DETECTION OF APPLE MARSSONINA BLOTCH DISEASE USING PARTICLE SWARM OPTIMIZATION

M. Shuaibu, W. S. Lee, Y. K. Hong, S. Kim

ABSTRACT. *Apple Marssonina blotch (AMB) is a devastating disease that is predominantly found in Asian countries, such as Japan, India, and South Korea. The disease has been known to cause huge economic losses in the regions where it has been found. AMB causes early defoliation, which ultimately leads to low quality and quantity of harvested apples. In this work, spectroscopic measurements were collected and analyzed for two datasets from 2014 and 2015. A stochastic algorithm called particle swarm optimization (PSO) was used to find optimal features for classification. A total of ten spectral features were found by the algorithm by selecting pairs of bands that resulted in the highest discrimination between every two classes. A support vector machine classifier resulted in 100% classification accuracy for both healthy and diseased samples. Abundance estimation and spectral unmixing analyses of early-stage AMB (ambE) samples were also conducted using PSO to extract symptomatic and asymptomatic endmembers. Results showed reasonable separation between healthy, seemingly healthy, and symptomatic classes. Quantitative analysis, using varying degrees of infection of ambE samples, was performed by applying a combination of partial least squares and stepwise multiple linear regression models, and coefficients of determination (R^2) of 0.76 and 0.71 were achieved for the calibration and validation datasets, respectively. The results demonstrate the potential of using spectroscopic technology as a non-invasive method for early detection of AMB disease.*

Keywords. *Apple, Detection, Marssonina blotch, Particle swarm optimization, Reflectance, Spectral unmixing, Spectroradiometer.*

The apple (*Malus domestica*) is one of the most important fruit crops in the world, mostly because of the numerous ways it can be consumed and the many health benefits it offers. According to the Food and Agriculture Organization (FAO), apple ranks second worldwide after banana in terms of production, with over 80.8 million metric tons of apples produced in 2013, and Asia alone accounted for over half that quantity (FAO, 2013). Despite this high figure, apple production has been declining in recent years in some of the top producing countries in the world, including China, Japan, India, Italy, and South Korea. A major part of this decline is attributed to a disease called apple Marssonina blotch (AMB). AMB is a severe fungal disease that primarily infects apple tree leaves and is caused by a pathogen called *Diplocarpon mali*. The first appearance of the disease was recorded in Japan in 1907, and by the 1980s, the disease had spread to some other countries in Asia, Europe, and North America (Harada et al., 1974; Lee et al., 2011; Lee and Shin, 2000; Tamietti and Matta, 2003). A case in point of the prevalence of the disease is South Korea, which has suffered major economic losses, with over 50% of apple orchards infected by the disease (Katti et al., 2015).

AMB disease occurs in the summer after periods of long rainfall, and it generally thrives in high humidity and warm temperature climates. It is dispersed by wind and rain and spreads in two major ways during the apple growing season. The primary form of infection is caused by ascopores that are released from overwintered apothecia in fallen leaves, while the secondary infection is caused by asexually produced fungal spores in the acervuli (EPPO, 2013). A long latency period of two to five weeks is typical for AMB, and symptoms begin to develop after this period. At the early symptomatic stage of the disease, small grayish black or brownish spots appear on the surface of the leaves. The disease then progresses to a stage where the spots coalesce, and necrotic and chlorotic blotches appear. The size of the blotches keeps growing until leaves turn yellow and prematurely fall off the tree. Leaf defoliation affects the quality and quantity of apples on a tree, including a reduction in fruit size and starch content.

The major preventative methods adopted by apple growers whose orchards have been infected by the disease are burning and burying of defoliated leaves. Treatments for AMB, including thiophyphanate-methyl fungicide, exist for control of the disease; however, there have been reports of the disease pathogen being resistant to some of these treatments (Tanaka et al., 2000). At this point, growing AMB-resistant cultivars might be one of the few ways to economically, reliably, and effectively control the spread of this disease, and some researchers have been working on finding disease-resistant cultivars and species (Yin et al., 2013).

Submitted for review in December 2015 as manuscript number ITSC 11700; approved for publication by the Information, Technology, Sensors, & Control Systems Community of ASABE in January 2017.

The authors are **Mubarakat Shuaibu, ASABE Member,** Former Graduate Student, and **Won Suk Lee, ASABE Member,** Professor, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida; **Young Ki Hong,** Agricultural Engineer, and **Sangcheol Kim,** Agricultural Engineer, South Korean Rural Development Administration, Jeonju, South Korea. **Corresponding author:** W. S. Lee, 1741 Museum Road, P.O. Box 110570, Gainesville, FL 32611-0570; phone: 352-392-1864 ext. 207; e-mail: wslee@ufl.edu.

Vol. 60(2): 303-312 ISSN 2151-0032 DOI 10.13031/trans.11700

Checking for the occurrence of disease by visually inspecting each tree in the field is a challenging task, not only because this is a time-consuming process but also because there is a high chance that infected leaves may not be spotted, given the long latency period of the disease.

Because AMB poses so many management challenges for growers, developing methods for early diagnosis is imperative. Studies focused on AMB detection, using existing technologies, are still relatively rare, and to the best of our knowledge only one technique has been explored in the literature. Lee et al. (2012) reportedly diagnosed the disease at the early and asymptomatic stages using a technique called optical coherence tomography (OCT). From two-dimensional (2D) and three-dimensional (3D) imaging scans created by the OCT system, they were able to find distinctive differences between the inner cross-sectional layers of healthy and diseased leaves. They concluded that an early-stage AMB detection tool could be developed based on an upgraded version of the system. However, it is not known if this technique can distinguish between leaves at different stages of infection.

Spectroscopy is extensively used in precision agriculture for assessing the general health of crops (Del Fiore et al., 2010; Gómez-Sanchis et al., 2008; Graeff et al., 2006; Qin et al., 2008). It is preferred to some other tools currently used for qualitative and quantitative analyses in agriculture because it is inexpensive, accurate, fast, and non-destructive (Li et al., 2007; Roggo et al., 2002; Sankaran et al., 2010). Plant disease detection is one area that has been successfully tackled using spectroscopy. Jones et al. (2010) and Xu et al. (2007) showed the potential of spectral technology in the detection of bacterial leaf spot and leaf miner diseases in tomatoes, respectively. They successfully created disease prediction models capable of diagnosing tomato diseases at different severity levels. The notorious Huanglongbing (HLB) disease in citrus has also been detected with accuracies of 60% to 80% using spectral features developed from reflectance data (Katti et al., 2015).

