Data collection

An InfraSpec VFA-IR spectrometer (Wilks Enterprise Inc., East Norwalk, CT), was used to collect the mid-infrared spectra in the range of 5.15–10.72 μm (1942–933 cm^{-1}) with 0.04 μm resolution. This portable instrument (Fig. 1) was interfaced with a computer via a USB connection and operated in attenuated total reflection (ATR) mode. The operational set-up of the spectrometer was controlled using Igor Pro 6.01, a program provided by the manufacturer. The program allows preprocessing of the spectral data such as baseline correction, peak area estimation, and quantitative analysis among others. Both the reflectance and absorbance spectra can be acquired from the instrument. Some of the features of the instrument are summarized in Table 1. The processed healthy, nutrient-deficient, and HLB-infected leaf samples (Fig. 1) were placed on the top of the ATR crystal window (50 mm × 16 mm) of the spectrometer (without any pressure), covered with a polycarbonate sample cover, and spectral reflectance data were collected. Each spectrum represented 30 complete scans (for all wavelengths) of 1 s each (total time for scan was 30 s). The blank crystal (without any sample) was taken as the sample background.

Table 1
Features of mid-infrared spectrometer.

Parameters	Features
Dimensions	7 cm × 15.25 cm × 16.5 cm
Wavelength range	5.15–10.72 μm (1,942–933 cm^{-1})
Detector array	128 Pixel linear pyroelectric array
ATR crystal	Zinc selenide
ATR surface size	50 mm × 16 mm
No. of reflections/measurement	10

For each sample (collection of 4–6 leaves), about four to five reflectance spectra were collected and each spectrum was considered as a replicate. The data was not balanced manually to include as many as data points as possible. Few healthy and HLB-infected trees were sampled twice. The number of spectra collected from leaf samples processed using process-1 was 263, 84 and 327 for healthy, nutrient-deficient, and HLB-infected leaf samples, respectively. Similarly 134 healthy, 50 nutrient-deficient, and 193 HLB-infected sample spectra were collected from process-2 based leaf samples. The overall data consisted of 128 spectral reflectance or absorbance values for each sample.

2.3. Starch analysis

Selected leaf samples were analyzed for starch content to determine whether there was a difference in starch concentrations among the leaf samples. The starch analysis protocol was obtained from the Plant Cell Physiology research group at CREC, Lake Alfred, FL. For quantification of starch in leaves [12], leaf tissue samples (circle of 27.3 mm^2) were homogenized at 6500 rpm using a tissue homogenizer. Homogenization was performed for a total of 80 s in two cycles of 40 s using disposable 2 mL plastic tubes and 4 metal beads of 2.38 mm diameter (Mobio Laboratories, CA) with 500 μL of distilled water. After homogenization of the leaf tissue, the metal beads were removed and the tubes with the homogenized tissue boiled at 100 °C for 10 min, after which the samples were centrifuged for 2 min at 2500 rpm in order to precipitate any residual leaf tissue debris. From the supernatant, 300 μL were transferred to an Eppendorf tube, and 900 μL of pure ethanol were added to precipitate the starch by centrifuging the tube for 10 min at 10,000 rpm.

The supernatant was discharged and the pellet was resuspended in 1 mL of distilled water by agitating the mixture in a vortex mixer for 4 min. After re-suspension of starch pellet in 1 mL of water, 50 μL of iodine solution (8.8 g potassium iodide + 2.2 g iodine per liter of water) was added and absorbance of 250 μL of the sample (pipette into enzyme linked immune-sorbent assay, ELISA plates) was measured at 595 nm using ELISA microplate Reader Model 680 (BIORAD, California). The concentration of starch in the sample was calculated using a standard curve of starch-iodine made with pure rice starch from SIGMA. In addition to the quantitative analysis of starch content in the leaves, spectral signatures of different types of starch (potato, rice, and wheat) were also acquired. Water content of few healthy, nutrient-deficient, and HLB-infected leaves (Hamlin) was also measured. Water content in the leaves was measured by monitoring the percent loss of moisture by drying the leaf samples for 48 h at 70 °C.

