Comparing Apples and Oranges: Hyperspectral Imaging Techniques for Fruit Detection and Determination of Plant Health

NSF Summer Undergraduate Fellowship in Sensor Technologies
DaVonne Henry, Sunfest Fellow (Mechanical Eng.) Carnegie Mellon University

Advisors: Camillo J. Taylor, Ph. D & Jnaneshwar Das, Ph. D

ABSTRACT

Remote sensing has grown to be an attractive strategy for monitoring agriculture in places such as orchards and farms because it is nondestructive and a quick and simple method of data acquisition. The aim of this research is to use hyperspectral imaging techniques to estimate crop yield and determine plant health based on data taken from the vegetation canopy, which consists of the visible, outer layer of vegetation on a tree. To determine the viability of this method, an Ocean Optics USB 2000+ Vis-NIR spectrometer was used to collect preliminary data on leaves and fruits. Reflectance measurements were taken for wavelengths ranging from 400-1000nm. To analyze the data, strategies involve using vegetation indices such as NDVI, PRI, and MCARI and the application of machine learning techniques to determine how to efficiently and accurately represent remotely sensed data. Using a quadratic Support Vector Machine (SVM) on a subset of reflectance data from fruit, a detection accuracy of over 80% and it was determined that PRI and NPCI vegetation indices can be used for fruit detection.

Table of Contents

1 Introduction 3
2 Background.... 3
2.1 Characteristic Leaf Reflectance Spectrum 3
2.2 Vegetation Indices.... 4
3 Materials and Methods 6
3.1 Comparison of Live and Dead Leaves 6
3.2 Simulation of Variability.... 7
3.3 Collection of Fruit Samples.... 8
4 Results 8
4.1 Characteristic Spectra of Live versus Dead Leaf 8
4.2 Variance in Reflectance Data 9
4.3 Classification Using a Support Vector Machine 10
4.4 Comparison of Vegetation Indices.... 11
4.5 Use of Vegetation Indices to Detect Stress 12
5 Discussion and Conclusions 13
6 Recommendations 13
7 Acknowledgements 14
8 References 14

1 INTRODUCTION

This research aims to tackle two important challenges in the field of agriculture. First, it is often difficult to accurately estimate the projected yield for a season. Second, given the scale of many farms, assessing biological or abiotic stress on a region cannot be feasibly done by humans. For these reasons, our aim is to use remote-sensing techniques coupled with unmanned aerial vehicles (UAV's) to efficiently collect data.

2 BACKGROUND

Because of its nondestructive nature and its ability to collect data over a wide region in a short amount of time, aerial hyperspectral imaging is seen as an attractive method for determining yield and plant health in orchards and crop fields [1]. Many researchers have previously used reflectance data to estimate crop yield with some success [2]. However, not much work outside of that of Xujun Ye et al. was done on orchard crops.

2.1 Characteristic Leaf Reflectance Spectrum

Reflectance in leaves can be attributed mainly to structural and physiological conditions within the leaf [3-4]. In general, reflectance in the visible region tends to be caused by physiological conditions such as water stress, while reflectance in the infra-red (IR) spectral region tends to be caused by leaf structure. The reflectance spectrum of a leaf is characterized by relatively low reflectance in the visible region and significantly higher reflectance in the near-infrared (NIR) region. The green color of most leaves can be explained by a peak around 550 nm [5]. Also characteristic of a leaf spectrum is the so-called "red edge" which is a positive, high slope region near 700 nm. This is caused by high absorption by chlorophyll *a* at 670 nm and high internal scattering of NIR radiation by the leaf structure [6].