Spectroscopic data usually possess high numbers of spectral bands, with some datasets containing hundreds or even thousands of bands. With so many spectral bands, the feature space of a given spectral dataset could potentially contain tens of thousands of features (Thenkabail et al., 2011). In most situations, some of these features hold little or no information about the target of interest, and including them in information extraction processes such as classification could slow the process and cause inaccurate classification results. One way of dealing with this problem is to apply preprocessing techniques such as feature selection and extraction to the spectral data before classification is performed. Various methods for feature selection and extraction exist, with the most popular being spectral band selection and projection pursuit measures. Many researchers have investigated several dimensionality reduction methods based on these measures (Bruce et al., 2002; Jia and Richards, 1999; Martínez-Usó et al., 2007; Wang and Angelopoulou, 2006; Yang et al., 2012; Zhang et al., 2005). Feature selection methods are generally preferred to feature extraction methods because the latter do not preserve the original information; instead, they produce transformed features using all the original data. This is especially undesirable in situations where the goal is to build a multispectral sensor based on a subset of bands or features.

A spectroradiometer is one of the most popular spectroscopic systems used to detect plant stress by measuring crop spectral reflectance. It has several advantages, including ease of use, efficiency, and portability. However, some spectroradiometer systems, such as the one used in this work, are only capable of extracting one spectral signature of a wide region on an object being sensed. This poses a problem when the aim is to analyze regions with different chemical properties separately. Optimization algorithms, such as particle swarm optimization (PSO), have been successfully used by some researchers in extracting endmembers from spectrally mixed data (Omran et al., 2006; Zhang et al., 2011). PSO is preferred to some of the other well-known endmember extraction algorithms due to its ease of use, efficiency, and robustness in solving optimization problems.

In this study, PSO was used for selecting optimal spectral bands and for spectrally unmixing early-stage diseased samples. The algorithm is versatile and can resolve many optimization problems, but it has not been used extensively in agriculture. The objectives of this work were to determine the optimal spectral features for AMB disease detection and to develop methods for qualitative and quantitative detection of the disease.

MATERIALS AND METHODS

DATA COLLECTION

The leaf samples used in this work were acquired from Fuji apple trees in an experimental apple orchard located in Gunwi City, South Korea. A test area measuring 40 m × 60 m, with a total of 260 trees, was set aside for the experiment. Datasets were acquired during the fall season in two consecutive years (2014 and 2015). Before leaves were plucked from the trees and analyzed, the trees were inoculated with AMB spores to facilitate disease development. Leaf samples were collected on different days to ensure that healthy leaves and leaves with different degrees of AMB infection were included in the dataset. Molybdenum and manganese nutrient deficient leaves were also included in the dataset due to their similarity in color with some of the other classes.

Indoor spectral measurements were obtained from the samples immediately after they were plucked from the trees to minimize damage to the cell structure of the leaves. Spectral reflectance information was acquired between wavelengths 350 nm and 2500 nm using a system that consisted of a spectroradiometer (Field Spec 3, ASD, Inc., Boulder, Colo.) and a plant probe with an attached leaf clip (Leafclip Assembly A122325, ASD, Inc., Boulder, Colo.). The leaf clip had a spot size of 10 mm, and it was used to hold leaves in place while spectral measurements were obtained. Before clipping leaves to the plant probe, a white polytetrafluoroethylene (PTFE) reflector with 99% reflectance was attached to the probe and used for calibrating the system to reflectance. After spectral measurements of leaf samples had been

acquired, the data were transferred and stored on a laptop computer (Intel Core i7-3720QM, HP). The spectroradiometer system had a spectral resolution of 1 nm and included three detectors: VNIR (350-1000 nm), SWIR1 (1000-1800 nm), and SWIR2 (1800-2500 nm). The system operated with a scanning time of 100 ms, and stable illumination was produced throughout the spectral range using a halogen lamp (ASD Illuminator, 70 W, ASD, Inc., Boulder, Colo.).

A total of 621 and 751 samples were acquired in 2014 and 2015, respectively. Spectral measurements from both years contained 2151 spectral bands, and five different classes were defined for the datasets, as shown in figure 1. The classes were healthy mature (healthM), healthy young (healthY), early-stage AMB diseased (ambE), advanced-stage AMB diseased (ambA), and molybdenum/manganese nutrient deficient (nd). Aside from the healthy classes, all other classes had at least two different-colored regions on the leaves. As stated earlier, a leaf clip with a spot size of 10 mm was used in the spectral data acquisition process. This large spot size made it impossible to capture spectral information of individual symptomatic and asymptomatic regions; instead, the average reflectance spectra of all the colors enclosed within the clipped region was captured, resulting in spectrally mixed data for the AMB diseased and nutrient deficient samples. The ambE spectral class was a spectral mixture of the reflectances of brown and green portions of early-stage AMB diseased leaves. The ambA samples mostly contained spectrally mixed green, yellow, and brown colors, while the nd spectral class contained a spectral mixture of light and dark green portions of nutrient deficient samples.

The sampling area of each leaf was the same size as the diameter of the spectroradiometer leaf clip, and sampling areas were strategically chosen to ensure that different variations of the stressed samples were analyzed. Some samples contained very clear symptoms, while symptoms on other samples could hardly be seen. Only one area on a single leaf was scanned because the goal was to extract different stress levels from the leaves. As a result, a uniformity check between different regions on the same leaf was not performed, as the size and quantity of AMB blotches varied from one region of a leaf to another.

Particle Swarm Optimization

Particle swarm optimization (PSO) is a methodology in evolutionary computation that was invented over two decades ago (Eberhart and Kennedy, 1995). PSO is a stochastic algorithm inspired by the social behavior of fish schooling and bird flocking. PSO simulates the social behavior of humans and insects. Individuals interact with one another while also learning from their own experiences; with time, the population members move into more suitable regions in the problem space. Just like genetic algorithm (GA) and other heuristic tools, PSO is randomly initialized with a set of potential solutions and iteratively searches for an optimal solution by updating the population in each iteration. Unlike GA, however, PSO does not make use of the selection, crossover, and mutation evolution operators in its implementation. In the PSO algorithm, each particle represents a potential solution; each particle flies through a multidimensional search space by following the current optimum particles and keeps a record of its current position as well as the best position it has achieved so far. Aside from the personal best positions achieved by each particle, the algorithm also finds a global best position for all particles in the search space, a position also referred to as the "best" of the personal bests. Particles are initialized with random positions and velocities, with the velocity of each particle adjusted according to its own flying experience and those of other particles. Each particle position is updated using Euler's integration equation. The velocity and position of each particle are modified according to the following equations:

$$v_{(i+1)} = \omega v_{i(t)} + C_1 rand\,(p_i best - x_{i(t)}) + C_2 rand\,(gbest - x_{i(t)}) \quad (1)$$

$$x_{i(t+1)} = x_{i(t)} + v_{i(t+1)} \quad (2)$$

where

v = velocity of the ith particle
x = position of the ith particle
i = particle index
t = discrete time index
$p_i best$ = personal best position of the ith particle
$gbest$ = global best of all particles
$rand$ = uniformly distributed number between 0 and 1
C_1, C_2 = weighting factors. In most cases, $C_1 = C_2 = 2$
ω = inertia function. Values close to one facilitate global exploration, while values close to zero facilitate a local exploration. The algorithm performs best if the inertia function linearly decreases through the course of the implementation of the algorithm.

