HYPERSPECTRAL IMAGE CLASSIFICATION FOR DETECTING FLOWERING IN MAIZE

by

Karoll Quijano

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Environmental and Ecological Engineering

Division of Environmental and Ecological Engineering

West Lafayette, Indiana

May 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Melba M. Crawford, Chair

School of Civil Engineering

Dr. Mitchell R. Tuinstra

College of Agriculture

Dr. John A. Howarter

School of Materials Engineering

Approved by:

Dr. John A. Howarter

To my family, the one biology gave me
and the one given by life.
I'm blessed for their love and unconditional support.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Melba Crawford, for her invaluable advice and guidance throughout the past years. I have profound respect and admiration for her work and commitment. I would also like to give special thanks to my other committee members, Dr. Mitchell Tuinstra and Dr. John Howarter for their valuable time and guidance towards my academic life.

I am extremely grateful to the TERRA project, staff and principal investigators for providing the opportunity to learn from them. To Andy Linvill, Eugine Glover, and all the people who assisted in the fields and made this project possible.

I would like to acknowledge my lab mates Ali Masjedi, Taojun Wang, Ruya Xu, and Claudia Avilés, from the Laboratory for Applications of Remote Sensing (LARS), and Yi Chun Li, Meghdad Hasheminasab, and Tian Zhou from the Digital Photogrammetry Research Group (DPRG). I really enjoyed working with you all and appreciate the time you took to share your knowledge with me.

I extend a special thanks to Shelby Gruss for her invaluable help in growing the maize experiment at the greenhouse, and Dr. Yang Yang, Dr. Augusto Souza, and Chris Hoagland from the Controlled Environment Phenotyping Facility (CEPF), for their assistance with the collection and processing of the data.

Finally, I would like to thank my family and friends for supporting me throughout my life. My Mom and Grandma, for teaching me the value of hard work and discipline. Thanks to the Nelsons for helping me discover love, joy, and compassion out of any moment. Romika K. for your invaluable friendship and companionship during this time at Purdue. My friends from Bogotá, Doris R., Paola R., Luisa H., Paula R., Laura P, Laura L., Genny T, and Bibian R. for your unrelenting support always. This would have never been possible without any of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SSM	Site-Specific Management
PA	Precision Agriculture
GPS	Global Positioning System
GNSS	Global Navigation Satellite System
UAV	Unmanned Aerial Vehicle
UAS	Unmanned Aerial System
AVIRIS	Airborne Visible/InfraRed Imaging Spectrometer
RGB	Red Green Blue
HSI	Hyperspectral Imaging
VNIR	Visible and Near InfraRed
SWIR	Short Wave InfraRed
FS	Feature Selection
FE	Feature Extraction
VI	Vegetation Index
MCARI	Modified Chlorophyll Adsorption Reflectance Index
PCA	Principal Component Analysis
LLE	Locally Linear Embedding
ISOMAP	Isometric Feature Mapping
SAM	Spectral Angle Mapper
SVM	Support Vector Machine
RBF	Radial Basis Function
CEPF	Controlled Environment Phenotyping Facility
DN	Digital Number
ELM	Empirical Line Method
IMU	Inertial Measurement Unit
FOV	Field of View
MNF	Minimum Noise-Fraction Transform
NIR	Near Infrared

ABSTRACT

Maize (*Zea mays L.*) is one of the most important crops worldwide for its critical importance in agriculture, economic stability, and food security. Many agricultural research and commercial breeding programs target the efficiency of this crop, seeking to increase productivity with fewer inputs and becoming more environmentally sustainable and resistant to impacts of climate and other external factors. For the purpose of analyzing the performance of the new varieties and management strategies, accurate and constant monitoring is crucial and yet, still performed mostly manually, becoming labor-intensive, time-consuming, and costly.

Flowering is one of the most important stages for maize, and many other grain crops, requiring close attention during this period. Any physical or biological negative impact in the tassel, as a reproductive organ, can have significant consequences to the overall grain development, resulting in production losses. Remote sensing observation technologies are currently seeking to close the gap in phenotyping in monitoring the development of the plants' geometric structure and chemistry-related responses over the growth and reproductive cycle.

For this thesis, remotely sensed hyperspectral imagery were collected, processed and, explored to detect tassels in maize crops. The data were acquired in both a controlled facility using an imaging conveyor, and from the fields using a PhenoRover (wheel-based platform) and a low altitude UAV. Two pixel-based classification experiments were performed on the original hyperspectral imagery (HSI) using Spectral Angle Mapper (SAM) and Support Vector Machine (SVM) supervised classifiers. Feature reduction methods, including Principal Component Analysis (PCA), Locally Linear Embedding (LLE), and Isometric Feature Mapping (Isomap) were also investigated, both to identify features for annotating the reference data and in conjunction with classification.

Collecting the data from different systems allowed the identification of strengths and weaknesses for each system and the associated tradeoffs. The controlled facility allowed stable lighting and very high spatial and spectral resolution, although it lacks on supplying information about the plants' interactions in field conditions. Contrarily, the in-field data from the PhenoRover and the UAV exposed the complications related to the plant's density within the plots and the variability in the lighting conditions due to long times of data collection required. The experiments implemented in this study successfully classified pixels as tassels for all images, performing better

with higher spatial resolution and in the controlled environment. For the SAM experiment, nonlinear feature extraction via Isomap was necessary to achieve good results, although at a significant computational expense. Dimension reduction did not improve results for the SVM classifier.

1. INTRODUCTION

Concerns about food security have become a major challenge for agricultural production around the world. As a result, the United Nations has declared it as one of the 17 Sustainable Development Goals [1]. In order to address the challenge of ensuring sustainable food production, plant scientists have focused on production strategies such as Site-Specific Management (SSM), Precision Agriculture (PA), and plant breeding. These strategies have doubled the productivity and are more environmentally friendly, producing more with fewer inputs [2].

Precision agriculture and Site-Specific Management use information from technology to monitor the outcomes, while contributing to a long-term sustainable production [2]. Plant scientists and researchers from different disciplines have studied and integrated the use of multiple technologies and devices for agricultural purposes and benefits. For example, the inclusion of global positioning systems GPS-GNSS, passive and active remote sensing technologies with high spatial and spectral resolution, and computational algorithms are being adopted both for experimental and production agriculture.

Moreover, plant breeders are actively using DNA sequencing that reduces the cost of genotyping in crop development [3]. To analyze the genetic connection with the performance of each variety, hundreds of field trials must be conducted in real-world environments, and key traits must be measured in different locations and at different times [4]. This field-based phenotyping adds many man-hours that could potentially be replaced by data acquired from remote sensing platforms with the ability to deliver the necessary throughput for large scale experiments.

The study of plant development and phenotyping has primarily focused attention toward certain attributes that influence final yield. For maize, and many other grains crops, flowering is one of the most important stages as it initiates the stage of reproduction. Any external stress, physical or biological, can cause plant damage and result in production losses. Traditional ways to monitor tasseling in maize are often subjective, time consuming, labor-intensive, and expensive. Additionally, estimating flowering time using calendar days may have disadvantages caused by environmental conditions. Due to the reduction of the personnel cost and availability of high spatial-spectral resolution imagery that can increase area coverage, monitoring flowering and tassel development via remote sensing is of increasing interest in agriculture, and is the focus of this study.

The main objective of this research is to evaluate whether the additional spectral information in hyperspectral data provide useful information for detecting tasseling. To achieve this goal, hyperspectral imagery was acquired in a controlled facility and in field conditions by a wheeled-based vehicle and a low altitude UAV. The exploration of the data includes different dimension reduction approaches, and Support Vector Machine (SVM) and Spectral Angle Mapper (SAM) as pixel-based classifiers.

The thesis is organized as follows:

Chapter 2 provides a review of the literature related to candidate remote sensing technologies, an overview of the phenology of maize, and the approaches that have been explored for tassel detection in maize and headed grains. Additionally, related work on classification and flowering detection on maize and other similar grains is reviewed.

Chapter 3 describes the experimental design used for acquiring, processing, and analyzing the data in this study.

Chapter 4 includes experimental results from datasets acquired by the Purdue TERRA team to evaluate the feasibility of the proposed approach.

Chapter 5 contains an overall evaluation of the research results, concluding remarks, and discussion of future work.

2. LITERATURE REVIEW AND BACKGROUND

This chapter provides background material related to remote sensing in agriculture, and particularly the use of hyperspectral imagery for image classification. A review of the life cycle of maize explains the importance of tasseling related to the final yield. Additionally, this chapter includes the foundations of the methodology that was used to conduct the data analysis in this research.

2.1 Remote Sensing for Vegetation Monitoring in Agriculture

Remote sensing involves the use of sensing technology to observe, collect data, and analyze information about an object, area or phenomenon of interest without being in contact with it [5]. The applications of optical remote sensing in agriculture rely on the basic phenomena in chemistry, physics, and biology, where the optical properties are related to morphological characteristics of the plants [1], identifying spectral features associated with plant growth conditions and stages, nutrient deficiency, pests and diseases, abiotic stress, and yield prediction [2], [6], [7]. Ground-based measurements are combined with data collected from sensors mounted on vehicles, Unmanned Aerial Systems (UAS), manned airborne platforms or spaceborne platforms, pursuing non-destructive monitoring methods of plant growth and development.

Spaceborne hyperspectral missions have been extremely limited, and the medium resolution imagery has been used primarily for land monitoring applications and agriculture. The Hyperion instrument on board on the Earth Observing-1 (EO-1) satellite, successfully acquired data for nearly two decades (2000-2017), collecting 220 bands with 10 nm spectral resolution and 30 m spatial resolution. Spaceborne missions typically have a low revisit time, making it difficult for monitoring crops. For passive sensors in satellites, the problem is exacerbated by cloud cover over the area of interest at the time of the overflight. As hyperspectral cameras became more affordable, manned aircraft, helicopters and zeppelins are serving as platforms for hyperspectral data acquisition. They offer adjustable spectral and spatial resolution as well as flexibility with repetition of the data acquisition; this temporal flexibility to adjust for environmental conditions is advantageous relative to spaceborne imagery. AVIRIS (Airborne Visible/InfraRed Imaging Spectrometer), a NASA airborne hyperspectral sensor provide images with pixel size of 4m and

224 bands with spectral resolution of 10 nm with a range from 380 nm to 2500 nm. Hyperspectral data are also available commercially, although on a limited basis, by companies such as HyMap and SpecTIR. The complications for manned aircraft-based sensors rely on the technical preparation and specific personnel required for each mission, making it expensive and operationally complicated [8].

Both sensors and UAV platforms have radically improved agricultural monitoring in recent years, particularly for agricultural applications requiring high spatial, and spectral resolution data over limited areas, with acquisitions targeted for favorable weather in a cost-effective way compared with aircraft, or satellite systems [7]. The information that UAV remote sensing provides enables farmers and managers to identify the plants' response to their local environment and make decisions for early detection, diagnosis and corrective actions of agricultural management problems [7]. The limitations for UAVs systems include payload size, power requirements, need for technical expertise for flight operations (particularly for advanced sensors), and large quantity of data to store, process, and analyze.

Other methods for remote sensing used in agriculture include wheel-based systems, serving to capture data at plot scale exploiting more advanced global positioning systems (GPS), providing data with a very high spatial and spectral resolution. The challenges with these systems are related to a) the low platform speeds when acquiring the data, often making near simultaneous acquisitions of data over large fields impossible for plot comparisons, b) wet soils and soil compaction, and platform/sensor vibration due to the terrain [9].