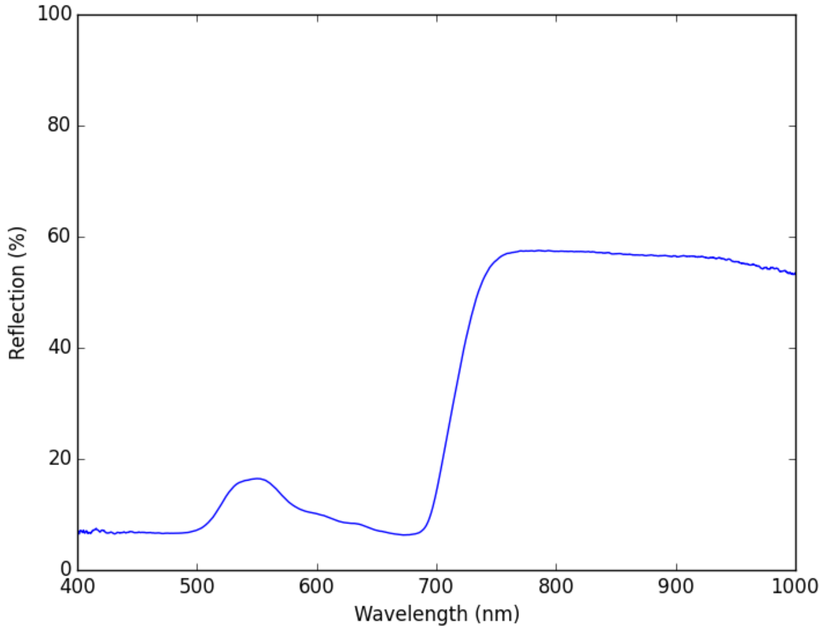


Figure 1: Example reflectance spectrum of a live leaf.

2.2 Vegetation Indices

There is a challenge in analyzing hyperspectral data because of the large amount of data taken and the large amounts of collinearity between adjacent wavelengths [2]. The simplest method of reducing reflectance data into comparable statistics is by calculation vegetation indices (VIs). These indices are functions that take two or more wavelengths from a reflectance spectrum and perform simple arithmetic on the reflectance values to give a number. Two commonly used indices are the Photochemical Reflectance Index (PRI) and the Normalized Differential Vegetation Index (NDVI). While there is some small variance in what wavelengths are reported for these calculations, for these two measures, we used the formulae used by Rascher et al. [4]:

$$PRI = \frac{R_{531} - R_{570}}{R_{531} + R_{570}}$$

$$NDVI = \frac{R_{780} - R_{670}}{R_{780} + R_{670}}$$

Where R_λ stands for the reflectance at wavelength λ.

In addition to these two, four other indices were used: The modified chlorophyll absorption in reflectance index (MCARI), and the optimized soil-adjusted vegetation index (OSAVI) as described by Daugtry et al [7], the Transformed Chlorophyll Absorption in Reflectance Index (TCARI) described by Haboudane et al. [8], and the normalized pigments chlorophyll ratio index (NPCI) described by Peñuelas et al. [9]

$$MCARI = [(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})]\frac{R_{700}}{R_{670}}$$

$$TCARI = 3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})\left(\frac{R_{700}}{R_{670}}\right)]$$

$$OSAVI = \frac{(1 + 0.16)(R_{800} - R_{670})}{R_{800} + R_{670} + 0.16}$$

$$NPCI = \frac{R_{680} - R_{430}}{R_{680} + R_{430}}$$

Ye et al. and Uno et al. determined independently that while vegetation indices such as the ones described above can do a reasonably good job at determining plant health, when trying to apply them to predicting crop yield, they do not perform very well [1-2]. These teams both continued their work to predict crop yields. While Uno's team could do a fair job at prediction corn yields, Ye's work with citrus in Japan was far less successful. In the same research with corn, Uno et al. also applied principal component analysis (PCA) and artificial neural networks (ANN) to prediction models with significantly better yield prediction results with the same corn data. The work of Ye et al. also continued by applying multiple linear regression models to their citrus data. This produced significantly better results than vegetation indices. In an earlier work, Ye et al. used citrus plants and hyperspectral data and attempted to use neural networks to predict yield [10]. Their results were mixed, but seemed to suggest that the accuracy their method was largely dependent on what time during the growing season measurements were made.

In our research, we hope to start with a modular spectrometer to take preliminary reflectance data on both leaves and fruits. Eventually we would like to mount a hyperspectral camera onto a UAV to automate the assessment of crop health and estimation of crop yield in apple orchards.

3 MATERIALS AND METHODS

To capture the reflectance spectra of the samples, an Ocean Optics 2000+ Vis-NIR spectrometer was used. Connected to the spectrometer was a fiber optic cable with a collimating lens attachment, which was well-suited for open-air experiments since the lens allowed for a reasonably small sampling area. All experiments were standardized with a Labsphere white reflectance standard, and the light source was an incandescent light bulb that could provide broad spectrum light.

3.1 Comparison of Live and Dead Leaves

Samples of superficially healthy (green) leaves were picked off of plants grown and maintained on the campus of the University of Pennsylvania, as well as dead (brown) leaves and a few leaves that appeared to be somewhere in-between with significant yellow and brown patches. The lamp was placed in the opening of a box while the collimating lens attachment was fixed and focused onto a point at the base of the box. Reference spectra were gathered with the white reflectance standard, and background (dark) spectra were gathered after removing all samples from the box and turning off the light source. Before taking each spectrum, a piece of black cloth was draped over the apparatus to reduce the amount of ambient light entering the experiment.