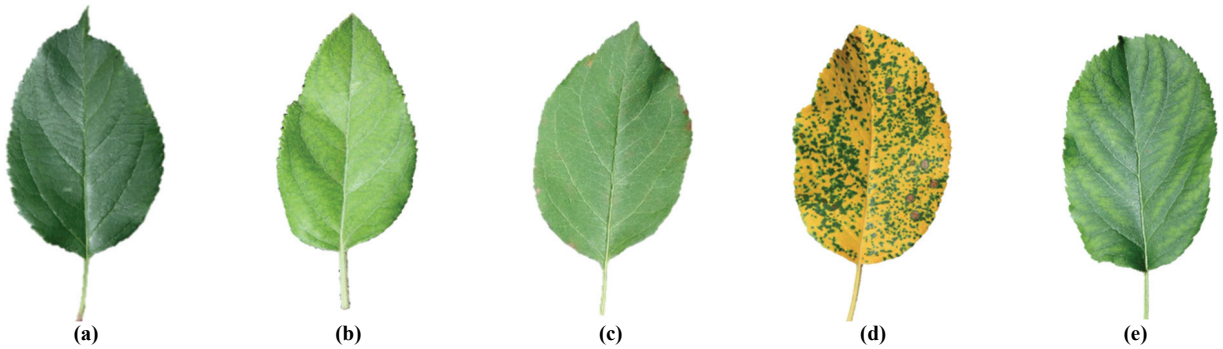

Figure 1. Apple leaves used in indoor spectroradiometer analysis: (a) healthy mature (healthM), (b) healthy young (healthY), (c) early-stage AMB diseased (ambE), (d) advanced-stage AMB diseased (ambA), and (e) molybdenum/manganese nutrient deficient (nd).

PSO Band Selection

The traditional method of finding the most discriminatory pairs of spectral bands for any two classes involves an exhaustive search in the feature space. This can be a computationally expensive process, especially when thousands of features are involved. The PSO-based band selection method used in this work resolves this problem by applying a heuristic approach to finding optimal bands. The goal was to find two bands that best maximize the difference between each pair of classes. An objective function (B) was created using the norm of class vectors, and it is mathematically expressed as follows:

$$B = abs\left[\left(norm(X_i) - norm(Y_i)\right) - \left(norm(X_j) - norm(Y_j)\right)\right] \quad (3)$$

where X and Y represent two sets of class vectors, i and j are two pairs of spectral bands and the function, and *abs* is the absolute value of the norm difference between the bands.

Abundance Estimation and Spectral Unmixing of Early-Stage AMB Diseased Samples

Early-stage AMB diseased samples had both brown and green colored regions. A couple of steps were taken to arrive at the abundance estimation for both regions (fig. 2). After spectral measurements had been taken, a white ring was used to mark the regions where spectral data were acquired. A digital camera (EOS 5D, Canon, Japan), with a focal length of 24 mm and an exposure time of 0.04 s, was used for acquiring color images of leaf samples. Hough transform was used in extracting only regions marked by the white ring. To allow easy segmentation of brown pixels, green pixels were first masked using the ratio of the red channel to the green and setting green pixel extraction to less than or equal to a threshold of 0.5. This threshold was chosen because it gave the best separation between green pixels and other colors. Nineteen color and texture features were then created and used as input for a supervised k-nearest neighbor (kNN) classifier. The classifier used a Euclidean distance metric and a neighborhood size of four for segmenting ambE color images.

PSO was again used for spectrally unmixing brown and green endmembers' reflectance from the original spectral range. The linear unmixing model was used in relating the abundances of each colored region to the reflectance spectrum generated by the spectroradiometer. The linear mixing model for a mixed spectrum (X_{mixed}) containing two endmembers is given by:

$$X_{mixed} = f_{c1}a_{c1} + f_{c2}b_{c2} \quad (4)$$

where F_{c1} represents the fractional abundance for class $c1$, and F_{c2} represents fractional abundance for class $c2$, while a and b represent the unmixed reflectance data for class $c1$ and class $c2$, respectively.

In order to get the spectral reflectance for each of the classes, the objective function (J) had to be minimized using the least squares equation given below:

$$J = \frac{1}{2}\left[X_{mixed} - (f_{c1}a_{c1} + f_{c2}b_{c2})\right]^T \times \left[X_{mixed} - (f_{c1}a_{c1} + f_{c2}b_{c2})\right] \quad (5)$$

where T represents the transpose function.

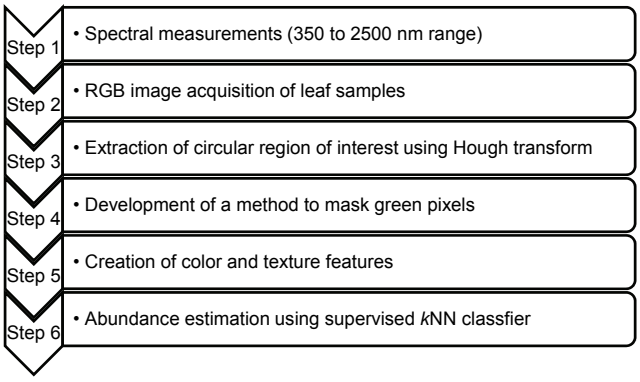

Figure 2. Flowchart of steps for extracting brown and green colored pixel abundances in early-stage AMB diseased samples.

Regression Models

Both partial least squares regression (PLSR) and stepwise multiple linear regression (SMLR) statistical methods were used in building prediction models for the quantitative assessment of early-stage AMB samples. One major difference between PLSR and SMLR is that the former makes use of all the input variables in developing a predictive model, while SMLR selects a subset of variables by successively adding or removing them based on a selection criterion, such as the p-value. SPSS Statistics 23 (IBM, Armonk, N.Y.) was used in developing the prediction models.