With a goal of developing crops with high yield potential, heat and drought tolerance and disease resistance, plant scientists and breeders seek to identify the best genotypes (or the candidate genes) [2]. An effective and efficient variety can produce higher yields than others with the same quantity of agricultural inputs (water, nutrients, fertilizers, pesticides, etc.), is tolerant to changing environmental conditions, and is resistant to biotic and abiotic stressors [9]. Each variety needs to be proven in field conditions, evaluating the performance with field trials, and observing the physical characteristics of the plants as a function of the interactions of genetics, the environment, and management practices (GxExM), as represented in (2.1) [4], [10]. This critical component is referred to as phenotyping, which is typically performed manually, requires many man-hours of labor, resources, and often involves destructive sampling. Phenotyping methods have improved

very slowly over the years and are expensive, time consuming, and limited to subjective analysis [3], [10].

$$Genotype \times Environment \times Management = Yield \quad (2.1)$$

Combining new platforms and sensors, remote sensing is playing a very important role in agriculture, providing the opportunity to monitor the development of the plants' geometric structure and chemistry-related responses over the growth cycle. These physiological and spectral traits, captured as changes occurring when plants mature, experiencing modifications in their chemical, physical and biological aspects, are an expression of the genetic factors [11], [12]. Whether the platform is UAV or wheel-based, remote sensing is beginning to fill the gap between genotyping and phenotyping in plant breeding, providing researchers the opportunity to monitor edaphic factors and crop status while measuring the environment and phenotype reactions with a non-invasive solution and the requirement of high spatial, spectral and temporal resolution data at a low cost.

2.1.1 Maize Phenology (life cycle)

Maize (*Zea mays L.*) is one of the most important crops internationally for commercial applications, including food, animal feed, biodiesel and fibers, among other uses [13]. In the U.S., maize is the most widely produced feed grain with a planted area of 89.1 million acres reported in 2018 [14]. Maize production is increasing at 1.6 % per year; this rate is insufficient for the projected required demand increase of 67 % needed by 2050 [11], unless changes can be made to accommodate environment or consumptions patterns. Research on this grain is critically important for agriculture, economic stability and food security on a global basis.

Maize plants are C4 carbon fixation, monoecious (Greek for "one household") grasses, meaning that they have separated male inflorescence (tassel) and female inflorescence (silk) on the same plant [15] (Figure 2.1). The tassel is located at the top of the main stem, with a central spike and 10 to 50 lateral spikelet flowers. Each spikelet contains a pair of glumes that contains two florets with 2-3 anthers, that after extrusion, shed the pollen necessary for reproduction. The silk is located in the middle of the stalk

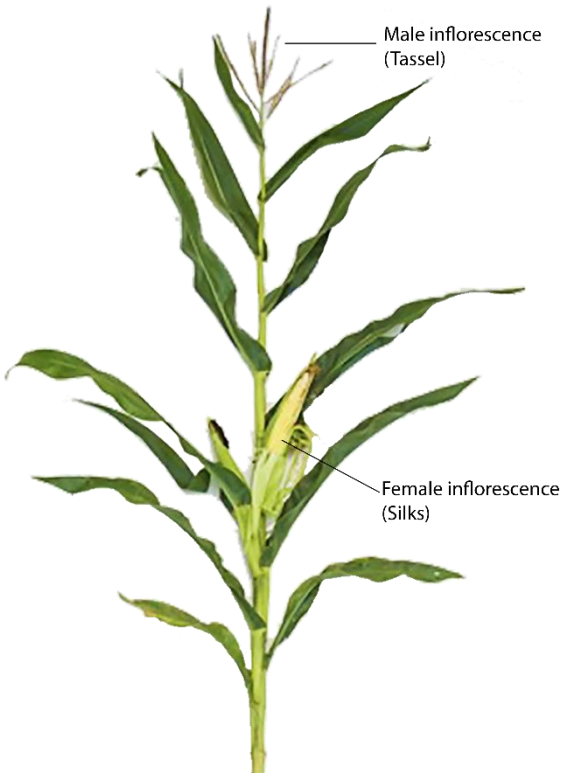

Figure 2.1. Maize plant with tassel (male inflorescence) and ear with silks (female inflorescence).

The maize growth is divided into the Vegetative Stage (VS) and Reproductive Stage (RS) (Figure 2.2). The VS starts when the plant is emerging from the soil surface and can continue for approximately 8 weeks (V1 to V16), ending when the tassel is fully developed (VT). At this point the plant has reached maximum height. The RS begins when silks are visible and pollination, when pollen grains contact the new moist silks (R1 to R6). The reproductive stage ends when the plants have reached physiological maturity and all the kernels have attained maximum dry weight.

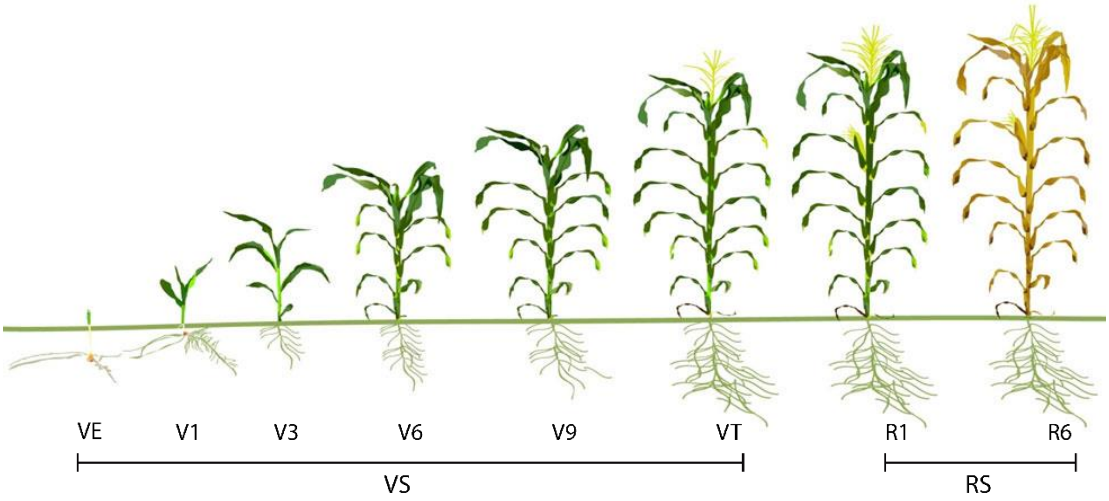

Figure 2.2. Growth stages of maize. Reprinted from Staging Corn Growth - Field Facts written by *Pioneer Agronomy Sciences*. Retrieved on April 2020, from https://www.pioneer.com/us/agronomy/staging_corn_growth.html

In the dynamics of plant growth and development, the emergence of the tassel (VT) changes the focus of the corn plant development, switching from vegetative to reproductive stage, redirecting the plant storage resources from growth to formation of the ear, which determines the grain production and weight at the end of the season [16]. At this stage of pollen shedding and silking, any physical or biological stress can cause more yield loss than at any other period in the crop's maturity [16]. To achieve high yields on commercial corn, adequate pollen production is an essential prerequisite [17].

Tasseling, when the central spike of the tassel is completely visible, usually occurs two to three days before silking, when silks have not yet emerged from the ear shoot. Anthesis, when 50% of the tassel is shedding pollen, occurs when about 80% of the tassel is visible [18]. At this point, the plant has reached full height and the pollen shed begins when the tassel has fully emerged from the whorl. This process initiates one of the most important stages of the reproductive crop development, corn pollination, and is what ultimately creates the grain. Any negative impact to a corn plant at this time can have significant consequences to overall grain production.

The number of flowers per plot is a relevant trait to measure in maize crops, as it is related directly to the final yield [17]. Current research on remote sensing for tassel detection exploits the high spatial resolution of RGB imagery using machine vision techniques, machine learning, and deep learning algorithms with data obtained in controlled facilities or laboratories [17]–[21]. Although very little research has been completed in the area of tassel detection for maize using

hyperspectral imagery in field conditions, in part due to the extremely high resolution of imagery that is required, it could be advantageous to exploit the chemical and biophysical changes of the plant in natural conditions with different management practices.

2.2 Hyperspectral Imaging (HSI)

The electromagnetic energy incident on earth occurs across a wide range of wavelengths, including visible light, radio waves, heat, ultraviolet rays, and X-rays. The interaction of the electromagnetic energy with objects on the Earth's surface involves absorption, occurring when radiation is absorbed into the target, transmission happening when radiation passes through the target, and reflection, when radiation is reflected from the target. Depending on the material and conditions of the target, the amount of energy absorbed, transmitted or reflected may be wavelength dependent. The spectral reflectance of an object as a function of wavelength is referred to as its spectral signature; the specific pattern of features reflected and absorbed provide insights chemistry-related to the characteristics of the object [5].

Current hyperspectral imaging (HSI) sensors commonly capture data from the visible and near infrared (VNIR) and short-wave infrared (SWIR), covering the range of the spectrum from ~400 nm to ~2500 nm with continuous and contiguous narrow bands (<10 nm) generating a data cube (Figure 2.3). The associated patterns of the spectral signatures provide capability to characterize materials and potentially distinguish land cover classes, vegetation state, and discriminate phenomena such as disease in agricultural classes, among others.

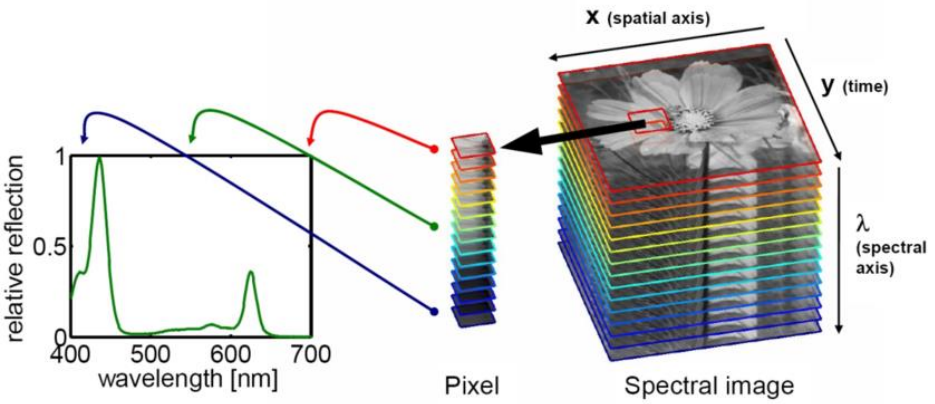

Figure 2.3. Hyperspectral image data cube representing reflectance from the VNIR range of spectrum. Adapted from "A Spectral Imaging System for Detection of Botrytis in Greenhouses", by Polder, Gerrit & Pekkeriet, Erik & Snikkers, Marco, *Conference "Sustainable Agriculture through ICT Innovation"*, Turin, Italy, 24-27 June 2013

The reflectance of plants is determined by morphological and chemical characteristics of leaves or other organs varying with plant type, water content within tissues, and stage of growth. Remote sensing of vegetation is mainly achieved by obtaining the reflectance signature from canopies using passive sensors [22]. The visible reflectance of the leaves and canopies is contributed by the plant pigments, chlorophylls (a and b), carotenoids and anthocyanin [8]. As Figure 2.4 shows, the red edge and near-infrared reflectance signatures are a characteristic contribution by the chlorophyll strong absorption pattern and leaf internal scattering of radiant energy.