2.4. Data preprocessing

The two spectral datasets collected from leaf samples were preprocessed before further analysis. Visual observation of the spectral data showed the presence of two absorbance peaks, one in the wavelength range from 5.5 to 6.5 μm (1538–1818 cm^{-1}) and another in the wavelength range from 9 to 10.5 μm (952–1112 cm^{-1}). The 5.5–6.5 μm peak was confirmed to be due to the presence of water. The MIR spectra of the water are presented in Fig. 2a. The peak between 9 and 10.5 μm (952–1112 cm^{-1}) was attributed to the carbohydrates present in the leaves [6,8,13]. The peak absorbance of HLB-infected samples in the spectral range 5.5–6.5 μm (water) was found to be higher than that of healthy samples. A preliminary study was performed to determine if there was a noticeable difference in the water content between healthy, nutrient-deficient, and HLB-infected leaves (Hamlin). Results indicated that there was no noticeable pattern in moisture content between the healthy and HLB-infected leaves (Fig. 2b). Therefore,

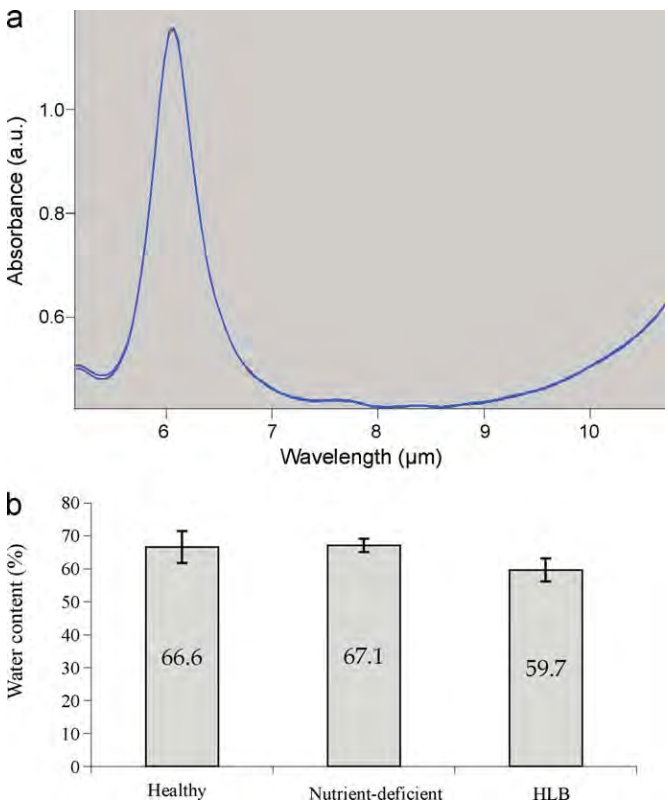

Fig. 2. (a) Water spectra and (b) water content in citrus leaves.

the MIR spectra of the samples were baseline corrected using Igor Pro 6.01 from 6.82 to 10.72 μm (1467–933 cm^{-1}), excluding the water absorbance spectra. The program corrects each spectrum such that a linear fit is drawn across the selected region and then, the spectrum is shifted so as to convert calculated fit line with zero slope and intercept. The 90 spectral features from wavelength range of 6.82–10.72 μm were considered for further analysis. A few representative spectra of healthy, nutrient-deficient, and HLB-infected leaves processed by two methods are presented in Fig. 3. After the baseline collection, the area of the region, peak location (wavelength), and peak absorption values were acquired for each spectrum in the range 9–10.5 μm (1110–952 cm^{-1}) using Igor Pro 6.01 program.

The baseline correction was required to bring the absorbance values to a comparable scale. After baseline correction, as some of the absorbance values were below zero, an offset correction was performed to convert the spectral absorbance greater than or equal to zero for better presentation. This was done using simple mathematical manipulation by estimating the minimum value and adding to all the absorbance values in the spectrum. The preprocessed MIR spectra were used for further analysis. The preprocessed spectral data acquired from leaf samples, processed using process-1 and process-2 is termed dataset-1 and dataset-2, hereafter. The first and second derivatives were calculated from the preprocessed datasets using the Savitsky–Golay filter [14]. In Savitzky–Golay filtering, an unweighted linear least-square fit based on polynomial equation is used to calculate the filter coefficients. The Savitzky–Golay filtering performs data smoothing, in addition to calculating the derivatives [15]. The derivatives were calculated using a window size of five and a second order of polynomial (quadratic). For each leaf sample, 85 first and second derivatives were derived from 90 spectral absorbance values.