Results and Discussion

Spectral Feature Analysis

The mean reflectance spectra of the five classes from 2014 and 2015 are shown in figures 3a and 3b, respectively. Inspection of the plots shows that the signatures of the five classes are notably different. The slight differences in signatures for the same class pairs in both years can be attributed to varying fractional abundances of different leaf colors. At the nitrogen absorption band of 550 nm, the samples that contained dark green regions (healthM, ambE, and nd), and hence more chlorophyll and nitrogen, had lower peaks when compared to samples with light green or yellowish regions (healthY and ambA). Higher reflectance values in this band were only experienced by severely stressed and healthy young samples as a result of their color. The reflectance in the near-infrared (NIR) region, between 700 nm and 1000 nm, was high for all classes due to the internal scattering of light within the leaf structure, which is a phenomenon present in all vegetation. As the water content in the leaf increases, so does the absorption strength at 1450 nm and 1950 nm. The plots show that healthier samples had stronger absorption at these bands than stressed vegetation because the healthier samples contained more water. Further study of both sets of plots reveals slight differences between the spectral amplitudes of the data acquired in 2014 and in 2015. Before subsequent analyses were performed, both datasets were

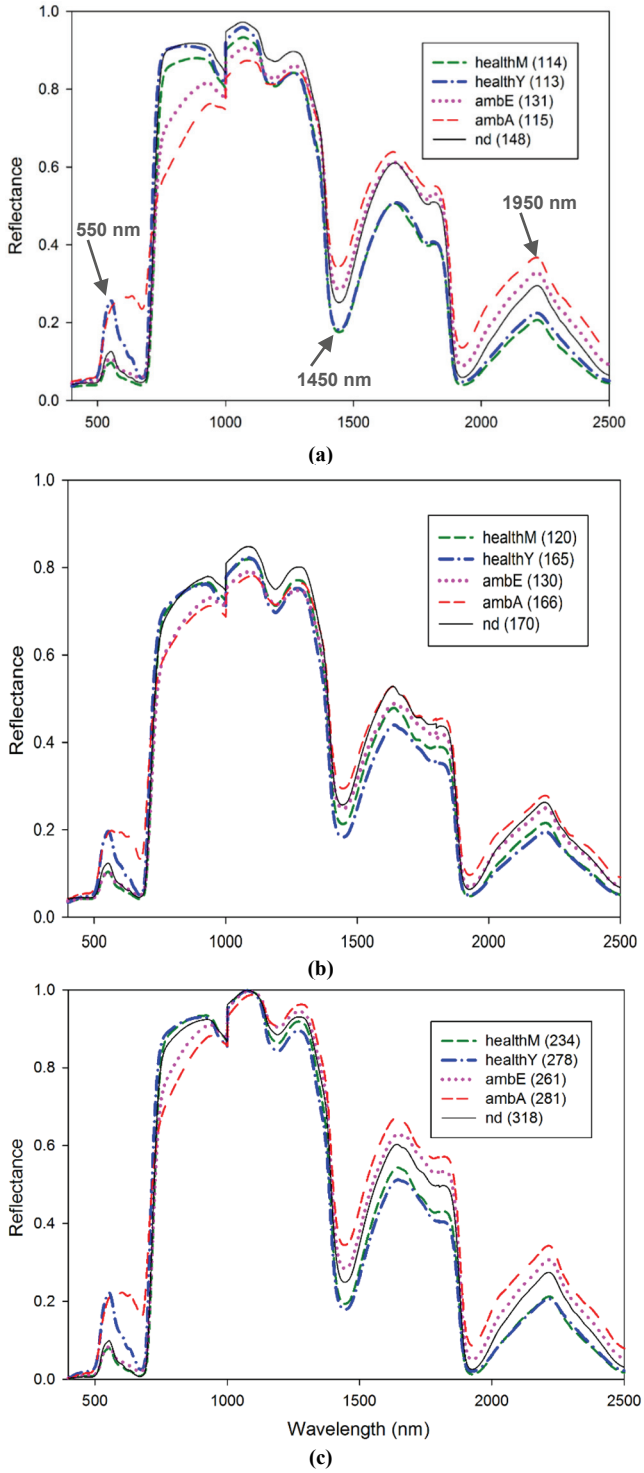

Figure 3. Mean reflectance spectra of healthy mature (healthM), healthy young (healthY), early-stage AMB diseased (ambE), advanced-stage AMB diseased (ambA), and molybdenum/manganese nutrient deficient (nd) samples in the spectroradiometer dataset for (a) 2014, (b) 2015, and (c) 2014 and 2015.

combined and converted to the same scale using values between 0 and 1. A representation of the transformed data is shown in figure 3c.

PSO-Selected Spectral Bands

The pairs of spectral bands that maximized the PSO objective function were extracted, and they resulted in ten spectral features (table 1). Two-waveband scatter plots were created to visualize the separation between classes, and they are shown in figure 4. It can be seen from the figure and table that bands between wavelengths 623 nm and 772 nm are important for detecting the advanced stage of the disease (ambA). Most of the bands chosen to discriminate between ambA and other classes are concentrated in the visible range of the spectrum due to the distinct and mostly yellowish color of the ambA samples. Wavelengths around 700 nm, which represents the red-edge region, were also selected for ambE and nd discrimination against other classes, indicating the importance of this region for analyzing crop stress. The wavelength at 477 nm was one of the bands selected for differentiating between healthM and nd.

Table 1. Optimal PSO spectral features created from 20 wavebands.

Feature	Wavelengths (nm)	Class Pairs
1	718, 1470	healthM, healthY
2	477, 1194	healthM, nd
3	765, 1751	healthM, ambE
4	636, 772	healthM, ambA
5	713, 1382	healthY, nd
6	721, 1833	healthY, ambE
7	690, 742	healthY, ambA
8	764, 1896	ambE, nd
9	623, 768	ambA, nd
10	700, 771	ambE, ambA

The blue spectral range is a region where plants are sensitive to the loss of chlorophyll, and because the nd class consists of some light green regions, this band indicates loss of nutrients (Thenkabail et al., 2011). Mature and young healthy apple leaves can be separated using wavelengths at 718 nm and 1470 nm. The wavelength at 1470 nm is within the water absorption range. While both classes seem absorb the same quantity of water, healthY has higher values in the NIR band at 718 nm. This waveband has been indicated as important for studying the growth of crops (Thenkabail et al., 2011). The other selected wavebands in the near-infrared region are important for distinguishing between vegetation of similar color. Another thing figure 4 reveals is that AMB diseased samples are generally more scattered around the plots than the healthy and nutrient deficient samples. The scattering effect was a result of the mixed reflectance spectra and varying ratios of colors that were discussed earlier.

Classification Based on PSO Optimal Bands

To avoid bias and ensure that the analysis results from this work would generalize well in an independent dataset, three-fold cross-validation was applied to the combined dataset from 2014 and 2015. Each class had an average of 274 samples. The classification results achieved at individual stages of the validation process were summed to create one confusion matrix. The PSO-selected features served as input for a support vector machine (SVM) classifier. Because SVM is inherently a binary classifier, a one-versus-one SVM approach was used in creating the training dataset, totaling ten SVM training models for the five classes. The kernel function of the group of classifiers was set to a third-order polynomial. A multi-class classification was first performed using the four selected wavebands for each class. The overall accuracy of the analysis was 100%, indicating that all the samples were correctly mapped to their respective classes.