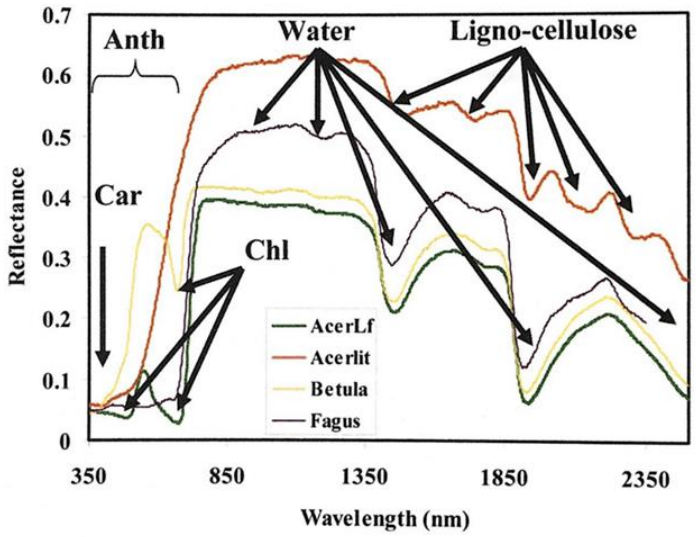

Figure 2.4. Reflectance spectra of leaves from a senesced birch (Betula), ornamental beech (Fagus), and healthy and fully senesced maple (AcerLf, Acerlit) illustrating carotenoid (Car), anthocyanin (Anth), chlorophyll (Chl), water, and ligno-cellulose absorptions. Reprinted from Hyperspectral Indices and image Classifications for Agriculture and Vegetation, in *Hyperspectral Remote Sensing of Vegetation*, vol. II, CRC Press, 2019. By P. S. Thenkabail, J. G. Lyon and A. Huete [3]

2.2.1 Feature Selection and Feature Extraction

It is tempting to say that the large number of spectral features in HSI automatically leads to a greater classification accuracy when classifying data. Unfortunately, HSI data have a lower signal-to-noise ratio than broad band sensors, and many bands are highly correlated, providing redundant information, and potentially degrading the classification of the data [8]. This is referred to as the Hughes phenomenon and is of concern for parametric classifiers which typically require estimation of the covariance matrix of the bands. Hughes indicates the importance of having a

correct dimensionality of data and complexity for a model; therefore, an increased dimensionality over the optimum will decrease the classification accuracy [23].

Feature selection (FS) and feature extraction (FE) techniques have been proposed to manage the feature space, reduce the data dimensionality, and to extract more meaningful features of the original HSI, along with reduction in computational costs. The goal for FS methods is to reduce the original data dimensionality by selecting the ideal minimum subset of features. This is inherently a combinatorial optimization problem, and most typically involves development of a model where features are added and/or removed. Common approaches to feature selection include Jeffries–Matusita (J–M) distance, Bhattacharyya distance, Mutual Information (MI), and signal-noise ratio [24].

Alternatively, FE methods seek to transform and/or project the original feature space onto a new space with small number of features. In this process the original set is presented in a new compressed version [8], [25]. Feature extraction approaches include knowledge-based, such as vegetation indices, and statistical approaches that can be unsupervised (linear or nonlinear) or supervised (parametric or nonparametric) combinations of the original feature. Some of the most widely applied unsupervised techniques include linear methods such as principal component analysis (PCA), independent component analysis (ICA), minimum noise fraction (MNF), and nonlinear approaches including isometric feature mapping (ISOMAP), and locally linear embedding (LLE). Supervised FE techniques include, for example, linear discriminant analysis (LDA), local Fisher discriminant analysis, and nonparametric discriminant analysis (NDA) [24].

To manage the feature space and dimensionality reduction via FS and FE, the following methods are investigated for the tassel classification problem.

2.2.1.1 Vegetation indices

Valuable information can be extracted from the large number of wavelengths, by exploiting absorption features in the signature via ratios of band differences and sums. Vegetation indices (VI), which are widely used in agricultural studies, are usually divided into three categories: structure, biochemistry, and plant physiology [8]. The first category measures properties such as green biomass, leaf area index (LAI), senescence, and fractional cover. The second category provides information on water content, pigments, nitrogen compounds, lignin and cellulose; and the third category can be used to evaluate changes in pigments, chlorophyll content, fluorescence

and leaf moisture. The continuous spectrum and narrow bands offered by HSI enhance the capability to characterize biophysical and chemistry-based properties in a more efficient way than broad-band indices which can be derived from multispectral imagery, gaining interest in the plant sciences community for high-throughput phenotyping [3].

A wide range of VIs have been created and are widely used for monitoring, analyzing and mapping temporal and spatial distributions of vegetation state, such as the normalized difference vegetation index (NDVI), used in [13] with a threshold of 0.1 as segmentation process to follow up the growth stages of cereal crops. Some VIs are proposed specifically to capture the photochemical processes associated with photosynthesis activity, such as light use efficiency or to estimate leaf pigment content [26], as for example the modified chlorophyll absorption in the reflectance index (MCARI) defined as shown in equation (2.2).

$$\boldsymbol{MCARI} = [(R_{700} - \mathrm{R}_{670}) - 0.2(\mathrm{R}_{700} - R_{550})]\left(\frac{R_{700}}{R_{670}}\right) \quad (2.2)$$

The MCARI indicates the relative abundance of a chlorophyll feature, compensating for non-photosynthesizing materials by quantifying the depth of the absorption at 670 nm relative to the reflectance at 550 nm and 700 nm [27]. This index is a resourceful method of vegetation feature extraction from soil and other backgrounds. To relate tassel condition and pollen release during flowering season, [28] used VI from canopy reflectance measured with a spectroradiometer.

2.2.1.2 Principal component analysis

Principal Component Analysis (PCA) is one of the most widely used unsupervised methods for dimensionality reduction and feature extraction in high dimensional data. PCA transforms the data through an axis rotation, in the direction of maximum variance [29]. Successive PCs are the linear combinations of the variables with maximum variance, which are orthogonal to the previously computed components. The total variance is typically represented in a small number of components, which are selected as "extracted variables". The principal components are the eigenvectors of the covariance matrix of the data, and the associated variance is represented in the corresponding eigenvalues. The PCA bands are orthogonal and have successively ordered variances. PCA is a global transformation, which does not accommodate unique patterns in individual classes, which may have local characteristics on a spectral neighborhood graph.

2.2.1.3 Isometric feature mapping

Isometric feature mapping (Isomap), is a manifold based approach for nonlinear dimensionality reduction of a set of high-dimensional data points [30]. It attempts to recover a low-dimensional nonlinear embedding structure by determining the k nearest neighbors and computing the shortest path distance between pairs of nodes on a neighborhood graph using Dijkstra's algorithm (Figure 2.5) [31].

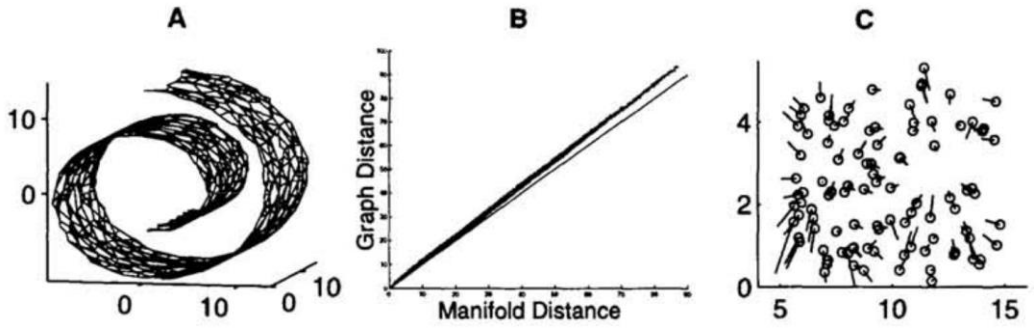

Figure 2.5. Steps to implement Isomap algorithm, (A) Discrete representation of manifold. (B) Correlation between measured graph and true distances. (C) Correspondence of recovered two-dimensional feature points. Reprinted from: 'Mapping a Manifold of Perceptual Observations," by J. B. Tenenbaum, 1998, *Advances in Neural Information Processing Systems 10*, MIT Press, pp. 682-688 [4]

Global manifold methods seek to maintain the fidelity of the overall topology of the data set at multiple scales of the data [32]. Generally, Isomap is less susceptible to overfit the data, which can be beneficial for classification, but is computationally intensive and does not exploit relationships in local neighborhoods of the graph.

2.2.1.4 Locally Linear Embedding

In 2000, Roweis and Sau [33] proposed an unsupervised learning algorithm that computes low dimensionality, while preserving neighborhood embeddings and geometry of highly nonlinear dimensional data. The LLE algorithm, illustrated in Figure 2.6, first identifies k nearest neighbors per data point i measured by Euclidean distance, then computes the weights W_{ij} minimizing the cost function, and subsequently, constructs the low dimensional vector that represents the global internal coordinates on the manifold, preserving the angles and geometry of the data. Because the weights are invariant to translation in the vector space, it is computationally efficient requiring sparse matrix computations only.

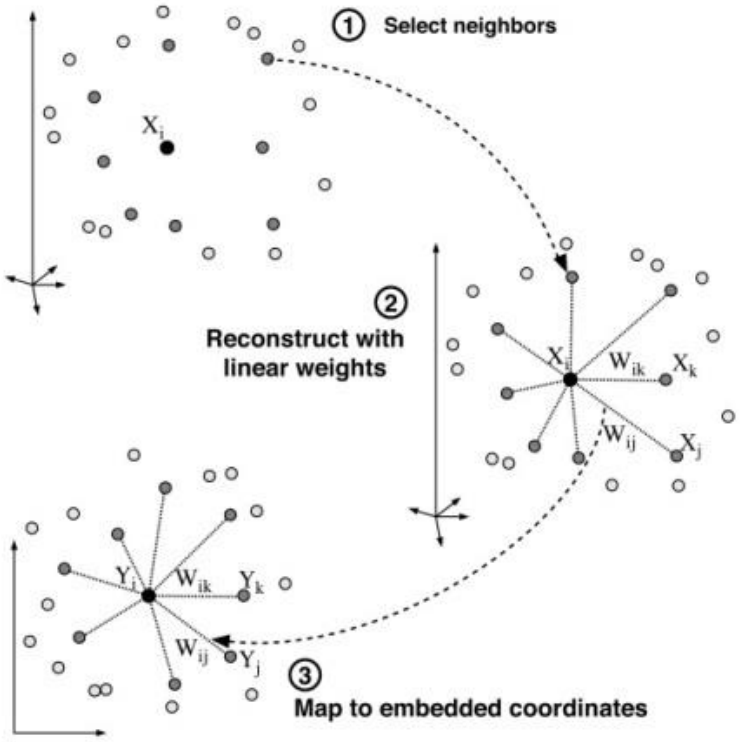

Figure 2.6. Steps of a locally linear embedding: (1) Assigning neighbors to each data point, (2) Compute the weights that best lineally reconstruct the data point from its neighbors, (3) Compute the low-dimensional embedding vectors. Reprinted from: "Nonlinear Dimensionality Reduction by Locally Linear Embedding," by S. T. Roweis and L. K. Saul, *Science, vol. 290*, no. 5500, pp. 2323-2326, 2000 [5]

2.3 Hyperspectral Image Classification

In remote sensing, a wide range algorithms are used to classify data, extract useful information, and assign labels to images related with certain land cover. The methods seek to leverage specific phenomenon assuming a unique structure that can be represented by a spatial pattern, or spectral signature, that is consequently assigned to a class [24]. In agriculture, the reflectance information of the reflectance from canopies and leaves is measured using passive sensors [22].