The dead and dying leaves were imaged separately with a standardization (reference and background) taken between each new sample. A live leaf was put into the apparatus while the provided OceanView software took one reflectance image of the leaf every 10 minutes over the course of an hour as a simulation of water stress.

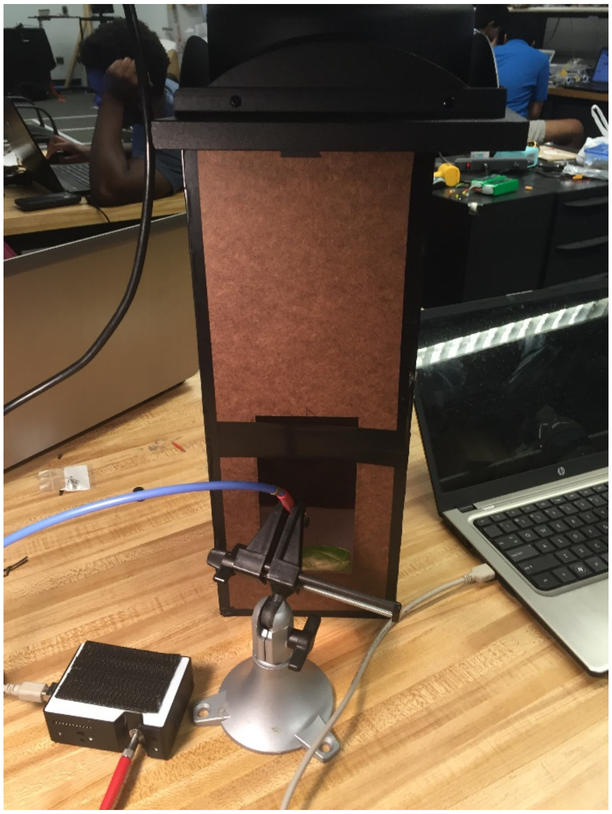


Figure 2: Experimental setup for comparison of live and dead leaves

3.2 Simulation of Variability

To gain an understanding of what kinds of variance could be seen when taking hyperspectral data of a canopy from a device that is not fixed in space, more leaf samples were taken. Similar to the method described above, a variety of healthy, dead, and yellowing leaves were sampled. Instead of using a box, samples were left in the open while the area was kept relatively dark to simulate taking data at night, which is a future goal for this work. In addition, all samples were placed on a dark cloth background. To take a background measurement, the light source was covered as opposed to switched off to preserve its thermal equilibrium. The collection was standardized before each new sample with reference and background measurements throughout the experiment.

“Standard” samples were taken, where the live leaves and dead leaves were clustered by their health and placed on the cloth background. Each of the yellowing leaves were taken individually. Afterwards, the clustered samples (including the yellowing leaves, which were grouped together for this part of the experiment) were manually manipulated to simulate varying heights and orientations of leaves.