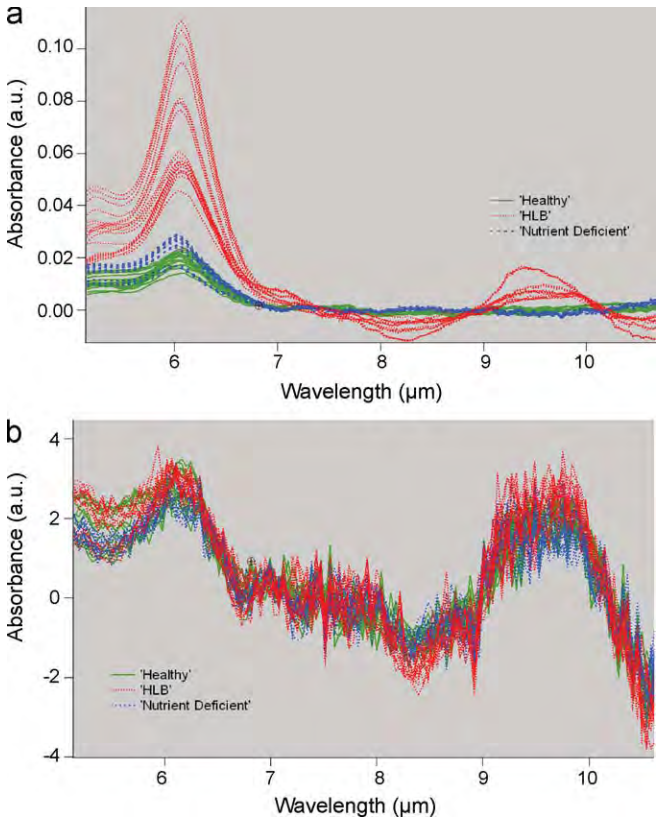

Fig. 3. Representative baseline corrected absorbance spectra (raw data) of leaf samples processed using (a) process-1 and (b) process-2.

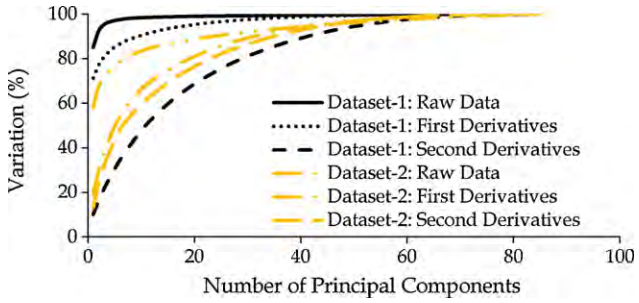

Fig. 4. The variation in the data as explained by the number of principal components in each dataset.

2.5. Pattern recognition

The preprocessed raw dataset, first derivatives dataset, and second derivatives dataset derived from dataset-1 and dataset 2 were analyzed separately. In each dataset, principal component analysis (PCA) was performed to reduce the dimensionality of the data, and number of principal components (PCs) were selected such that the PCs represented >99% variability within the data. The coefficients or loadings of PCA are generated based on the original variables that yield the PCs. The scores generated using selected PCs (selected based on variability each PC accounts for and total variability) were used as input features in pattern recognition algorithms. The datasets containing PCs were randomized and separated into training and testing datasets such that their ratio was about 3:1.

Two different multivariate classification algorithms, quadratic discriminant analysis (QDA) and k-nearest neighbors (kNN) were used for classifying spectral features of the leaf samples (healthy, nutrient-deficient and HLB-infected leaves). Both these technique have been used for various classification applications in literature [15–18]. The kNN-based algorithm classifies the unknown samples in the test dataset by determining the group/class of its 'k' nearest neighbors (shortest distance). In QDA, multivariate normal densities are derived from the covariance estimates stratified from each sample group. In case of larger number of PCs, a negative covariance matrix was generated that cannot be classified using QDA. In such cases, PCs accounting for 90% variability within the data were used as input in the QDA-based classification to generate a positive covariance matrix. The classification algorithms were tested five times, and the overall and individual class (healthy, nutrient-deficient, and HLB) classification accuracies were determined. All the preprocessing and classification was performed using the MATLAB® 7.6 program (The MathWorks Inc., Natick, MA).

In summary, the preprocessed raw dataset, first derivatives dataset and second derivatives dataset derived from the dataset-1 (process-1) and dataset-2 (process-2) were analyzed using two different algorithms (QDA and kNN). The principal components were calculated from each of the six datasets (2 processes × 3 types of dataset) and the number of PCs explaining the percent variation in the data is shown in Fig. 4. Based on the variation as described by the number of PCs, the PC scores were selected, randomized, and separated into training and testing datasets. The training and testing datasets from dataset-1 consisted of 506 and 168 data points, while those from dataset-2 consisted of 283 and 94 data points, respectively. Henceforth, class 1, class 2, and class 3 refer to healthy, nutrient-deficient and HLB-infected leaf sample spectra, respectively. For each dataset and each test run, a confusion matrix (sample matrix as shown in Table 2) was generated to calculate the overall and individual class classification accuracies. Due to the large amount of data available, the confusion matrices of all the test runs are not shown in the manuscript. Once the classification accuracies (overall and individual class) of the five test runs were determined, the average classification accuracies were calculated.