Machine learning refers to a broad range of tools to manage and analyze data, providing answers to questions from predictions or inferences. In image classification, unsupervised approaches identify homogenous groups within the data set, while supervised classification algorithms learn from labeled data (training data), iteratively, to map classes with complex attributes [34].

Detecting tassels in field conditions is difficult due to structure of the tassels, illumination, shadows, occlusion due to plant density, and differences in the background. Other applications for HSI classification are for crop and range management, weed science, plant pathology, and insect pest management.

Within the broad methods to classify data in machine learning, supervised and unsupervised methods, parametric or non-parametric, pixel or object-based, the Support Vector Machine (SVM) nonparametric classifier has been widely utilized for supervised, nonparametric pixel-based classification of hyperspectral data. SVM classifiers were adopted for this exploratory study.

2.3.1 Support Vector Machine – SVM

Proposed by Cortes and Vapnik (1995), the Support Vector Machine (SVM) was designed to identify a linear boundary for two classes, or the best separating hyperplane between classes, learning from training data and accurately predicting unknown data. SVM has become one of the most widely used non-parametric classifier for hyperspectral data.

To address the limitation of nonlinear separability, the use of the kernel trick has been implemented to transform the data onto a higher dimensional feature space [34], where the separation is linear. Common kernels used in remote sensing are polynomial kernels and the radial basis function (RBF) kernel.

SVM was used successfully by [35] to perform object identification of tassels in RGB images taken in-field with stationary cameras, reporting precision a rate of 98% and recall rate of 99%. The results are better when the tassels are fully developed, and the training samples are over 10.000 images.

More recently, [36] used an SVM classifier for image binarization to detect the cutting location for tassels in a natural canopy from a corn detasseling machine. The authors used color images and implemented morphological operations (shape and texture) to determine the potential tassel locations. They then implemented hierarchical clustering methods to merge multiple detections from the same tassel.

2.3.2 Spectral Angle Mapper – SAM

Another supervised classification algorithm that compares a known spectral signature, as a vector equal to the number of bands, with the vector of each pixel in the image calculating the spectral angle as in equation (2.3) [37]. The spectral angle between the reference pixel and the candidate pixel are calculated. Using the direction of the vector (but not the length), SAM is effective in reducing illumination problems.

$$\boldsymbol{\alpha} = \cos^{-1}\left(\frac{t \cdot r}{\|t\| \cdot \|r\|}\right) \tag{2.3}$$

where t is the spectrum from the image and r is the reference used for tassels.

[38] experimented with a hyperspectral image of 64 bands in the VNIR and NIR region of the spectrum, classifying 4 classes, clear water, dense cloud, dense vegetation and fallow land. SAM was determined to be insensitive to changes in illumination and unknown gain factors.

2.3.3 Other Methods

In [18], the authors analyzed the VNIR reflectance of corn canopies to estimate tasseling and pollen shed stages, collecting the hyperspectral data with a spectrophotometer. They conclude that the canopy reflectance changes during tasseling occurs in the VNIR spectrum and can be detectable by the algorithms partial least squares (PLS), artificial neural networks (ANN) and operator-enabled genetic algorithms (ROE-GA). The authors also noted that a simple correlation analysis produced poor results.

[20] presented an image-based phenotyping system to measure the tassel morphology (length, branch number, tassel area, tortuosity, compactness, fractal dimension, skeleton length, and perimeter length). The tassels were manually removed from the plants and images were collected in a photography box in controlled conditions.

Corn tassel detection on image processing was presented in [39] with the goal of quantifying morphological features of maize tassels, such as tassel length, branch number, tassel area, tortuosity, compactness, fractal dimension, skeleton length, and perimeter length. The Tassel Image-based Phenotyping System (TIPS) imaged tassels inside a white box and computed the traits using a series of algorithms using the tassel image binarization provided using the HSI color space. Image segmentation was applied to extract morphological traits from the tassels, such as

length, spike length, and branch number. The correlation between the traits measured by images and manual measurements ranged from 0.66 to 0.89.

While deep learning techniques are increasing in popularity and are now found more frequently in the remote sensing literature related to RGB imagery, classical machine learning approaches potentially provide important capability to derive insights related to tassel detection. To our knowledge, there have not been any studies to date that exploit the hyperspectral signature of the tassels imaged from wheeled vehicles or UAVs for in-field classification using machine learning algorithms.

3. EXPERIMENTAL METHODOLOGY AND DESIGN

This chapter describes the hyperspectral imagery selected for this study and the two platforms used to collect the data. Two experiments were included in the analysis, data from a controlled facility using plants grown in the greenhouse, and data from in-field conditions that were acquired with hyperspectral sensors mounted on a UAV and a converted sprayer platform with a custom boom, referred to as the PhenoRover. Additionally, the data pre-processing and the label selection using ENVI® 5.5 for ground truth is described. Python programing language using Spyder and Jupiter Notebook as integrated development environments (IDE), and a series of packages, including NumPy, Pandas, Matplotlib, Osgeo, Scipy, and Sklearn, were used to process the data, train, test, and evaluate the classification experiments.

3.1 Hyperspectral Data Acquisition and Pre-Processing

3.1.1 Controlled Facility Data

Plants were grown in the Controlled Environment Phenotyping Facility (CEPF) at Purdue University in Indiana, USA, and imaged by a hyperspectral sensor. The CEPF (Figure 3.1) includes a plant growth facility with precise environmental controls and automated conveyors for imaging plants up to 5m height, using top view and side view RGB and VNIR hyperspectral cameras to provide high-throughput plant measurements in complex plant traits assessment [40].

a) b)

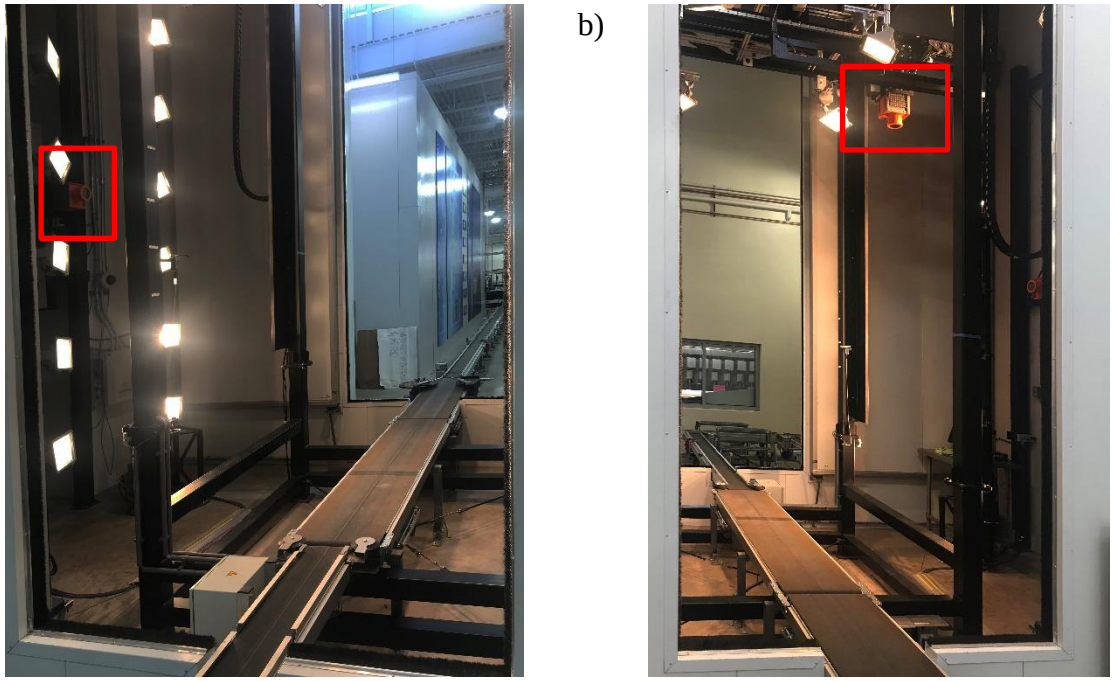

Figure 3.1. CEPF automatic conveyor for plants imaging. a) Side view camera, b) Top view camera.

The hyperspectral camera MSV-500 VNIR scanner (Middleton Spectral Vision, Middleton, WI), provides top-view and side-view imagery from the plants with a range of 400 – 1000 nm, with spectral resolution of 1.3 nm (473 bands), and spatial resolution of 2.08 mm/pixel. The data acquired for this study were binned to 4.8 nm spectral resolution (119 bands) and processed to reflectance (with support of Jin lab in the Agricultural and Biological Engineering Department at Purdue University).

The test involved plants of the Hybrid Mo17xB73 planted on July 15, 2019 and transplanted into a chamber 15 days later, with a soil mixture 50:50 mixture of profile greens grade and Sungrow 360 and Nitrogen rate of 100 ppm. The plants were watered to saturation each day and observed during the flowering time for this study. The hyperspectral images were calibrated using a dark and white reference to calculate the spectral reflectance using the equation (3.1) [41].

$$\boldsymbol{X_\lambda} = \frac{I_\lambda - B}{W_\lambda - B} \tag{3.1}$$

where $\boldsymbol{X_\lambda}$ is the pixel calibrated reflectance for the image, I_λ is the original measured radiance value for the pixel, W_λ is the spectral radiance of the white reference, and B is the black reference.

3.1.2 Field Data

The field data for this study were collected during the 2019 growing season at the Indiana Corn and Soybean Innovation Center (ICSIC) at the Agronomy Center for Research and Education (ACRE), Purdue University, Indiana, USA. The field experiment Genome to Fields (G2F) consisted of 255 experiment panels of maize planted north to south on June 5th, with two replicas for a total of 510 plots (15 ranges by 68 rows), and a population density of 30 K/acre. Students collected ground phenotypic data as reference data (ground truth), for flowering and tasseling time the data was collected every other day (Figure 3.2).

Figure 3.2. Field data collection crew.

The UAV-HSI data were collected with a multi-rotor Matrice 600 Pro UAV from DJI (Figure 3.3. a), using a Nano-Hyperspec® VNIR sensor (Headwall Photonics Inc., Bolton, MA) with a spectral range of 400-1000 nm divided into 270 spectral bands and 640 spatial bands (2.2 nm/pixel), at 1 cm spatial resolution. The data were aggregated to 135 bands of the original 270 via averaging, and noisy bands after 920nm were removed. The system includes an Applanix APX-15v3 GPS/IMU to accurately georeference the data, a LiDAR Velodyne VLP-16 Lite sensor, and an RGB Sony Alpha 7RIII camera. The UAV-HSI data were collected as close to solar noon as possible to reduce the impact of shadows, on August 08 with flight lines east to west, and August 10 with flight lines north to south.