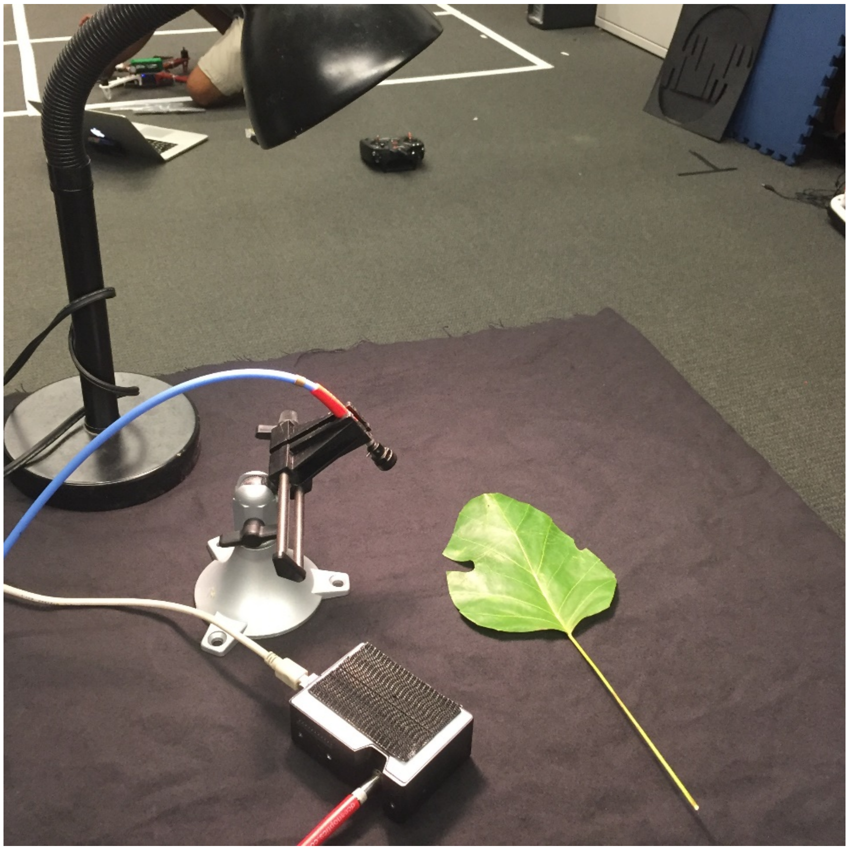


Figure 3: Experimental setup for simulating variability

3.3 Collection of Fruit Samples

A procedure very similar to the one described above in *Simulation of Variability* was used. A collection of fruits purchased from a local grocery store were sampled a few times in the apparatus. Multiple samples of each fruit were taken after manually manipulating the and orientation of the fruit to account for variability in the fruit's color, shape, and surface properties.

4 RESULTS

4.1 Characteristic Spectra of Live versus Dead Leaf

Live leaves tend to have a characteristically low reflectance in the visible range, save for a

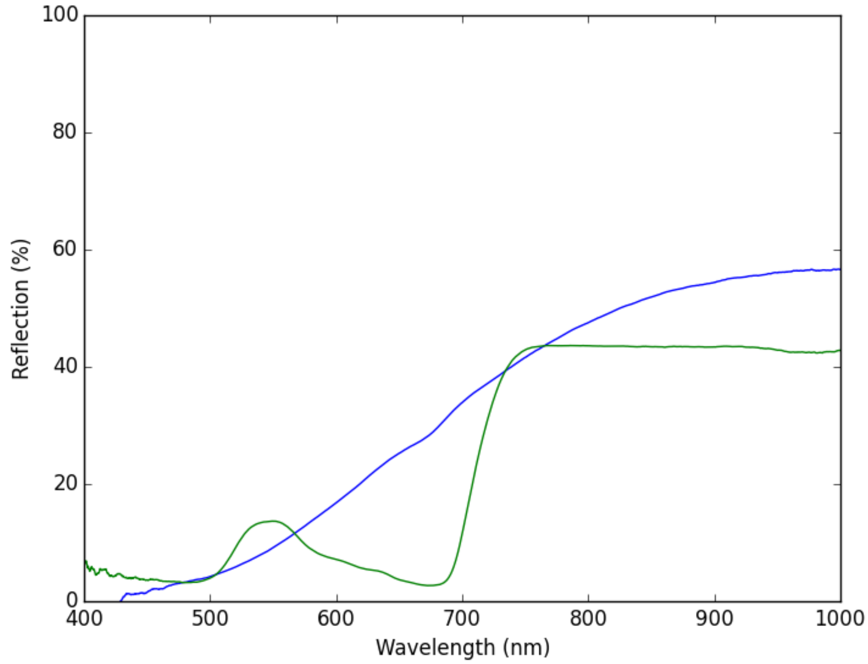


Figure 4: Comparison of a dead leaf (blue) and live leaf (green)

“green bump” around 570nm and the commonly-termed “red-edge” where the reflectance drastically increases between 670nm and 750nm. Dead leaves, on the other hand, have far less characteristic features, and instead have smooth curves that start low in the blue region and end up around equal to live leaves in the NIR region.

4.2 Variance in Reflectance Data

As a result of manually manipulating the position of the samples, significant variance was seen in the plots of the reflectance data of the samples. These tend to appear as vertical transformations of a similar shape, but sometimes different parts of the spectra appear to be shifted by different amounts.