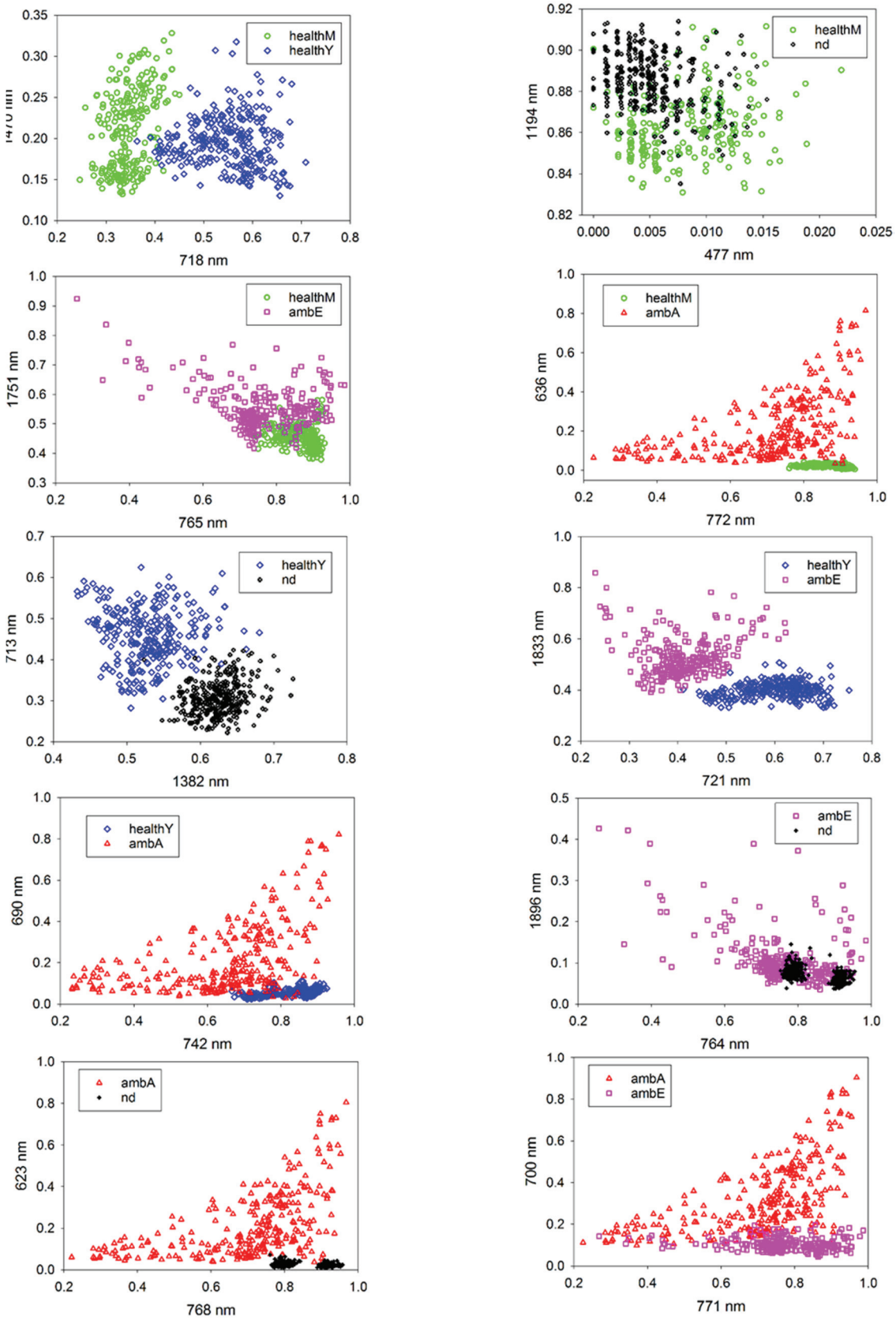

Figure 4. Two-waveband plots using PSO-selected spectral bands.

While some of the two-waveband plots in figure 4 showed some overlap between classes, a combination of all the features (four for each class) enhanced class separation even further and explains why there was perfect separation among the classes.

If the goal was to separate any two classes among the five,

Table 2. SVM classification result for each pair of classes for combined 2014 and 2015 spectroradiometer datasets using one PSO spectral feature (comprising two wavebands) and a normalized index created using both bands.

Class Pairs	Correctly Classified (%)	Misclassified (%)	Overall Accuracy (%)
healthM	96.6	3.4	98.2
healthY	99.6	0.4	
healthM	85.9	14.1	83.3
nd	81.4	18.6	
healthM	91.5	8.5	90.7
ambE	90	10	
healthM	100	0	99.4
ambA	98.9	1.1	
healthY	98.2	1.8	98.7
nd	99.1	0.9	
healthY	99.3	0.7	99.6
ambE	100	0	
healthY	97.1	2.9	96.6
ambA	96.1	3.9	
ambE	62.5	37.5	82.2
nd	98.4	1.6	
ambA	98.6	1.4	98.7
nd	98.7	1.3	
ambE	90.8	9.2	92.8
ambA	94.7	5.3	

the overall classification accuracy probably would not have been as high, given the overlap between classes in figure 4. In such a situation, the results may be improved by adding a normalized spectral index, or other indices, created using a combination of two pairs of PSO-selected bands for classification. To test how well the selected bands performed in discriminating between each pair of classes, another classification analysis was performed using two spectral bands in combination with a normalized index, and the results are given in table 2. As expected, the classification results between nd and healthM and between nd and ambE were the lowest due to their similarities in color. All nutrient deficient and early-stage AMB samples were previously healthy mature leaves and retained healthy regions even after developing stress symptoms. The highest overall accuracies were achieved between classes that were paired with either healthY or ambA. As mentioned earlier, these two classes contained regions that were partially or fully distinct in color (fig. 1).

Spectral Unmixing of Brown and Green Endmembers for ambE Samples

To enhance the separation between healthy leaves and early-stage AMB infected leaves, the spectra of brown blotches had to be separated from the asymptomatic regions. Endmember unmixing was achieved using a combination of a *k*NN classifier and the PSO algorithm. Before *k*NN classification was applied to the 19 color and texture feature images, a couple of preprocessing steps were applied to the original images. Hough transform was used in extracting regions of interest, after which green pixels were masked for easy segmentation of brown pixels using the ratio of the red channel to the green channel. After carefully studying the band ratio histogram, a threshold of 0.5 was chosen for green pixel masking. A summary of the steps is shown in figure 5 using the most infected ambE leaf as an example. Even after application of the Hough transform, there were still remnants of the white ring in some of the images. The green pixels were also not completely removed after applying the mask. As a result, five classes had to be defined for the extraction of brown pixels: green, brown, vein, background, and ring. The background class represented the previously extracted green pixels (masked regions). The *k*NN algorithm assigned a unique color to each class, and the light blue region in figure 5c indicates the region containing brown pixels.

Abundance estimation for the brown class was computed by dividing the sum of brown pixels by the total number of pixels within the region of interest. To simplify the computation, all other pixels representing veins and green regions were regarded as the green class. The abundance estimates for the brown class ranged roughly between 0.015 and 0.58 for the 2014 dataset and between 0.00006 and 0.21 for the 2015 dataset. The equation used in estimating the abundance (A) for each class is given below:

$$A = \frac{S_c}{S_T} \tag{6}$$

where S_c and S_T represent the total number of pixels enclosed by a particular class and the total number pixels enclosed by the entire region of interest, respectively.