2.6. Comparison of means

The average of specific spectral features (area of region, peak location, and peak absorption values) in the spectral range of 9–10.5 μm of the three sample treatments (healthy, nutrient-deficient, and HLB-infected leaf sample spectra) were tested using analysis of variance (ANOVA). In addition to these specific spectral features, statistical analysis of the classification accuracies was

Table 2
Sample confusion matrix generated while testing kNN-based algorithm on test dataset (preprocessed raw data).

	Predicted									
	Class 1	Class 2	Class 3	No. of samples	Accuracy (%)	Class 1	Class 2	Class 3	No. of samples	Accuracy (%)
Actual	Process 1 (k = 1, 2)					Process 2 (k = 5)				
Class 1	62	–	–	62	100.0	30	3	–	33	90.9
Class 2	1	26	–	27	96.3	2	9	–	11	81.8
Class 3	1	–	78	79	98.7	–	–	50	50	100.0
Total				168	98.8				94	94.7

Table 3
Average overall and individual class classification accuracies of different datasets as classified using QDA- and kNN-based algorithms.

Dataset	Dataset-1						Dataset-2					
	PCs	Overall	Class 1	Class 2	Class 3	*k*	PCs	Overall	Class 1	Class 2	Class 3	*k*
k-Nearest neighbor												
Raw	19	99.2	99.4	99.3	99.1	1,2	73	93.6	95.3	79.6	96.6	5
First derivatives	43	98.8	99.2	98.9	98.4	1,2	62	87.7	84.7	77.3	92.6	3
Second derivatives	68	87.9	90.8	93.1	84.5	1,2	62	78.7	74.0	61.4	86.1	3
Quadratic discriminant analysis												
Raw	19	79.6	82.9	92.1	73.9		24*	86.2	94.5	29.4	97.1	
First derivatives	43	85.4	92.6	33.9	93.2		32*	80.9	93.6	0.0	91.6	
Second derivatives	42*	74.6	79.3	16.7	87.5		36*	76.6	83.1	0.0	88.5	

Note: PCs contribute 99% variation, except those marked (*with 90% variation).

also performed. The average classification accuracies as a result of two processing methods (process-1 and process-2) and two algorithms (QDA and kNN) were compared using a two sample *t*-test. In addition, ANOVA was used to compare the average classification accuracies that resulted from each type of data (preprocessed raw data, first derivatives and second derivatives). During the statistical analysis (ANOVA), Duncan's multiple range test (DMRT) was also performed, which compares the treatment means while controlling the comparison-wise error rate. All the statistical analysis was performed with 5% level of significance ($\alpha = 0.05$) and using a statistical program, SAS® 9.2 (SAS Institute Inc., Cary, NC, USA).

3. Results and discussion

3.1. Performance of classification algorithms

The average classification accuracies resulting from QDA- and kNN-based algorithms for all six datasets are summarized in Table 3. In the kNN-based algorithm, optimum '*k*' was determined by varying '*k*' from 1 to 15 based on the maximum average classification accuracies for each dataset. The optimum '*k*' for dataset-1 was found to be one and two; while, that for dataset-2 using preprocessed raw and derivatives (both first and second) dataset was found to be five and three, respectively.

Comparing the average classification accuracies of the three datasets (raw, first derivatives, second derivatives) derived from datatset-1, as analyzed using the kNN-based algorithm, it was observed that the preprocessed raw and first derivatives datasets yielded similar results. However, the second derivatives dataset of the dataset-1 resulted in lower average overall classification accuracy (87%) than that of other two datasets (~98%). The results from DMRT indicated that the preprocessed raw and first derivatives datasets yielded similar overall and individual class classification accuracies, which differed from those of second derivatives dataset (probability, $p < 0.05$). The only exception was class 2 (nutrient-deficient) classification accuracies ($p = 0.10$), which were similar in all the three datasets.

Comparing the average classification accuracies of the three datasets (dataset-1) acquired from the QDA-based algorithm, it was found that the preprocessed raw dataset yielded the highest classification accuracies. The first derivatives and second derivatives datasets resulted in a very low average class 2 classification accuracies (16.7–33.9%), in spite of the first derivatives dataset giving a good average overall, class 1 (healthy) and class 3 (HLB) classification accuracies. Duncan's multiple range test through ANOVA indicated that all the classification accuracies were different in the three datasets, except for the class 1 classification accuracy which was similar in preprocessed raw and second derivatives datasets.