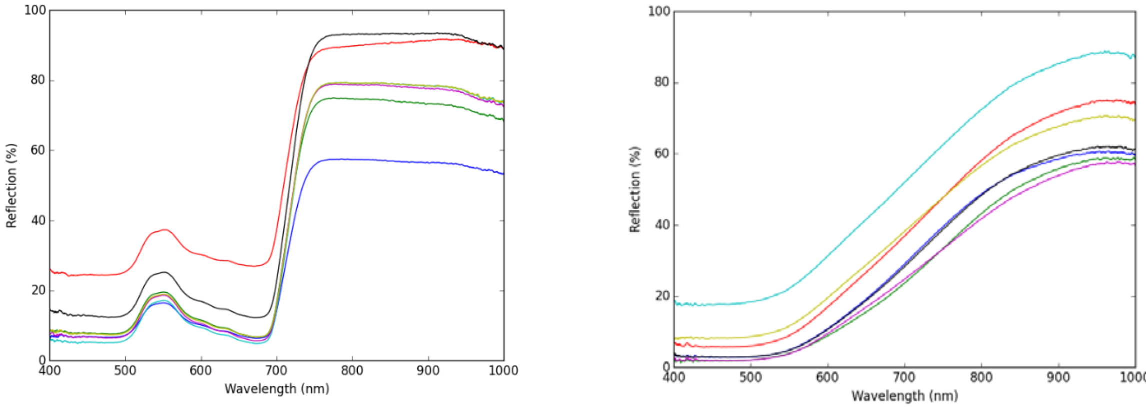


Figure 5: Variation in Data with (left) live leaves and (right) dead leaves

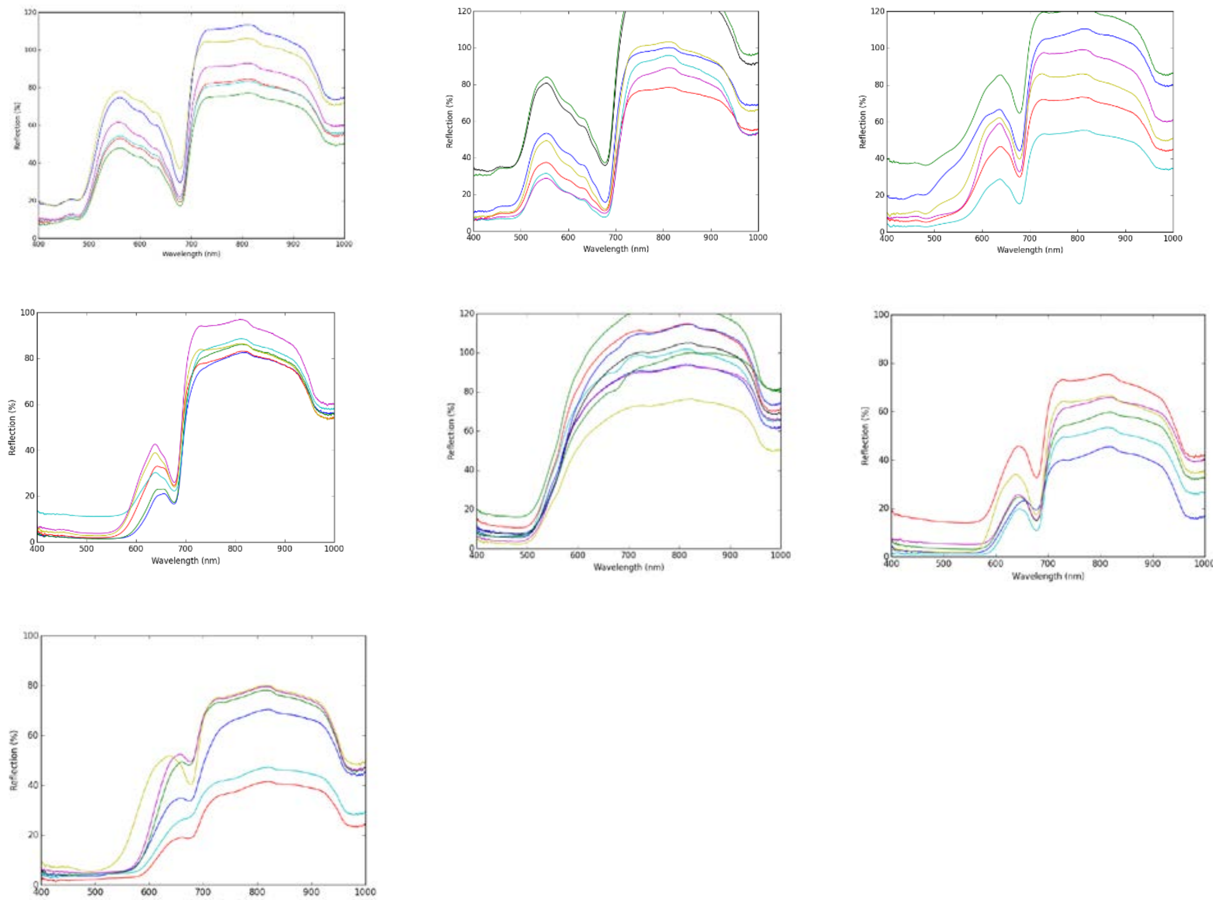


Figure 6: (left to right, top to bottom) Reflectance data of Golden Delicious, Granny Smith, Honeycrisp, Red Delicious, Orange, Plum and Strawberry

4.3 Classification Using a Support Vector Machine

A quadratic support-vector machine (SVM) was applied to the fruit data using the fruits as labels. An 81.2% overall accuracy was achieved using this model. The classes that caused the most issues for this model were Honeycrisp, Plum, and Strawberry. While the model correctly predicted strawberries with a true positive rating of 91% many false negatives for Honeycrisps and plums were incorrectly labelled as strawberries. Honeycrisps and plums had the poorest true-positive detection rates: 43% and 50% respectively.