Before the PSO algorithm was applied to the ambE dataset, its parameters were tuned using a set of hyperspectral data acquired from apple leaves. Abundances and mixed spectra were created by combining different ratios of endmember spectra. This made-up dataset was first applied to PSO so that the algorithm would produce accurate results for ambE endmember extraction. The parameters of the experimental dataset were tuned until an R^2 of at least 80% was achieved for each spectral band. To guarantee stable results, the algorithm was run 100 times for the full wavelength range. After PSO analysis, the mean reflectance signatures, shown in figure 6, were generated for the endmembers: early-stage AMB asymptomatic (ambEGr) and symptomatic (ambEBr). It can be seen from the spectral plot that the ambEBr class is very distinct in shape and can be easily distinguished from the healthy and asymptomatic classes at most wavelengths. The asymptomatic and healthM classes were fully overlapped in the visible range of the spectrum, but not in the NIR range because this region is effective in distinguishing between objects of similar colors. As with the spectra shown in figure 3, healthY is distinguishable from the other green classes in the green region of the spectrum. Overall, the results from the PSO analysis proved that mixed spectra acquired from a system, such as the one used in this study, can be separated and more efficiently analyzed.

Quantitative Analysis of ambE Samples

The combined normalized spectral data from 2014 and 2015 were used in building AMB prediction models, and the abundances estimated using the *k*NN classifier served as the ambE infection rate. A total of 20 healthy samples were included in the analysis to represent zero disease infection, while all 261 ambE samples were used. Table 3 shows some statistical measures that were computed for early-stage disease analysis. A total of 140 samples had disease severity

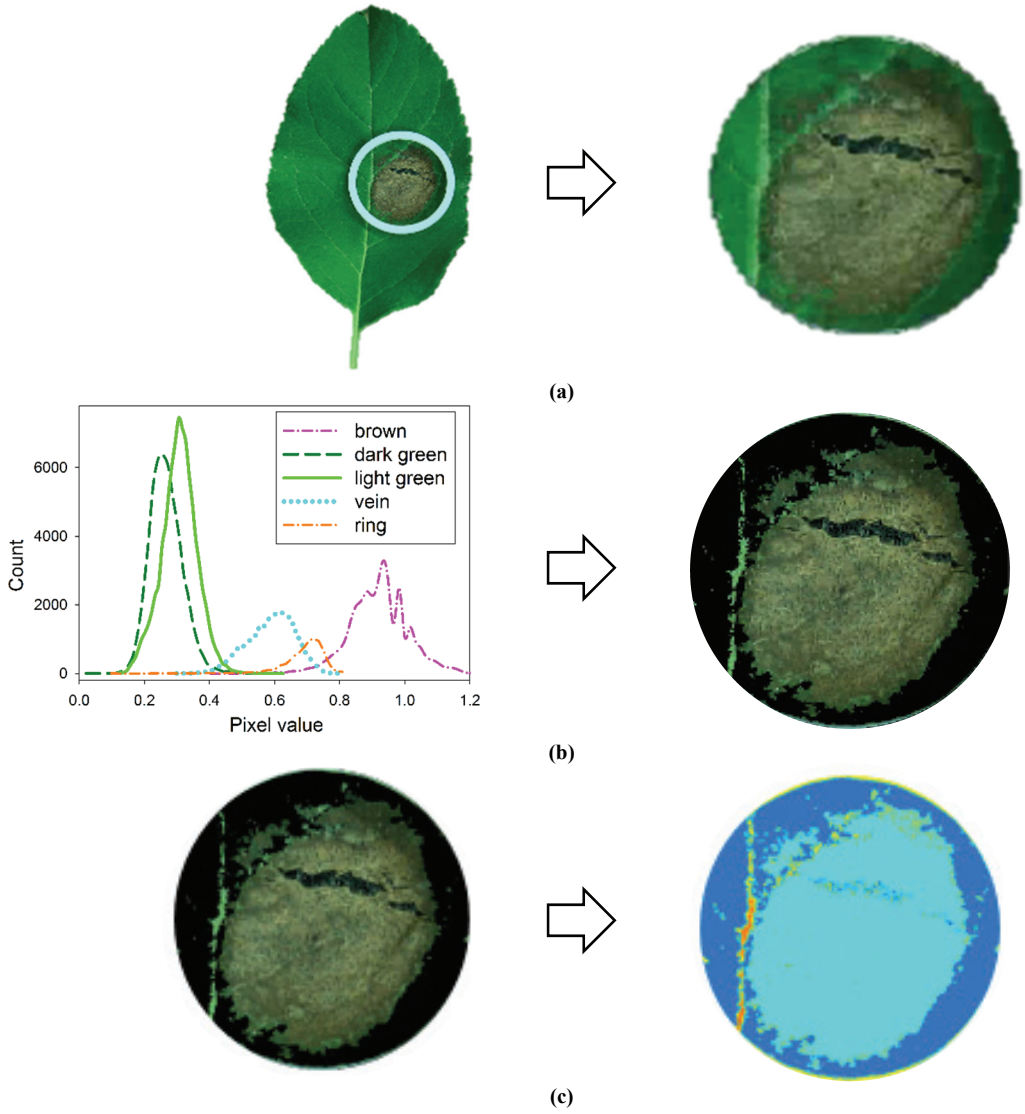

Figure 5. Steps taken in segmenting early-stage AMB diseased color images: (a) extraction of region of interest using circle Hough transform, (b) masking of green pixels using ratio of red and green channels, and (c) *k*NN image segmentation. The light blue color indicates the location of brown pixels, with a total abundance estimation of about 0.58.

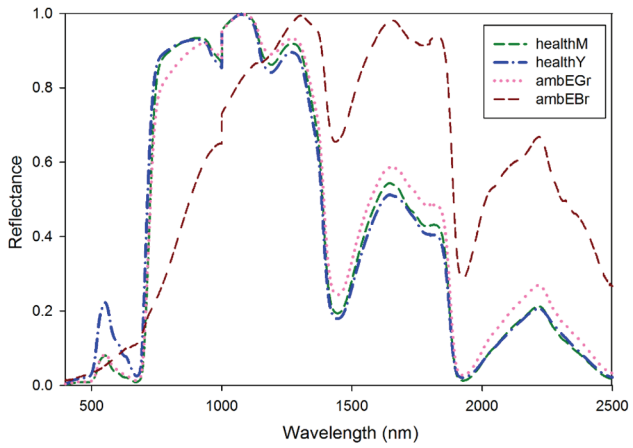

Figure 6. Mean normalized reflectance spectra of healthy mature (healthM), healthy young (healthY), early-stage AMB asymptomatic (ambEGr), and early-stage AMB symptomatic (ambEBr) classes for the combined 2014 and 2015 spectroradiometer datasets. Asymptomatic and symptomatic endmembers were calculated using a combination of *k*NN abundance estimation and PSO methods.

levels less than 4%, 92 samples had severity levels between 4% and 15%, while 49 samples had severity levels greater than 15%. Disease severity ranged between 0% and about 58%.