In dataset-1, comparing the kNN and QDA-based algorithms, it was found that the kNN-based algorithm performed better than QDA-based algorithm yielding higher classification accuracy. One-on-one comparison of classification accuracies resulting from kNN- and QDA-based algorithms using statistical analysis (two sample *t*-test) indicated that with an exception of class 3 classification accuracies of the second derivatives dataset, none of the classification accuracies were similar. Among the three datasets, in kNN-based algorithm, preprocessed raw and first derivatives datasets resulted in high average classification accuracies (>98%); while in the QDA-based algorithm, the raw dataset resulted in good average classification accuracy.

In dataset-2, both kNN- and QDA-based classification algorithms yielded higher average classification accuracies (overall and individual class) using the preprocessed raw dataset, with decreasing performance from the first derivatives to second derivatives datasets. However, Duncan's multiple range test indicated that in the kNN-based algorithm, the preprocessed raw and first derivatives datasets yielded similar average class 2 and class 3 classification accuracies. In the QDA-based algorithm, the first and

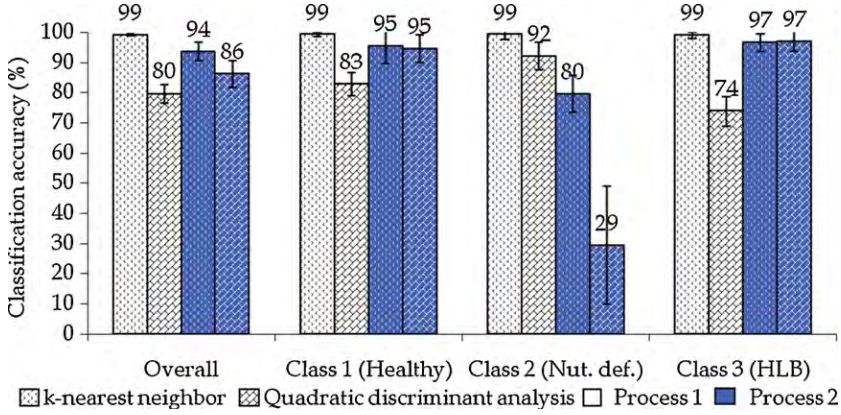

Fig. 5. Data analysis of preprocessed raw dataset using different classifiers.

Table 4
Comparison of different spectral features extracted from leaf sample spectra using statistical analysis.

Probability of null hypothesis (equal means) being true		
Spectral features	Process 1	Process 2
Peak absorbance	<0.0001[a]	<0.0001[c]
Peak location	<0.0001[a]	<0.0001[a]
Area of the region	0.0009[b]	0.0063[c]

[a] Three classes different through pairwise comparison.
[b] Classes 2 and 3 were similar.
[c] Classes 1 and 3 were similar; where class 1 refers to healthy, class 2 refers to nutrient-deficient, and class 3 refers to HLB.

second derivatives datasets yielded similar average overall, class 2, and class 3 classification accuracies; while preprocessed raw and first derivatives datasets resulted in similar average class 1 classification accuracies. Comparing the types of datasets, in both kNN- and QDA-based algorithms, the kNN-based algorithm (preprocessed raw dataset) exhibited higher classification accuracies than the QDA-based model, especially for the average overall and class 2 classification accuracies. Statistical analysis (two sample *t*-test) confirmed these results demonstrating that the average overall and class 2 classification accuracies were different ($p < 0.02$) between kNN- and QDA-based models; while the average class 1 and class 3 classification accuracies were similar ($p > 0.81$). In all the datasets, the preprocessed raw dataset resulted in good average overall and individual class classification accuracies (Fig. 5).

In addition to the different types of datasets (preprocessed raw, first and second derivatives) and the classification algorithms (kNN and QDA), the results from two different datasets processed in two different ways (process-1 and process-2) were also compared. A two sample *t*-test indicated that in the QDA-based algorithm, there was a significant difference in overall average and in individual class classification accuracies between dataset-1 and dataset-2. However, while analyzing the datasets with the kNN-based algorithms, although the average overall and class 2 classification accuracies showed some difference ($p < 0.02$), the average class 1 and class 3 classification accuracies were found to be similar ($p > 0.12$). This indicates that the process-1 was effective in classifying the nutrient-deficient (class 2) leaf sample spectra in comparison to the process-2. Moreover, the kNN-based algorithm showed a lower coefficient of variation in classification accuracies during dataset-1 analysis than those acquired from dataset-2 analysis.