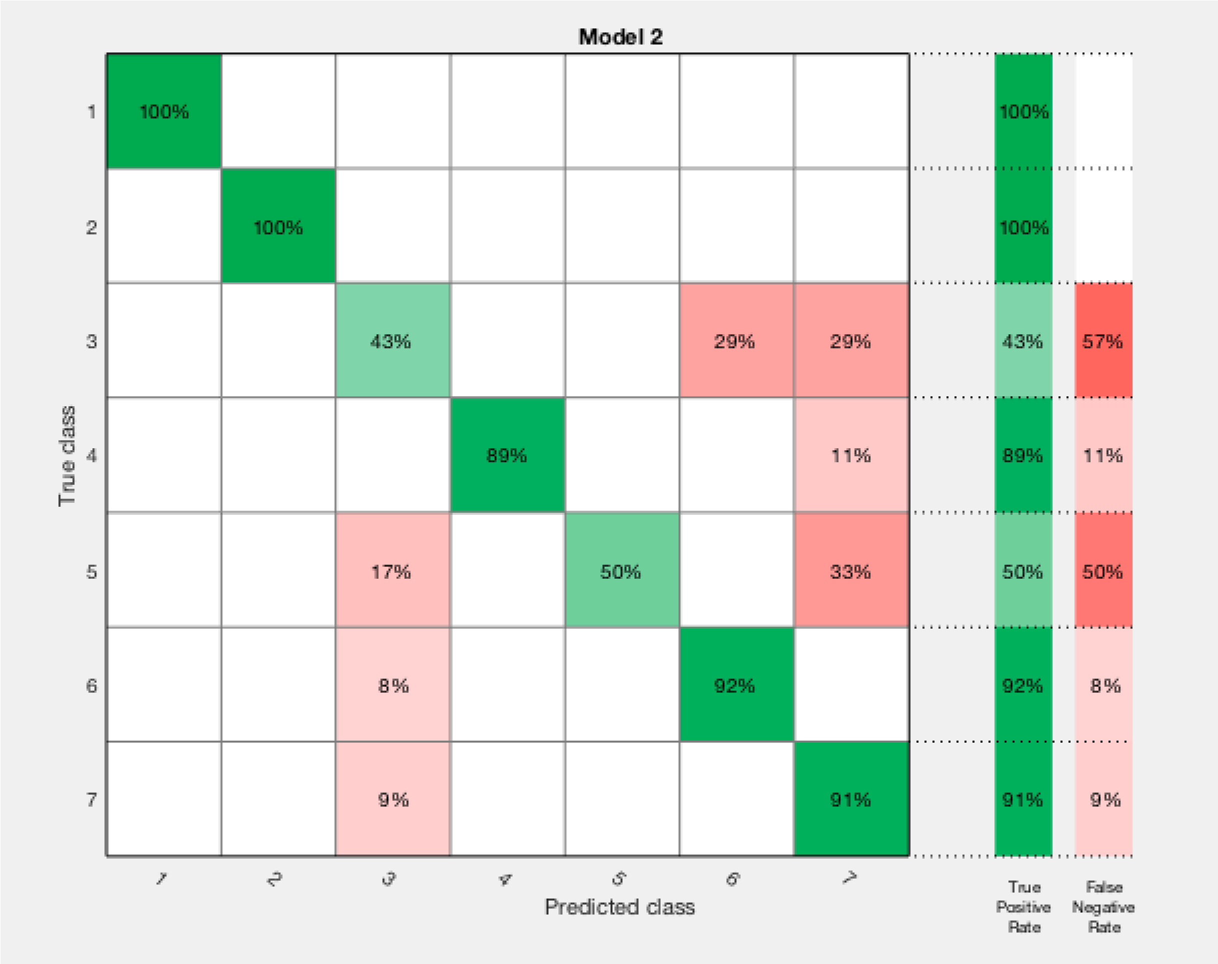


Figure 7: Confusion matrix for application of a quadratic SVM to fruit data. Classes are 1 – Golden Delicious, 2 – Granny Smith, 3 – Honeycrisp, 4 – Orange, 5 – Plum, 6 – Red Delicious, 7 - Strawberry