The wheel-based HSI data were collected using a commercial high-clearance spraying platform (LeeAgra Avenger) (Figure 3.3. b). The PhenoRover has a custom boom mounted on the chassis' hydraulically operated front lift, allowing the sensors to reach a height above ground of up to 4.2 meters. The HSI data were collected with a Headwall Machine Vision camera with a wavelength range of 400-1000 nm, spectral resolution of 2.2 nm (272 bands) and spatial resolution of 5 mm (from 2 m above the canopy). Additionally, the PhenoRover includes two Velodyne HDL-32E LiDAR units, and two FLIR GrassHopper3 GigE 9.1MP RGB color cameras. For direct georeferencing, the PhenoRover uses an Applanix POS LV 125 with dual GNSS antennas. The HSI data were down sampled to 136 bands via averaging. The PR-HSI data were collected on August 09th, during flowering time, and driving from north to south to mitigate the impact of shadows from the arms on the boom.

a)  b) 

Figure 3.3. Hyperspectral imagery field data carriers. a) UAV Matrice 600, b) PhenoRover.

Calibrated targets with diffuse near Lambertian reflectance coating (5%, 50%, and 80%), Labsphere – Permaflect® were deployed in the field for each UAV data acquisition (Figure 3.4). The reflectance of each target was measured with a field-portable spectroradiometer SVC XHR-1024i, Spectra Vista Corporation, using a foreoptic of 4° nominal field of view (FOV) lens, with a wavelength range of 250-2500 nm, and1024 bands. The data were processed considering a white diffuse reflectance target Spectralon 98% model CSTM-SRT-99-100, the sun angle, and the location of the field. All datasets were initially calibrated to radiance using SpectralView® (Headwall Photonics Inc., Bolton, MA) and to reflectance using PREDiCT, software developed by the Laboratory for Applications of Remote Sensing (LARS) at Purdue University. PREDiCT uses the empirical line method (ELM) to calibrate the images using the reflectance of the field calibrated targets which were imaged during the flight.

Figure 3.4. Calibrated targets for UAV data collection.

The PhenoRover had a hydraulic arm (Figure 3.5), with two Labsphere – Permaflect® targets (10% and 80%) and two Type 822 Polyester target fabric targets (30% and 56%) (Group 8 Technology, Inc.). The arm was rotated into the FOV of the sensor after each driving line, at the southernmost part of the field. Each driving line required ~5 minutes to be collected, and the PhenoRover drove back to the north side of the field to collect the next line (as noted earlier, to mitigate the impact of shadows from the PhenoRover arms due to sun angle). To process these datasets into reflectance, the spectral signatures of the calibrated targets in the arm and the Spectralon were used as reference for the PhenoRover and UAV data, respectively, following the ELM method in PREDiCT.

Figure 3.5. Calibrated targets on PhenoRover and Spectralon 98% used as white reference.

For both platforms (UAV and PhenoRover), data orthorectification was performed with PREDiCT using a Savitzky–Golay (SG) adaptive smoothing filter, on an 8 cm digital surface model (DSM). The DSM was developed from the LiDAR data collected in the fields from the UAV following the methodology used by the Digital Photogrammetry Research Group (DPRG) at Purdue University.

3.2 Data Processing and Classification

3.2.1 Data Selection

Two dates were selected for the CEPF-HSI data, September 04th at the beginning of the flowering time and September 10th during the peak of the flowering time. For the second date, both side view and top view data were analyzed. Table 3.1 summarizes the data analyzed for the experiments.

For the in-field experiment, two plots were selected for each collection system, based on the shortest time between flowering and the data acquisition date, and that the same plot was covered on the two dates. Plot 4553 from the pedigree PHW65_MoG_0148, source LH195:0807 and -plot 4529 PHW65_MoG_0108, source LH195:1639 both reported flowering time on August 07th. Examples of the images from the CEPF are shown in Figure 3.6, PhenoRover images in Figure 3.7, and UAV images for both dates August 08th and 10th in Figure 3.8.

Table 3.1. HSI data summary and geometry

System	Date	Data collection orientation
CEPF	20190904	Side view
CEPF	20190910	Side view and Top view
PR-HSI	20190809	North – South
UAV-HSI	20190808	East – West flight lines
UAV-HSI	20190810	North – South flight lines

a) b) c)

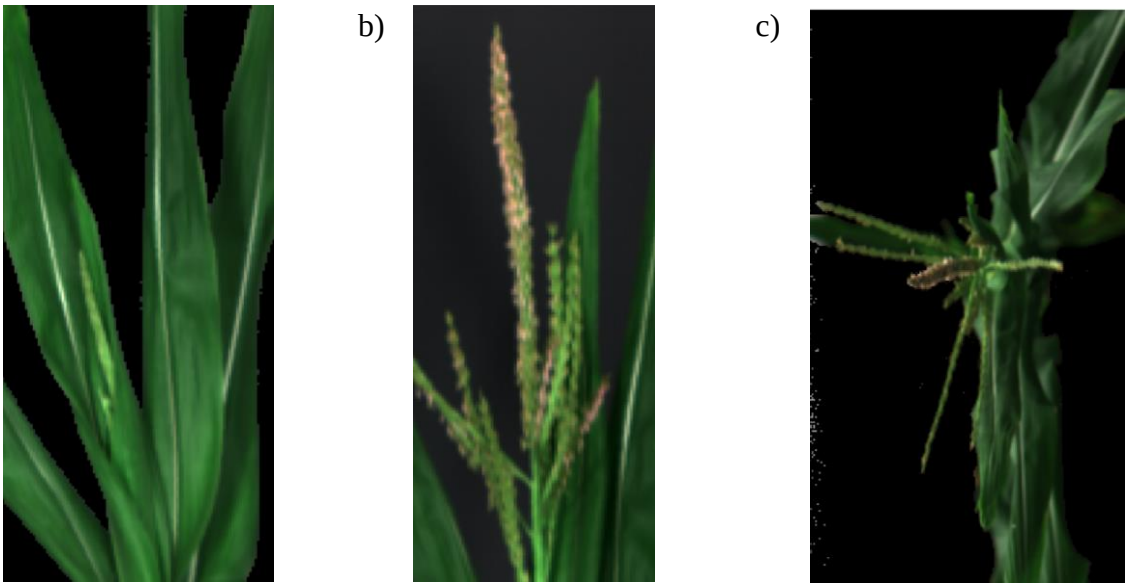

Figure 3.6. Hyperspectral imagery subsets for CEPF tests. a) Test 1: CEPF 20190904 – side view, b) Test 2: CEPF 20190910 – side view c) Test 3: CEPF 20190910 – top view.

a) b)

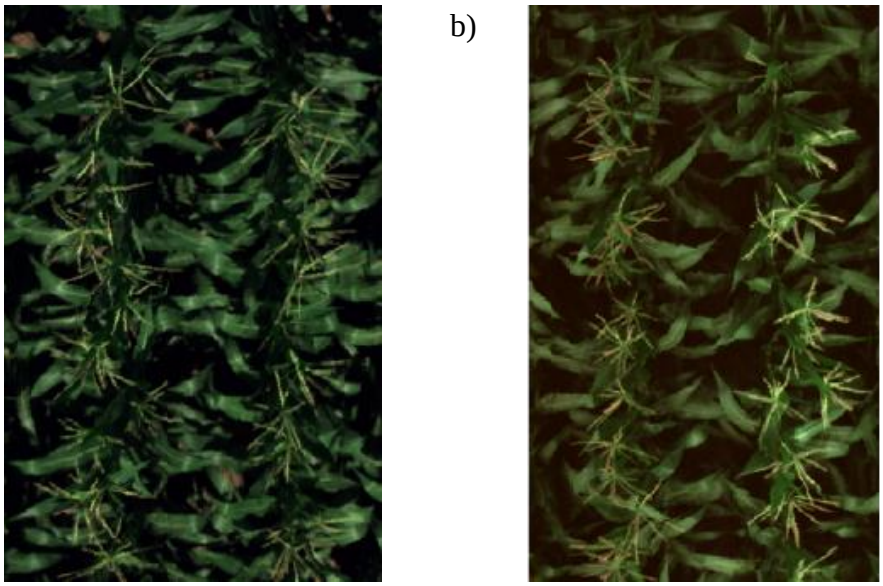

Figure 3.7. Hyperspectral imagery subsets for PhenoRover tests. a) Test 1: PR 20190809 – Plot 1, b) Test 2: PR 20190809 – Plot 2.

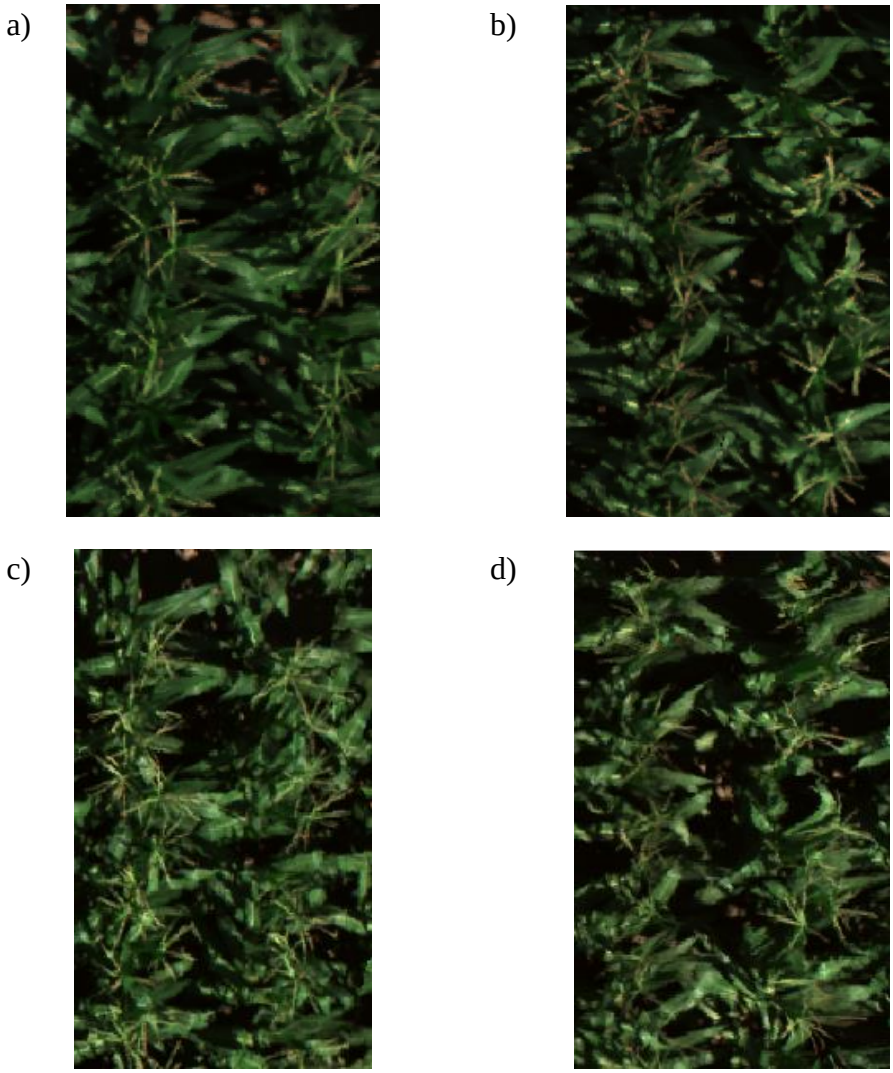

Figure 3.8. Hyperspectral imagery subsets for PhenoRover tests. a) Test 1: UAV 20190808 – Plot 1, b) Test 2: UAV 20190808 – Plot 2, c) Test 1: UAV 20190810 – Plot 1, d) Test 2: UAV 20190810 – Plot 2.