A PLSR analysis was performed using six components, and the regression coefficients (BETA) were plotted against the full wavelength range for extracting the bands with the highest discriminatory power. As was done by Jones et al. (2010), a threshold of the absolute value of 0.005 was set, and any spectral band above this limit was retained. A total of 766 bands from the original 2151 spectral bands met this criterion and were extracted for further analysis using SMLR. The dataset contained a total of 281 samples; two-thirds of the samples were used as calibration data, while one-third was used for validation. The stepping method criteria were set to a p-value of 0.05 for entry and a p-value of 0.1 for removal. Eight bands were selected based on the p-values: 671 nm, 1068 nm, 368 nm,

Table 3. Statistical analysis results of AMB disease severity for 20 healthy and 261 early-stage AMB diseased samples from the 2014 and 2015 datasets (SD is standard deviation).

Severity (%)	No. of Samples	Min. (%)	Max. (%)	Mean (%)	SD (%)
<4	140	0	3.9	1.4	1.2
4 to 15	92	4.1	14.8	8.5	3.2
>15	49	15.1	57.9	23.5	9.1

Table 4. SMLR prediction model summary for early-stage AMB detection. The letter R in the first column represents reflectance, while its subscript represents the wavelength in nanometers.

Variable	Estimate	SE	p-Value
Intercept	2.95	0.63	5.07E-06
R_{671}	14.24	4.32	1.20E-03
R_{1068}	-2.67	0.65	6.82E-05
R_{368}	-1.50	0.42	4.98E-04
R_{1658}	0.64	0.09	2.11E-10
R_{972}	-0.61	0.11	9.64E-08
R_{2451}	-0.97	0.22	1.52E-05
R_{383}	-2.61	0.94	5.93E-03
R_{675}	-11.05	4.32	1.15E-02

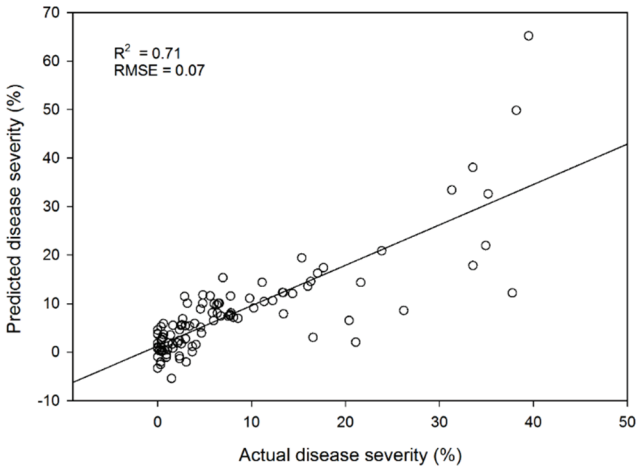

Figure 7. Predicted versus actual disease severity for validation dataset using a combination of PLSR and SMLR prediction models. The classes that were used in the analysis were healthy and early-stage AMB diseased samples from the 2014 and 2015 datasets.

1658 nm, 972 nm, 2451 nm, 383 nm, and 675 nm. The calibration dataset achieved an R^2 of 0.76 and an RMSE of 0.04, while the validation dataset achieved an R^2 of 0.71 and an RMSE of 0.07. The estimated coefficients for each variable in the SMLR model are given in the second column of table 4. The third and fourth columns of table 4 list the standard errors and p-values for each coefficient. As shown in figure 7, the model performed better at predicting leaves with lower disease severity than those with higher severity rates. This was a result of having more early-stage AMB samples than the advanced-stage in the dataset (table 3). Future work will include repeating the prediction analysis using more samples and exploring other efficient abundance estimation algorithms.

CONCLUSION

The main objective of this study was to detect AMB disease at the earliest stage possible using a subset of spectral bands from two consecutive years. A PSO-based band selection method was used to select ten spectral features created from 20 wavebands for discriminating five classes: healthy mature, healthy young, early-stage AMB diseased, advanced-stage AMB diseased, and nutrient deficient. Most of the bands selected to discriminate between the advanced stage of the disease and other classes were located in the red region of the spectrum. This range has been shown to be effective for detecting yellow plant disease symptoms (Jones et al., 2010). Spectral bands slightly beyond the red-edge region and close to the water absorption band at 1950 nm performed well in distinguishing between early-stage AMB infected leaves and other classes. Both regions are known to have a notable influence on plant diseases. Young and mature healthy leaves were efficiently separated using a combination of reflectance at 718 nm and 1470 nm. Most of the bands selected to separate green classes, including the nutrient deficient group, were found in the NIR region of the spectrum. The ten selected spectral features served as input for an SVM classifier for the classification of samples in all five classes. Results showed that the four features selected for each class were efficient in distinguishing it from the other classes and worked well for early detection of AMB. Based on the results achieved from this analysis, a multispectral camera using four spectral features consisting of eight bands can be applied for early detection of AMB disease in apple orchards.

Abundance estimation and spectral unmixing algorithms were developed using a combination of a *k*NN classifier and the PSO algorithm. PSO showed promising results in spectrally unmixing early-stage AMB diseased samples by creating symptomatic and asymptomatic endmembers from previously mixed spectra. Statistical prediction models (PLSR and SMLR) were combined to quantify various degrees of infection of early-stage AMB diseased samples. Eight spectral bands, between the visible and near-infrared wavelength range, were identified as important in predicting disease severity, with an R^2 of 0.71. Future work will include collecting spectral reflectance information of leaves using a system that can acquire data on a per pixel basis, such as a hyperspectral imaging system, and investigating more efficient methods for disease abundance estimation, spectral unmixing, and regression analysis. Overall, the results from this work indicate the potential of using spectroscopic technology as an efficient and non-invasive tool for early diagnosis of AMB disease.

ACKNOWLEDGEMENTS

The authors would like to thank the Rural Development Administration of South Korea for supporting this research. We would also like to thank Dr. Paul Gader, Dr. John Schueller, and all the members of the Precision Agriculture Laboratory at the University of Florida for their assistance in this study.