4.4 Comparison of Vegetation Indices

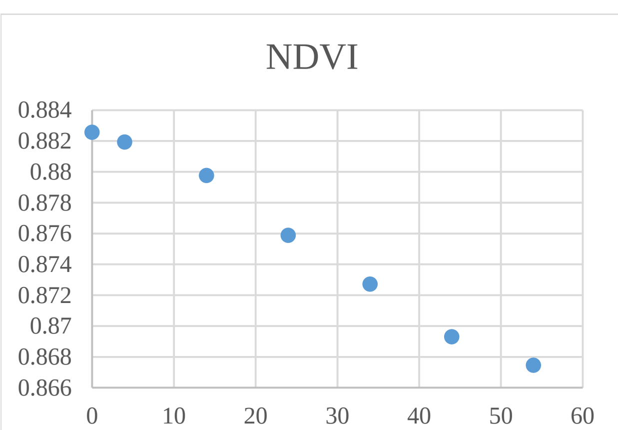
In trying to determine useful parameters for distinguishing between fruit and leaf, one index that stands out is NDVI which for live leaves averaged at about 0.81 with a standard deviation of 0.10. This is reasonably higher than the averages measured for many of the fruits. More promising still are PRI and NPCI. PRI was consistently positive for live leaves and negative for all other fruits except for black grapes, which averaged -0.0096 with a standard deviation of 0.021. NPCI, conversely, was consistently negative for live leaves and positive for everything else.

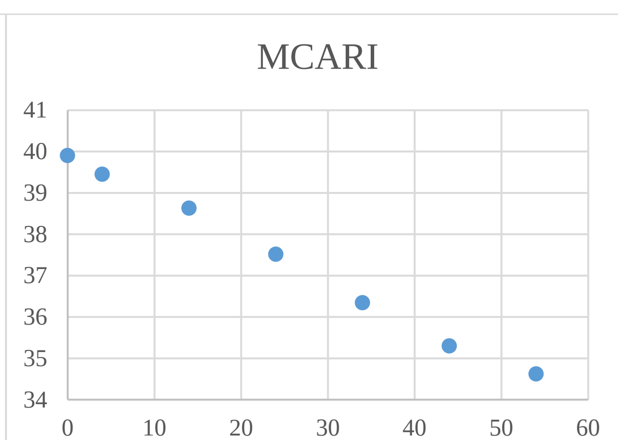
	NDVI	PRI	MCARI	TCARI	OSAVI	NPCI
Live Leaves	0.812419	0.049348	25.23622	33.08065	0.519919	-0.02851
Yellowing Leaves	0.549925	-0.15394	46.04491	40.46027	0.4189	0.594958
Dead Leaves	0.299663	-0.24115	1.16943	-0.46336	0.310968	0.711548
Black Grapes	0.750815	-0.00958	15.421	9.204434	0.556368	0.178077
Golden Delicious	0.538026	-0.06138	94.87345	112.2511	0.410153	0.329925
Granny Smith	0.70493	-0.00952	101.6263	99.07201	0.483228	0.155234
Green Grapes	0.594374	-0.01918	55.93051	57.81612	0.438334	0.144405
Honeycrisp	0.390704	-0.23781	45.05097	40.72342	0.330279	0.557848
Mango	0.778424	-0.14594	88.62648	29.85693	0.517042	0.227164
Orange	0.04918	-0.4228	-7.29976	-23.0881	0.062853	0.852615
Plum	0.496265	-0.10013	37.33859	16.92342	0.394133	0.59991
Red Delicious	0.539661	-0.17815	65.94121	29.80166	0.414314	0.587193
Strawberry	0.396925	-0.15227	13.02101	10.97943	0.329669	0.657702

Table 1: Averages of selected vegetation indices

4.5 Use of Vegetation Indices to Detect Stress

In an experiment where a freshly picked leaf was monitored over an hour, four of the used vegetation indices (NDVI, MCARI, TCARI, OSAVI) showed clear trends over time.