3.2.2 Data Labeling

Supervised machine learning algorithms for classification purposes require labeled data as training for all the categories. Generating these labels or ground truth data is a time-consuming task and requires previous knowledge of the images. The label selection/annotation for this study was performed in ENVI® 5.5, Harris Geospatial Solutions Inc., using three images as reference: a) the true color (RGB) bands of the hyperspectral image, b) band 2 from principal components analysis (PCA), and c) band 3 of the minimum noise-fraction (MNF) transform [42] (Figure 3.9). Every pixel was visually classified into two classes, one for tassels and one for the rest of the image, generating a two class labels (Figure 3.9. d). PCA and MNF were used to generate an initial region

of interest (ROI) for tassels, using a threshold of 0.7 to select the dark areas for both images. Subsequently, the spectral signature of pixels from the original HSI image was used to correct for tassel pixels that were missing from the initial ROI, and missing pixels for tassels that were not included. In this process, shiny leaves, the midribs, and shiny soil were usually corrected from the initial ROI, as the spectral signature tends to be similar to the tassels.

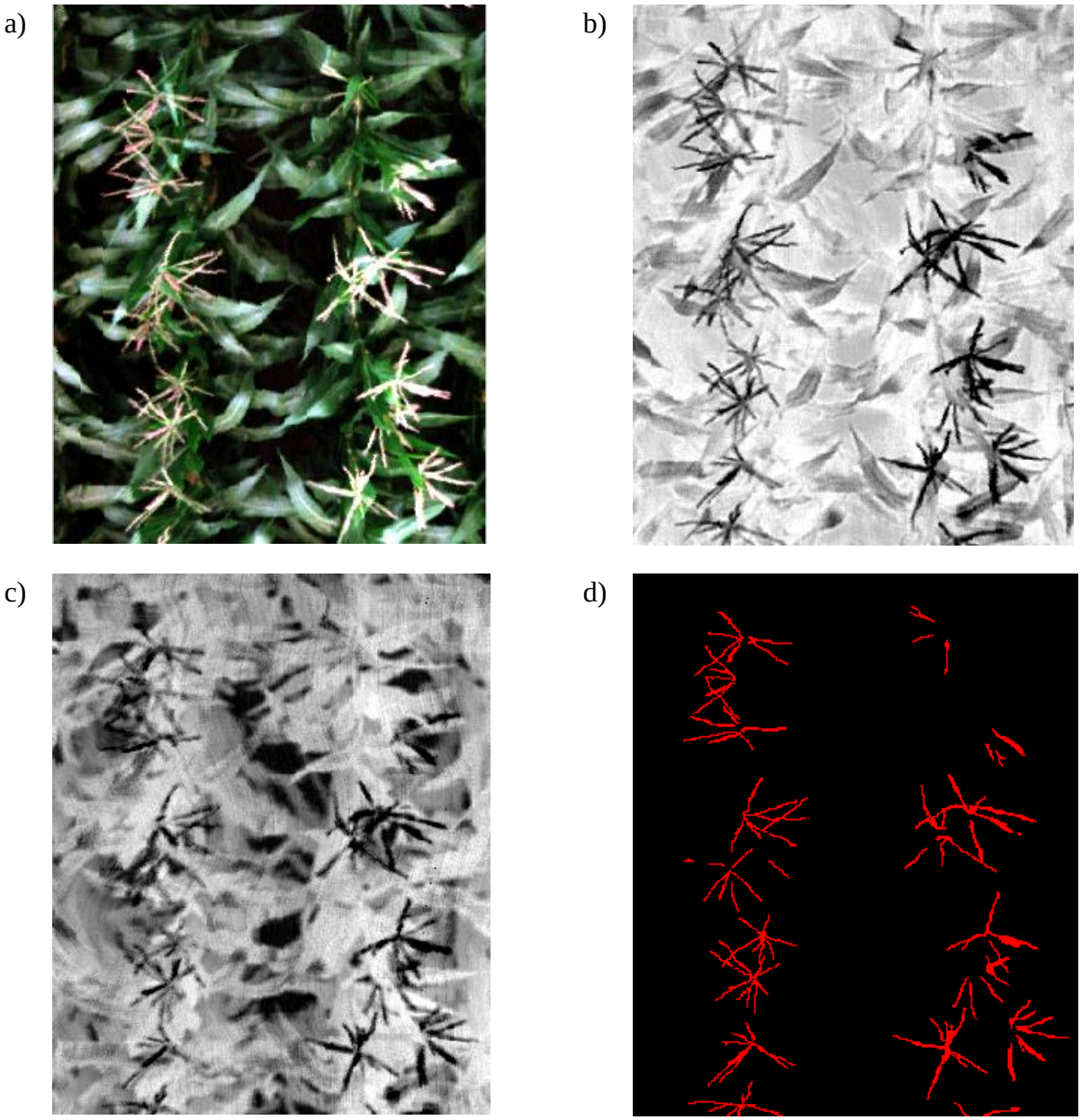

Figure 3.9. Hyperspectral imagery for PhenoRover test 1. a) Hyperspectral data – bands R642 G548 B468, b) Principal Components Analysis – band 2. c) Minimum Noise-Fraction – band 3. d) Training label – Red: Tassels, Black: Background

3.2.3 Data Visualization

The complete classification approach consisted of a workflow using ENVI® 5.5 and a series of functions written in Python, using several existing libraries. Each of the HSI images was loaded into Python as a NumPy array with a shape corresponding to the $Width \times Height \times Number\ of\ Bands$. The $Height$ and $Width$ represent the X, Y positioning of the pixels. The initial exploration of the HSI imagery includes plots of spectral signatures for all pixels to visualize the initial data. The series of plots allow identification of the spectral signature of each part of the maize plant when flowering. Figure 3.10 shows the spectral signature of data collected from the CEPF for the leaf area, leaf midrib, and tassel in early-stage form. The midrib and the tassel have very similar signatures, especially from 500 nm to the 650 nm in the green bands. The tassel and the leaves have a similar shape and slope of the spectral signatures from the 400 nm to the 650 nm, although with different magnitudes.

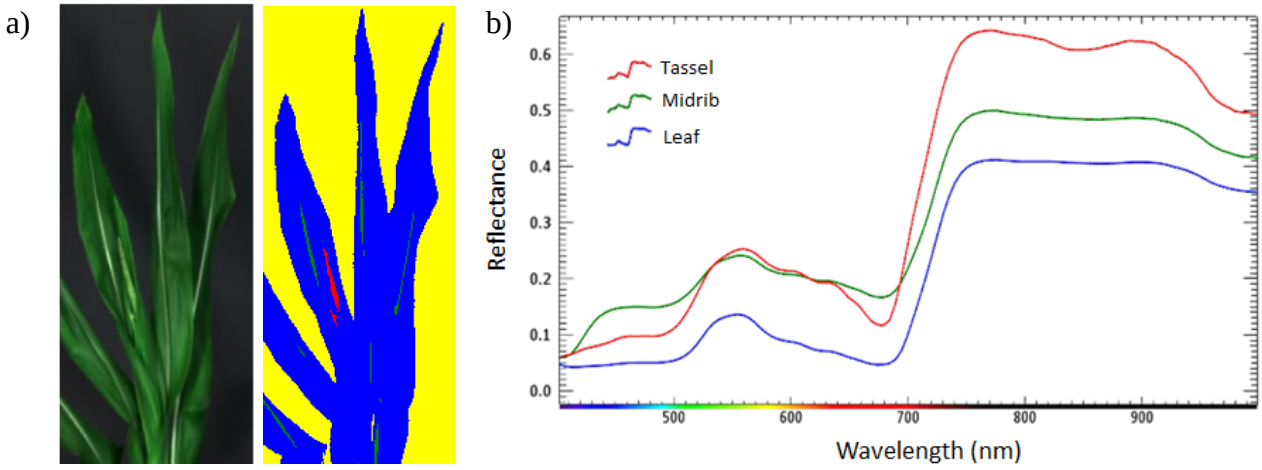

Figure 3.10. CEPF data for maize plant flowering. a) HSI side view collected on September 4 and the respective pixel labeling, b) Spectral signature plot for tassel in early stage, midrib and leaf area.

Univariate and multivariate plots were used explore the data. The correlation matrix of the spectral signatures (Figure 3.11) illustrates the high positive correlation between bands within the RGB region and the correlation between bands within the NIR region, resulting in two major groupings. Correlation between RGB bands and NIR bands is shown to be low, implying the different response to material of the visible range RGB and the NIR bands, explained in chapter 2.2.

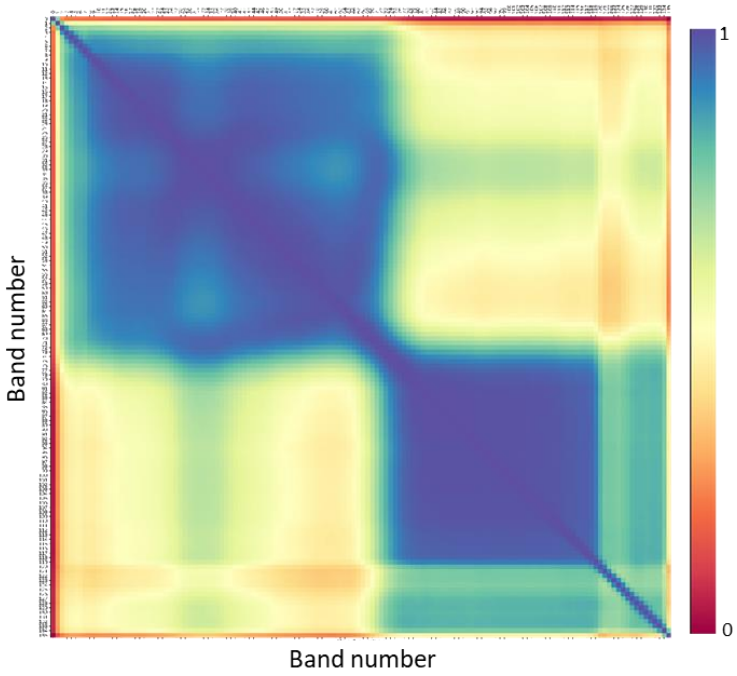

Figure 3.11. Correlation matrix for PhenoRover data Test 1 (136 bands)

3.2.4 Model Training and Validation

As explained in chapter 2.2 and shown in Figure 3.11, many bands in HSI are highly correlated. To achieve the best performance with the SAM and SVM classifiers, PCA, LLE, and Isomap were used to extract latent features using scikit-learn library in Python (Figure 3.12). For the linear dimension reduction PCA (Figure 3.12. a), the number of components was found identifying the elbow of a scree plot (component 3), where the proportion of variance explained was about 95%. LLE and Isomap (Figure 3.12. b and c), explained in chapter 2.2.1.3 and 2.2.1.4 respectively, were computed as nonlinear dimension reduction approaches. While LLE seeks to preserve the topology of the manifold using local spectral information in a lower dimension, Isomap attempts to preserve the geodesic distances in a lower dimension using a global similarity matrix from local information. For both methods, the number of neighbors used to compute the nonlinear reduction is important. In this study, the cost function plot was used to determine the optimal value for number of components and neighbors, selecting the neighbors where the reconstruction error was lower. The calculation of LLE and Isomap requires intensive computation. Thus, for this study the images of the full plots were subset to smaller cubes of 240x270 pixels for PhenoRover and 136x300 pixels for UAV.