To summarize the results, the kNN-based algorithm (in both dataset-1 and dataset-2) was found to be best suited for classifying the data with high classification accuracies (>95%) among the two algorithms. In addition, among the different datasets, the preprocessed raw dataset resulted in higher classification accuracies than first and second derivatives datasets, especially for dataset-2 using the kNN-based algorithm. When the false negative samples (HLB-infected samples classified as healthy or nutrient-deficient) were observed, it was found that most of the HLB-infected samples were classified as healthy rather than nutrient-deficient samples in all datasets except first derivatives-dataset 2 during kNN-based classification.

3.2. Spectral features

The statistical analysis of the spectral features (peak absorbance, peak location, and the area of the region) derived from the MIR spectral region from 9.0 to 10.5 μm indicated that at least average (comparison of means) of one of the classes were different from others (Table 4). Further analysis of pairwise comparison indicated that the peak absorbance and peak location of the three classes were different in process-1 spectral data; while peak location was different in process-2 spectral data among three classes. Thus, peak absorbance and peak location can be used as an indicator to identify HLB-infected leaf spectra, if the leaves are processed using process-1. In process-1, there a distinct difference in the absorbance of the starch peak (Fig. 3a). Thus, it can be stated that the difference in relative concentration of starch could be used as an indicator to HLB detection, if the leaves were processed in such manner (process-1). However, the leaf samples processed by process-2 did not show a difference in the absorbance in starch peak spectra (Fig. 3b). Statistical results demonstrated similar results, with no major difference in peak absorbance values, but some difference in the peak location values between the treatments. The peak location can be used for identifying HLB-infected spectra if the leaf samples are processed using process-2.

In addition to the spectral features derived from the MIR spectra, classification was performed using the kNN-based algorithm directly on the preprocessed raw spectral data in the range 9–10.5 μm (without principal component analysis). The '*k*' used for dataset-1 and dataset-2 was one and five, respectively. The overall and individual class classification accuracies are presented in Fig. 6.

The results indicated that the spectral absorbance values in the spectral range 9–10.5 μm could be used for classifying the different classes of leaf samples. High classification accuracies of >97% were achieved during the analysis of dataset-1, while the accuracies were lower with dataset-2 analysis. The results indicate that process-1 shows promises in detecting HLB under field conditions.

3.3. Starch analysis

Starch is known to accumulate in HLB-infected leaves compared to nutrient-deficient and healthy leaves [19–23]. The starch quantity in HLB-infected leaves can be about 8.5–20 times higher than that of healthy leaves in sweet oranges [24,25], which accumulate little or no starch. In addition, researchers have found that starch accumulation may occur even before the symptoms appear in HLB-infected leaves [12]. The starch build-up results from phloem blockages due to phloem limiting bacteria. Excessive starch leads to disintegration of the chloroplast thylakoid system resulting in leaf yellowing, a condition that may also occur due to branch breakage [25,26]. Thus, starch could be used as an indicator to HLB-infection in leaves. In this study, we monitored starch content in healthy and HLB-infected leaves of three different citrus varieties (Fig. 7a).

Starch concentration was consistently and significantly higher in HLB-infected leaves than in healthy leaves in all varieties tested. Our results also showed that starch content was also lower in nutrient-deficient leaves (similar to healthy leaves) compared to those of HLB-infected leaves. Therefore, it could be stated that starch detection above some threshold value using mid-infrared spectroscopy could aid in HLB detection.

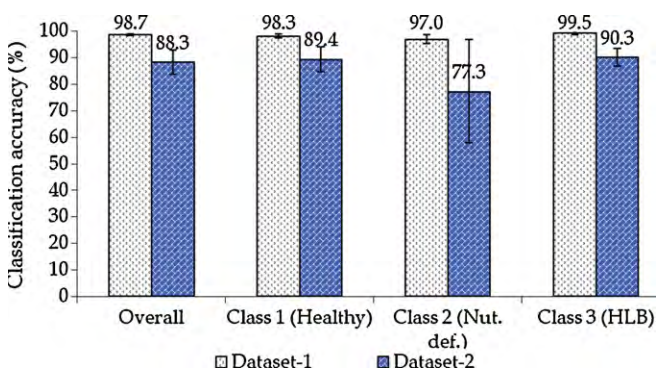

Fig. 6. Analysis of spectral data in specific wavelength range (9.0–10.5 μm) using k-nearest neighbor-based algorithm.