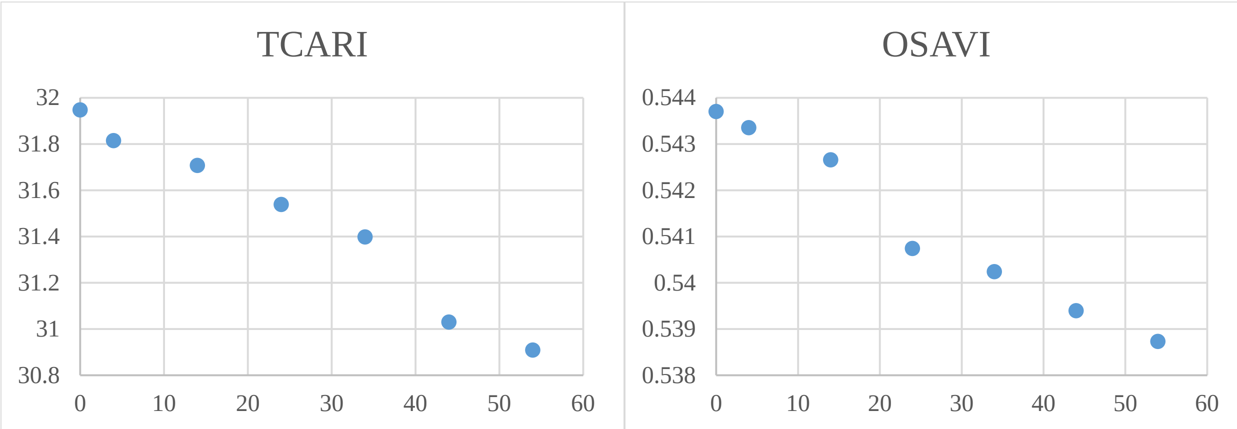


Figure 8: Comparison of four vegetation indices as a leaf under stress was monitored for an hour

5 DISCUSSION AND CONCLUSIONS

Hyperspectral imaging methods appear to be a viable method for determining both the health of plants and for distinguishing vegetation from crops. Using a preliminary set of labeled fruit spectra, a detection accuracy of 81.2% was achieved using a quadratic SVM. This shows that this method is robust against potential sources of error common to field measurements such as variation in sample orientation and position. Most error in this model was seen with redder fruits like plums and Honeycrisp apples.

Vegetation indices have also shown promise, both for fruit detection and for determination of plant health. For detection, both PRI and NPCI can provide a simple binary mask for distinguishing live leaves from other matter using the sign of the value. NDVI can also be used in this respect, but more work would need to be done to determine a threshold value to separate leaf from fruit.

Preliminary results show that vegetation indices can also be used to determine stress. Even over relatively short periods, NDVI, MCARI, TCARI, and OSAVI all showed strong, linear, negative correlations with time as a leaf sample dried out.

6 RECOMMENDATIONS

For further applications, we would like to do experiments on trees at night, bringing along a controlled light source to avoid variation from the sun’s position in the sky, and amount of ambient light. For these future experiments, it seems reasonable to mount an array of LEDs on a grayscale camera and use the two wavelengths of light for the NPCI calculation (680nm and

430nm) to illuminate the field for two separate images. An NPCI mask could then be applied to remove the leaves from the field, leaving only the fruit.

7 ACKNOWLEDGEMENTS

I would like to give a special acknowledgement to Shreyas Skandan for his help with the machine learning. I would also like to acknowledge Jnaneshwar Das and CJ Taylor for including me in their work and guiding me in my research.

8 REFERENCES

[1] Uno Y, Prasher SO, Lacroix R, Goel PK, Karimi Y, Viau, Patel RM (2005) Artificial neural networks to predict corn yield from compact airborne spectrographic imager data. Computers and Electronics in Agriculture 47: 149-161.

[2] Xujun Ye, Kenshi Sakai, Akira Sasao, Shin-ichi Asada (2008) Potential of airborne hyperspectral imagery to estimate fruit yield in citrus. Chemometrics and Intelligent Laboratory Systems 90: 132-144.

[3] Uwe Rascher, Caroline J. Nichol, Chris Small (2005) “Hyperspectral imaging of photosynthesis from the single leaf to the complex canopy - Understanding the spatiotemporal variations of photosynthesis within a drought-stressed tropical canopy” in New quality in environmental studies. Warsaw.

[4] Uwe Rascher, Caroline J. Nichol, Christopher Small, Leif Hendricks (2007) Monitoring spatio-temporal dynamics of photosynthesis with a portable hyperspectral imaging system. Photogrammetric Engineering & Remote Sensing 73: 45-56.

[5] Edward B. Knipling (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. 1: 155-159.

[6] Anatoly A. Gitelson, Mark N. Merzlyak (1996) Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. Journal of Plant Physiology 148: 494-500.

[7] Daughtry CST, Walthall CL, Kim MS, Brown de Colstoun E, McMurtrey, JE III (200) Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment 74: 229-239.

[8] Driss Haboudane, John R. Miller, Nicolas Tremblay, Pablo J. Zarco-Tejada, Louise Dextraze (2002) Integrated narrow-band vegetation indices for prediction of crop

chlorophyll content for application to precision agriculture. Remote Sensing of Environment 81: 416-426.

[9] Peñuelas J, Gamon JA, Fredeen AL, Merino J, Field CB (1994) Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sensing of Environment 48: 135-146

[10] Xujun Ye, Kenshi Sakai, Leroy Ortega Garciano, Shin-Ichi Asada, Akira Sasao (2006) Estimation of citrus yield from airborne hyperspectral images using a neural network model. Ecological Modelling 198: 426-432.