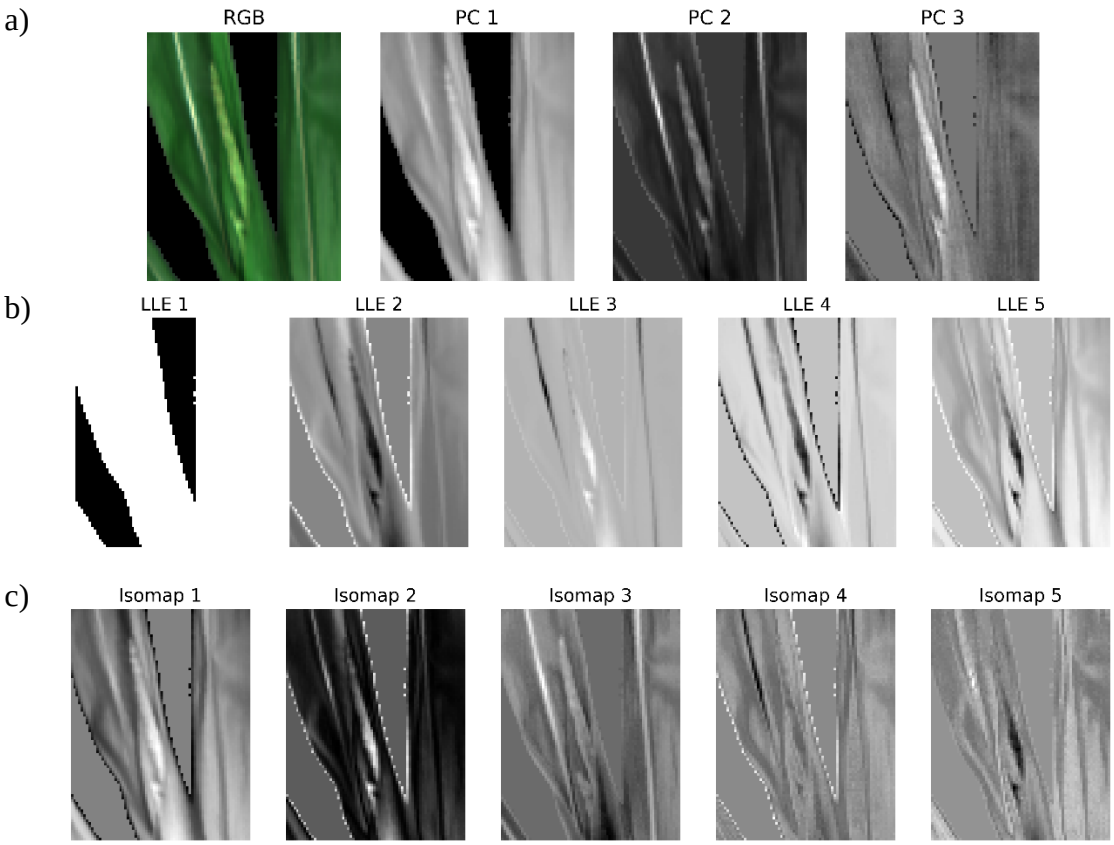

Figure 3.12. Feature Reduction for HSI – CEPF Test 1 data for maize plant at early flowering stage. a) RGB color composite from HSI and PCA with 3 components, b) LLE 5 components, c) Isomap 5 components.

After standardizing, the datasets were split into training, validation, and testing, where the training set was used to train the algorithm and validation was used to perform the model selection. For this experiment, each hyperspectral image was split into training (60 %), validation (10%), and 30% not included in the training process for testing

SVM classification is sensitive to unbalanced data and can produce inaccurate results by misclassifying the underrepresented dataset when the density of the majority class is higher than the minority class, resulting in a largely skewed hyperplane [43]. A learning solution proposed by [44] is to assign different misclassification costs (C), one for each class (C^+, C^-). The different error cost (DEC) method reduces the effect of the misclassification for the imbalanced data, setting the cost equal to the minority-to-majority-class ratio. The parameter estimation for the RBF kernel (C, γ) was obtained using a grid search with a 5-fold-cross-validation. The SVM-RBF model in this study used the parameters showing the best overall accuracy from the grid-search. Statistical analysis of the results was conducted based on the confusion matrix and the F1-score.

4. EXPERIMENTAL RESULTS

This chapter presents the results of two groups of experiments with nine different datasets collected from three systems (CEPF, UAV, and PhenoRover) during the flowering time in maize. The first experiment consisted of classifying the tassels using the SAM algorithm and the second experiment was completed using SVM as a classifier. For both experiments, the datasets were tested on the original HSI and after performing dimension reduction with PCA, LLE, and Isomap. Additionally, this chapter includes important findings related to the pre-processing steps.

4.1 Data Pre-Processing Results

4.1.1 Spectral Signature

The high spectral and spatial resolution of the CEPF data was useful to sample the spectral signature from different parts of the maize plants in an early stage and late season after flowering time. Figure 4.1 compares the spectral signatures of a tassel in the early stage versus the tassel after the anthers have extruded in the same plant, showing a significant reduction in the reflectance in the green bands and NIR bands. Additionally, a distinctive shape in the signature between the 670 and 700 nm was identified, where the slope of the curve for the tassel is greater than for the rest of the signatures. Furthermore, the plot shows the similarity of the spectral signature of the tassels with the midribs.

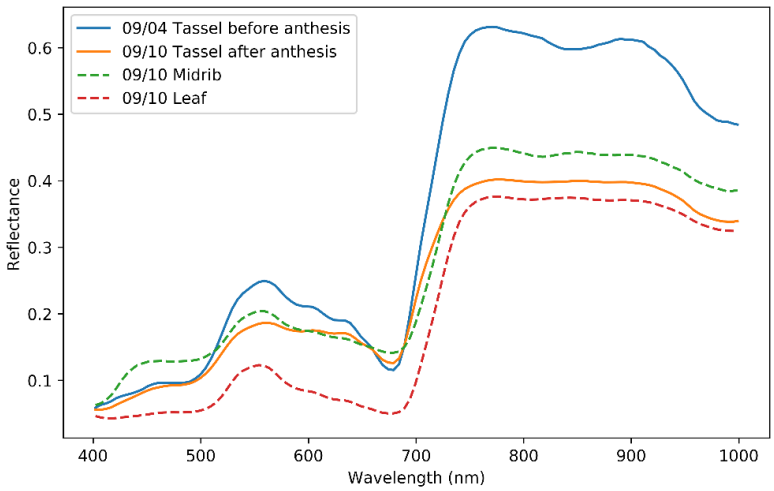

Figure 4.1. Spectral signature for a tassel in early stage before anthesis collected on 09/04 (Test 1) and after anthesis, data collected on 09/10 (Test 2). Data collected from the CEPF.

4.2 Classification Model Results

The classification performance of the models in this study was assessed using traditional metrics, including precision, recall, and the F1-score. These metrics are reported to address the problem of the imbalance data where one class can add more weight over the other class in a binary classification, resulting in high accuracy for the model with poor performance for the class with less weight [45]. Initially, a confusion matrix for the binary classification was computed to have better insight relative to the performance for each class, where the diagonal of the matrix represents the correctly classified pixels, True Positive (TP) and True Negative (TN), and the remaining cells representing the incorrectly classified pixels, False Positive (FP) or Type I error and False Negative (FN) or Type II error. Furthermore, the classification metrics were computed as follows:

$$\boldsymbol{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4.1}$$

$$\boldsymbol{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4.2}$$

$$\boldsymbol{F1} = 2 * \frac{Precision * Recall}{Presicion + Recall} \tag{4.3}$$

Precision is the positive predicted value and quantifies the correctly predicted pixels as tassels. Recall (or sensitivity) is the true positive rate that relates to the number of pixels belonging to tassels that were classified positive and those that the model incorrectly does not capture as tassels. The F1-score can be used to evaluate the entire model, balancing the precision and recall, making it a sensitive metric to changes in the data distribution and ratios.

4.2.1 Spectral Angle Mapper Classification

ENVI® 5.5 was used for mapping the spectral similarity of the image to the spectral reference data. As discussed in 2.3.2, SAM computes the angle between the spectral reference and the spectrum in the image and is not affected by the length of the vectors. The spectra collected from the images from the CEPF were used as reference to investigate the reflectance of the tassels and to later experiment with field data collected by UAV and PhenoRover. The SAM classification

was tested on the original hyperspectral image (HSI), hyperspectral image with PCA dimension reduction (HSI+PCA), hyperspectral image with LLE dimension reduction (HSI+LLE), and hyperspectral image with Isomap dimension reduction (HSI+ISO).

Table 4.1 shows the tassel classification performance using SAM for the CEPF data. Because tests 1 and 2, show the side view of the tassel, the F1-scores are significantly higher than in test 3. For test 1, the side view HSI shows the tassel forming, and the spectral signature has a significantly higher reflectance in the NIR region than the tassel in test 2, as previously shown in Figure 4.1.

Table 4.1. SAM classification results for tassel with CEPF data.

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	16	1864	1	9	0.94	0.64	0.76
HSI+PCA	17	1860	5	8	0.77	0.68	0.72
HSI+LLE	24	1858	7	1	0.77	**0.96**	**0.86**
HSI+ISO	25	1799	66	0	0.27	1.00	0.43
Test 2							
HSI	620	2641	56	967	0.92	0.39	0.55
HSI+PCA	562	3612	15	767	0.97	0.42	0.59
HSI+LLE	1016	3587	40	313	0.96	**0.77**	**0.85**
HSI+ISO	475	3611	16	854	0.97	0.36	0.52
Test 3							
HSI	233	5055	79	433	0.75	0.35	0.48
HSI+PCA	246	5013	121	420	0.67	**0.37**	**0.48**
HSI+LLE	235	4692	442	431	0.35	0.35	0.35
HSI+ISO	171	4963	171	495	0.50	0.26	0.34
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

It can be clearly seen that the tassel classification accuracy declines when the anthers are extruding, as observed in the recall score dropping from 0.96 to 0.77. When exploring the LLE components for each dataset, the spectral separability of the pixels for Test 1 (Figure 4.2. a) is greater than in Test 2 (Figure 4.2. b), where the spectral separability of the pixels is more challenging in the LLE transformed coordinate system. The biological and chemical composition in the anthers extruding could possibly lower the reflectance in the NIR bands at the point where is very similar to the midrib and the leaves.