REFERENCES

Bruce, L. M., Koger, C. H., & Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Trans. Geosci. Remote Sensing, 40*(10), 2331-2338. https://doi.org/10.1109/TGRS.2002.804721

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., ... Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *Intl. J. Food Microbiol., 144*(1), 64-71. https://doi.org/10.1016/j.ijfoodmicro.2010.08.001

Eberhart, R. C., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proc. 6th Intl. Symp. Micro Machine and Human Science*, (pp. 39-43). Piscataway, NJ: IEEE. https://doi.org/10.1109/MHS.1995.494215

EPPO. (2013). *Diplocarpon mali* (anamorph: *Marssonina coronaria*). Paris, France: European and Mediterranean Plant Protection Organization. Retrieved from https://www.eppo.int/QUARANTINE/Alert_List/fungi/Diplocarpon_mali.htm

FAO. (2013). Global fruit production in 2013, by variety (in million metric tons). Rome, Italy: United Nations FAO. Retrieved from http://www.statista.com/statistics/264001/worldwide-production-of-fruit-by-variety/

Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos-Herrero, C., Moltó, E., & Blasco, J. (2008). Hyperspectral system for early detection of rottenness caused by *Penicillium digitatum* in mandarins. *J. Food Eng., 89*(1), 80-86. https://doi.org/10.1016/j.jfoodeng.2008.04.009

Graeff, S., Link, J., & Claupein, W. (2006). Identification of powdery mildew (*Erysiphe graminis* sp. tritici) and take-all disease (*Gaeumannomyces graminis* sp. tritici) in wheat (*Triticum aestivum* L.) by means of leaf reflectance measurements. *Open Life Sci., 1*(2), 275-288. https://doi.org/10.2478/s11535-006-0020-8

Harada, Y., Sawamura, K., & Konno, K. (1974). *Diplocarpon mali*, sp. nov., the perfect state of apple blotch fungus *Marssonina coronaria*. *Japanese J. Phytopathol., 40*(5), 412-418. https://doi.org/10.3186/jjphytopath.40.412

Jia, X., & Richards, J. A. (1999). Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. Geosci. Remote Sensing, 37*(1), 538-542. https://doi.org/10.1109/36.739109

Jones, C. D., Jones, J. B., & Lee, W. S. (2010). Diagnosis of bacterial spot of tomato using spectral signatures. *Comput. Electron. Agric., 74*(2), 329-335. https://doi.org/10.1016/j.compag.2010.09.008

Katti, A. R., Lee, W. S., Ehsani, R., & Yang, C. (2015). Band selection using forward feature selection algorithm for citrus Huanglongbing disease detection. *J. Biosyst. Eng., 40*(4), 417-427. https://doi.org/10.5307/JBE.2015.40.4.417

Lee, C.-H., Lee, S.-Y., Jung, H.-Y., & Kim, J.-H. (2012). The application of optical coherence tomography in the diagnosis of *Marssonina* blotch in apple leaves. *J. Optic. Soc. Korea, 16*(2), 133-140. https://doi.org/10.3807/JOSK.2012.16.2.133

Lee, D.-H., Back, C.-G., Win, N. K., Choi, K.-H., Kim, K.-M., Kang, I.-K., ... Jung, H.-Y. (2011). Biological characterization of *Marssonina coronaria* associated with apple blotch disease. *Mycobiol., 39*(3), 200-205. https://doi.org/10.5941/MYCO.2011.39.3.200

Lee, H.-T., & Shin, H.-D. (2000). Taxonomic studies on the genus *Marssonina* in Korea. *Mycobiol., 28*(1), 39-46.

Li, H., Lv, X., Wang, J., Li, J., Yang, H., & Qin, Y. (2007). Quantitative determination of soybean meal content in compound feeds: Comparison of near-infrared spectroscopy and real-time PCR. *Anal. Bioanal.Chem., 389*(7), 2313-2322. https://doi.org/10.1007/s00216-007-1624-1

Martínez-Usó, Pla, F., Sotoca, J. M., & García-Sevilla, P. (2007). Clustering-based hyperspectral band selection using information measures. *IEEE Trans. Geosci. Remote Sensing, 45*(12), 4158-4171. https://doi.org/10.1109/TGRS.2007.904951

Omran, M. G., Engelbrecht, A. P., & Salman, A. (2006). Particle swarm optimization for pattern recognition and image processing. In *Swarm intelligence in data mining* (pp. 125-151). Berlin, Germany: Springer.

Qin, J., Burks, T. F., Kim, M. S., Chao, K., & Ritenour, M. A. (2008). Citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. *Sens. Instrum. Food Qual. Saf., 2*(3), 168-177. https://doi.org/10.1007/s11694-008-9043-3

Roggo, Y., Duponchel, L., Noe, B., & Huvenne, J.-P. (2002). Sucrose content determination of sugar beets by near-infrared reflectance spectroscopy: Comparison of calibration methods and calibration transfer. *J. Near Infrared Spectrosc., 10*(2), 137-150. https://doi.org/10.1255/jnirs.330

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric., 72*(1), 1-13. https://doi.org/10.1016/j.compag.2010.02.007

Tamietti, G., & Matta, A. (2003). First report of leaf blotch caused by *Marssonina coronaria* on apple in Italy. *Plant Disease, 87*(8), 1005-1005. https://doi.org/10.1094/PDIS.2003.87.8.1005B

Tanaka, S., Kamegawa, N., Ito, S., & Kameya Iwaki, M. (2000). Detection of thiophanate-methyl-resistant strains in *Diplocarpon mali*, causal fungus of apple [*Malus pumila*] blotch. *J. General Plant Pathol. (Japan), 66*(1), 82-85. https://doi.org/10.1007/PL00012926

Thenkabail, A., Lyon, P. S., & Huete, J. G. (2011). *Hyperspectral remote sensing of vegetation.* Boca Raton, FL: CRC Press. https://doi.org/10.1201/b11222

Wang, H., & Angelopoulou, E. (2006). Sensor band selection for multispectral imaging via average normalized information. *J. Real-Time Image Proc., 1*(2), 109-121. https://doi.org/10.1007/s11554-006-0014-9

Xu, H. R., Ying, Y. B., Fu, X. P., & Zhu, S. P. (2007). Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosyst. Eng., 96*(4), 447-454. https://doi.org/10.1016/j.biosystemseng.2007.01.008

Yang, C., Lee, W. S., & Williamson, J. G. (2012). Classification of blueberry fruit and leaves based on spectral signatures. *Biosyst. Eng., 113*(4), 351-362. https://doi.org/10.1016/j.biosystemseng.2012.09.009

Yin, L., Li, M., Ke, X., Li, C., Zou, Y., Liang, D., & Ma, F. (2013). Evaluation of *Malus* germplasm resistance to *Marssonina* apple blotch. *European J. Plant Pathol., 136*(3), 597-602. https://doi.org/10.1007/s10658-013-0190-y

Zhang, B., Sun, X., Gao, L., & Yang, C. (2011). Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm. *IEEE Trans. Geosci. Remote Sensing, 49*(11), 4173-4176. https://doi.org/10.1109/TGRS.2011.2131145

Zhang, M., Qin, Z., & Liu, X. (2005). Remote sensed spectral imagery to detect late blight in field tomatoes. *Prec. Agric., 6*(6), 489-508. https://doi.org/10.1007/s11119-005-5640-x