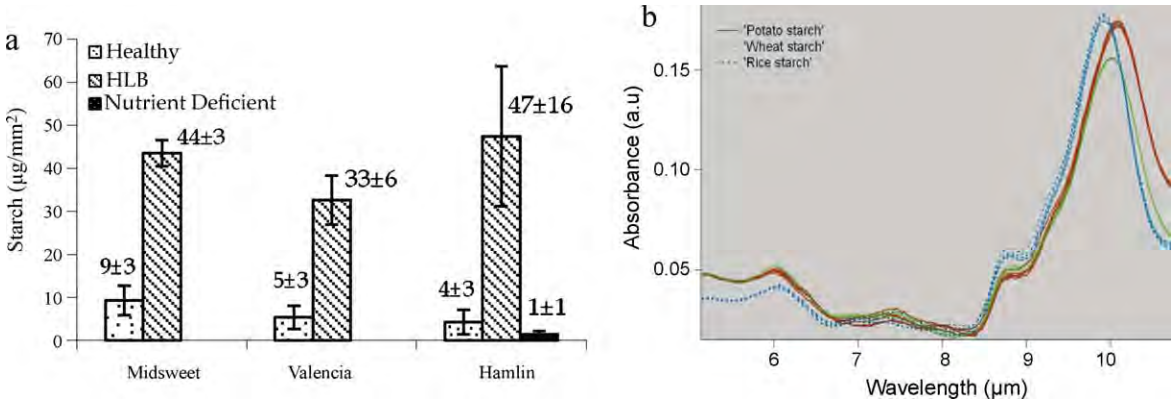

Fig. 7. Starch analysis: (a) quantitative analysis of starch in some varieties of oranges (mean ± standard deviation) and (b) spectral features of some types of processed starch.

The peak in the MIR spectral range 9.0–10.5 μm (952–1112 cm^{-1}) is attributed to the presence of carbohydrate (starch) in leaves. Several researchers have studied carbohydrates using MIR spectroscopy and found that carbohydrate vibration band occurs in this region [5,6,8,13,27]. In addition, in this study, mid-infrared spectra of different types of starch were collected (Fig. 7b). The spectral signature of processed starch types indicated the presence of peak in the spectral region from 9 to 10.5 μm. Studies have been conducted for detecting HLB in citrus leaves using Fourier transform infrared (FTIR) spectroscopy [13,27]. Some of the differences with this work are in the sample collection, sample preparation techniques, MIR instrument, and data analysis protocol. Hawkins et al. [13,27] found a similar peak in spectral range from 900 to 1150 cm^{-1}. Similarly, a study on the application of FTIR spectroscopy on selected carbohydrates [8] reported that the shape and intensity of the carbohydrate absorbance peak varied based on the type of carbohydrates, humidity, and other variables used in their study. Therefore, starch assessment in citrus leaves exhibits a good potential to detect HLB under field conditions.

4. Conclusions

This study investigates the applicability of mid-infrared spectroscopy for the detection of HLB in citrus leaves based on the massive accumulation of starch in affected leaves. A portable, field-based mid-infrared spectrometer operating in ATR mode was used to acquire spectral signature in the range 5.15–10.72 μm from processed leaf samples leaves. The results indicated that the spectra of HLB-infected citrus leaves can be classified from the spectra of healthy and nutrient-deficient samples using MIR spectroscopy. The kNN-based algorithm yielded classification accuracies of >90%. Further analysis of MIR spectra indicated that absorbance in spectral range 9.0–10.5 μm played a significant role in classification, which could be attributed to the starch present in the leaves. The method was also effective in classifying the nutrient-deficient leaves from that of healthy and HLB-infected leaves.

Future studies would involve the evaluation of this technique for detecting non-symptomatic HLB-infected leaves. There has been some indication that there might be one or more non-symptomatic tree for every symptomatic tree present in the field [2]. Therefore, it is critical to develop a field-based technique to not only detect symptomatic (having similar symptoms as that of nutrient deficiencies), but also to detect non-symptomatic HLB-infected leaves. Further work involves the determination of a starch threshold that would define HLB presence. One of the major benefits of this work can be the integration of such simple, portable MIR instrument with that of citrus mechanical harvesters for continuous monitoring of HLB.

Acknowledgements

The authors would like to thank the Florida Council Production Research and Advisory Council (FCPRAC) for their funding for this research. We would also like to extend our special thanks to Ms. Luba Polonik, Mr. Pedro Gonzalez, and Ms. Sherrie Buchanon. We would like to express our gratitude to Dr. Jose Reyes, Dr. Arnold Schumann, Ms. Laura Waldo, Dr. Juan Cevallos, and Ms. Shelley E. Jones for their support during the study.

References

[1] USDA Foreign Agricultural Service Report, 2008/09, Citrus: World markets and trade.
[2] T.A. Spann, R.A. Atwood, M.M. Dewdney, R.C. Ebel, R. Ehsani, G. England, S.H. Futch, T. Gaver, T. Hurner, C. Oswalt, M.E. Rogers, F.M. Roka, M.A. Ritenour, M. Zekri, B.J. Boman, K-R. Chung, M.D. Danyluk, R. Goodrich-Schneider, K.T. Morgan, R.A. Morris, R.P. Muraro, P. Roberts, R.E. Rouse, A.W. Schumann, P.A. Stansly, L.L. Stelinski, Publication no. HS1165, Horticultural Sciences Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, 2010.
[3] K.R. Chung, R.H. Brlansky, Fact sheet PP-210, Florida Cooperative Extension Service, Institute of Flood and Agricultural Sciences, University of Florida, 2009.
[4] S. Futch, S. Weingarten, M. Irey, Proceedings of the Florida State Horticultural Society 122 (2009) 152–158.
[5] N. Dupuy, C. Wojciechowski, C.D. Ta, J.P. Huvenne, P. Legrand, Journal of Molecular Structure 401–411 (1997) 551–554.
[6] M. Mascarenhas, J. Dighton, G.A. Arbuckle, Applied Spectroscopy 54 (2000) 681–686.
[7] J. Čopíková, A. Synytsya, M. Černá, J. Kaasová, M. Novotná, Czech Journal of Food Science 19 (2001) 51–56.
[8] M. Kačuráková, R.H. Wilson, Carbohydrate Polymers 44 (2001) 291–303.
[9] A.M. Allen, J.A. Foulk, G.R. Gamble, Beltwide Cotton Conferences, San Antonio, Texas, January 3–6, 2006, pp. 1938–1945.
[10] N. Hayashi, A. Hashimoto, K. Suehara, M. Kanou, T. Kameoka, T. Kumon, K. Hosoi, SICE Annual Conference, Kagawa University, Japan, September 17–20, 2007, pp. 220–223.
[11] R. Linker, IFIP International Federation for Information Processing, vol. 259, Computer and Computing Technologies in Agriculture, 2008, pp. 1137–1146.
[12] E. Etxeberria, P. Gonzalez, Plant Cell Physiology, Citrus Research and Education Center, Lake Alfred, FL, 2009–2010.
[13] S.A. Hawkins, B. Park, G.H. Poole, T. Gottwald, W.R. Windham, K.C. Lawrence, Applied Spectroscopy 64 (2010) 100–103.
[14] A.R. Mishra, D. Karimi, R. Ehsani, W.S. Lee, Transaction of ASABE, under review.
[15] S. Sindhuja, A. Mishra, R. Ehsani, Computer and Electronics in Agriculture, under review.
[16] L.R. Khot, S. Panigrahi, S. Woznica, Biological Engineering 1 (2008) 127–143.
[17] S. Panigrahi, Y. Chang, L.R. Khot, J. Glower, C.M. Logue, IEEE Sensors Applications Symposium, Atlanta, GA, February 12–14, 2008, pp. 85–88.
[18] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Analytica Chimica Acta 671 (2010) 27–35.
[19] M. Onuki, N.T.N. Truc, H. Nesumi, L.T.T. Hong, H. Kobayashi, Development of New Technologies and their Practice for Sustainable Farming Systems in the Mekong Delta, JIRCAS Workshop, Cantho, November 26–28, 2002.
[20] S. Taba, K. Nasu, K. Takaesu, A. Ooshiro, Z. Moromizato, Agricultural Department Academic Report, University of Ryukyus, Japan, 53 (2006) 19–23.
[21] E. Etxeberria, P. Gonzalez, W. Dawson, T. Spann, Publication no. HS1122, Horticultural Sciences Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, 2007.

[22] E. Etxeberria, P. Gonzalez, D. Archor, G. Albrigo, Physiological and Molecular Plant 74 (2009) 76–83.
[23] J-.S. Kim, U.S. Sagaram, J.K. Burns, J-.L. Li, N. Wang, Phytopathology 99 (2009) 50–57.
[24] T. Takushi, T. Toyozato, S. Kawano, S. Taba, K. Taba, A. Oshiro, M. Numazawa, M. Tokeshi, Japanese Journal of Phytopathology 73 (2007) 3–8.
[25] H. Schneider, Phytopathology 58 (1968) 1155–1160.
[26] A.A. Schaffer, K-.C. Liu, E.E. Goldschmidt, C.D. Boyer, R. Goren, Journal of Plant Physiology 124 (1986) 111–121.
[27] S.A. Hawkins, B. Park, G.H. Poole, T.R. Gottwald, W.R. Windham, J. Albano, K.C. Lawrence, Journal of Agricultural and Food Chemistry 58 (2010) 6007–6010.