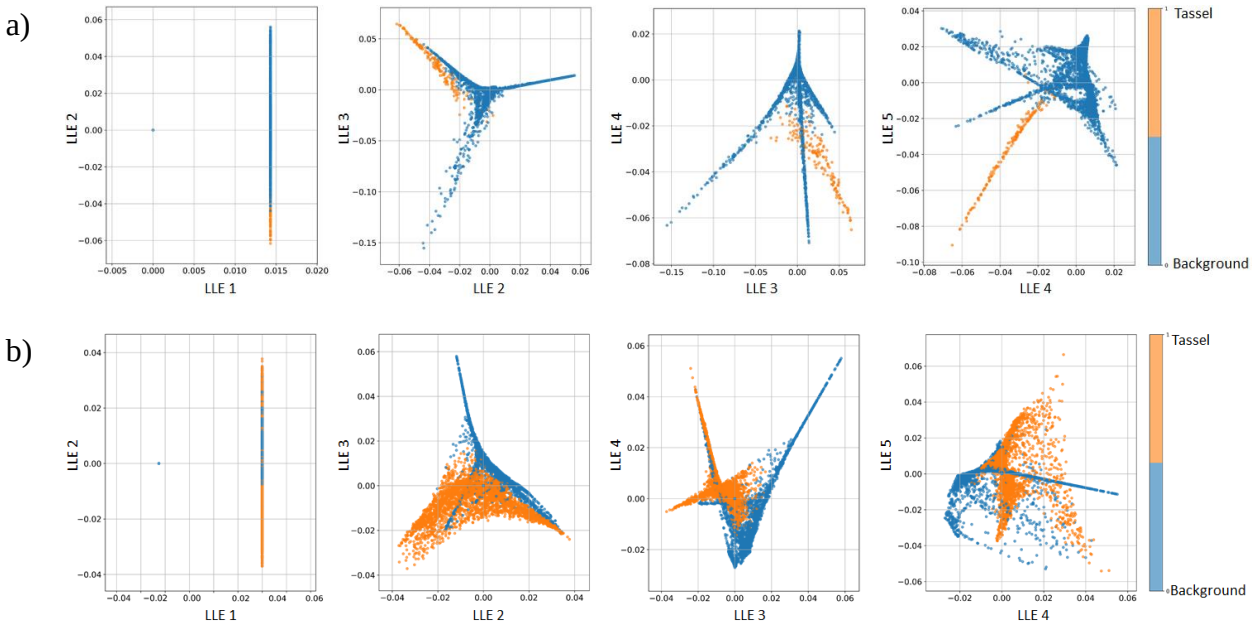

Figure 4.2. LLE components for CEPF data showing the separability of the tassels (orange) and background (blue). a) LLE components for Test 1, b) LLE components for Test 2

Figure 4.3 shows the classification labels for the Test 1 and the classified pixels using LLE as a dimension reduction method. The early-stage tassel has a very well-defined spectral signature, especially in the NIR region that allows a reasonable pixel classification.

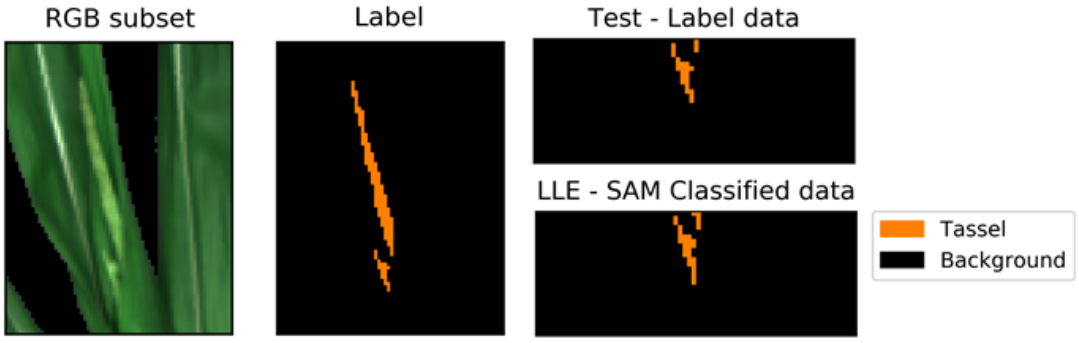

Figure 4.3. SAM classification results for CEPF HSI test 1 data using LLE as dimension reduction method. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

In figure 4.4, the results for Test 2 shows the complexity of the tassel diminishing the number of pixels identifies as TP and misclassifying the midrib as tassel.

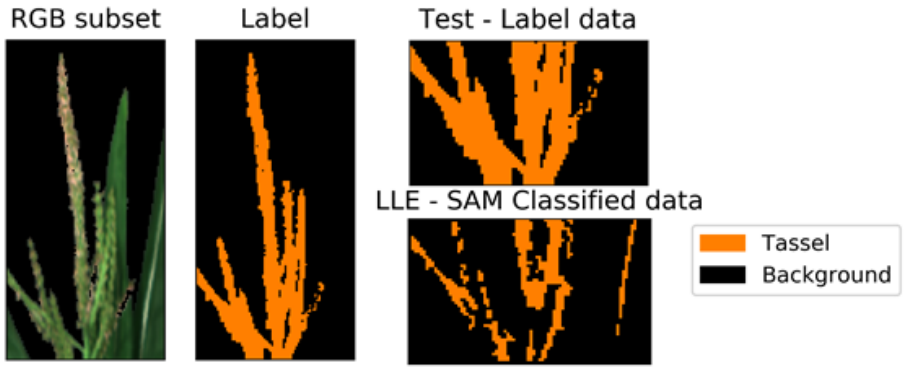

Figure 4.4. SAM classification results for CEPF HSI Test 2 data using LLE as dimension reduction method. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

For test 3 in the CEPF (Figure 4.5), the data were collected simultaneously with Test 2 but the top view of the tassel is shown when anthesis is happening. This image was processed to simulate the view of the sensors in the field, but in the controlled facility. The F1-score and the recall values are even lower than the results in Test 2 (side view) since the topmost of the tassels belong to the anthers extruding and the tassel without anthers is not seen by the sensor. For this test, the dimension reduction that best classified the pixels as tassels was PCA with an F1-score of 0.48 and recall of 0.37, followed closely by the results from this data. From this test, it can be inferred that the linear dimension reduction performed better than the nonlinear dimension reduction methods.

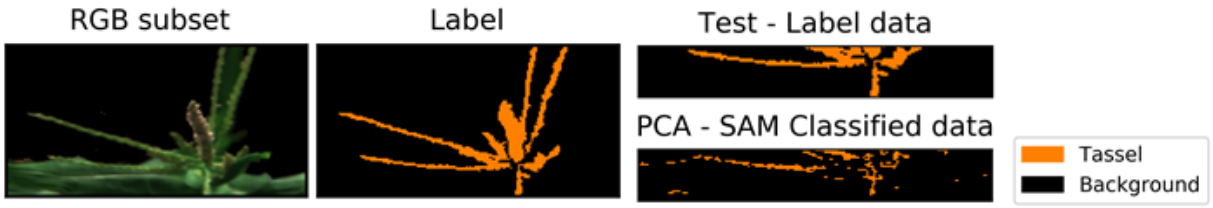

Figure 4.5. SAM classification results for CEPF HSI test 3 data using PCA as dimension reduction method. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

The experimental results from the PhenoRover data using SAM as classifier are summarized in Table 4.2, where the nonlinear dimension reduction LLE for Test 1 and Isomap for Test 2 resulted in higher F-1 scores, with 0.42 and 0.59 respectively. The lower score in these results, compared with the experiments in the CEPF, can be explained by the complexity of the image in a natural environment, where the light is not constant, and the plant density of the plots exacerbates the challenge of classifying tassels.

Table 4.2. SAM results for tassel classification with PhenoRover data

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	285	16287	506	402	0.36	0.41	0.39
HSI+PCA	253	16469	324	434	0.44	0.37	0.40
HSI+LLE	244	16554	239	443	0.51	**0.36**	**0.42**
HSI+ISO	184	16652	141	503	0.57	0.27	0.36
Test 2							
HSI	494	17631	623	452	0.44	0.52	0.48
HSI+PCA	723	17319	935	223	0.44	0.76	0.56
HSI+LLE	388	18028	226	558	0.63	0.41	0.50
HSI+ISO	601	17758	496	345	0.55	**0.64**	**0.59**
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Figure 4.6. shows the in-field image complexity in the RGB color composite of the hyperspectral subset. Additionally, the figure compares the reference labels with the final classification for Test 2 computing SAM after the Isomap transformation, where the number of true positives is 64% of the total amount of pixels for the tassels, and the false negative (FN) is about 37%. This can be explained by the similarity of the tassel's spectral signature with the midrib, given that SAM calculates the angle between the reference spectra and the image spectra.

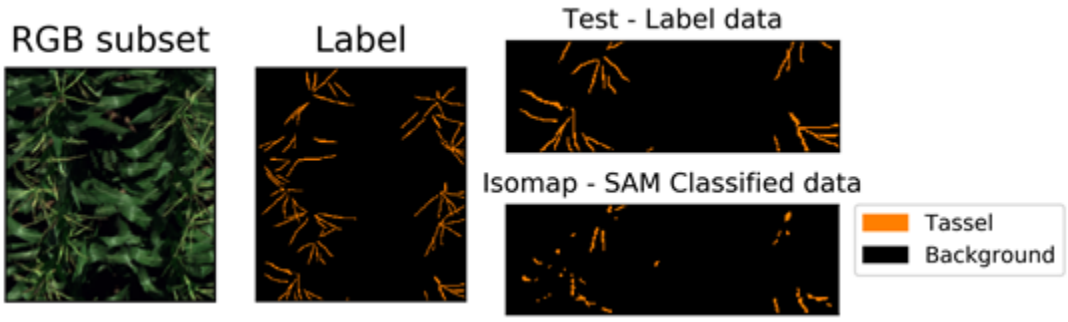

Figure 4.6. SAM classification results for PhenoRover HSI test 2 data using Isomap for dimension reduction. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

The UAV data had similar results for both days, August 08th (Table 4.3) and August 10th (Table 4.4). The experiment using Isomap resulted in higher F1-scores for all tests with 0.60 and 0.44 for the data on August 8th, and 0.54. and 0.47 for August 10th. It is important to note from these tables that the models using PCA have higher recall scores and produce fewer FN than the

other models for the specific tassel class, showing better performance at classifying the tassels, but lacks on precision in having more FP than the rest of the models.

Table 4.3. SAM results for tassel classification with UAV data 20190808

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	353	17262	247	345	0.59	0.51	0.54
HSI+PCA	506	16745	764	192	0.40	0.72	0.51
HSI+LLE	169	17362	147	529	0.53	0.24	0.33
HSI+ISO	477	17103	406	221	0.54	**0.68**	**0.60**
Test 2							
HSI	169	16247	147	342	0.53	0.33	0.41
HSI+PCA	361	15265	129	150	0.24	0.71	0.36
HSI+LLE	284	15655	739	227	0.28	0.56	0.37
HSI+ISO	303	15822	572	208	0.35	**0.59**	**0.44**
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Table 4.4. SAM results for tassel classification with UAV data 20190810

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	379	16767	352	590	0.52	0.39	0.45
HSI+PCA	392	16812	307	577	0.56	0.40	0.47
HSI+LLE	429	16883	236	540	0.65	0.44	0.53
HSI+ISO	553	16595	524	416	0.51	**0.57**	**0.54**
Test 2							
HSI	178	19575	386	469	0.32	0.28	0.29
HSI+PCA	350	19257	704	297	0.33	0.54	0.41
HSI+LLE	327	19254	707	320	0.32	0.51	0.39
HSI+ISO	281	19690	271	366	0.51	**0.43**	**0.47**
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Figures 4.7 and 4.8 illustrate the results for Test 1 on both dates using hyperspectral data with Isomap, where the contribution of FP is visible for both results. The high density of the plants, the similarity of the tassels' spectral signature with the midrib, and the lower resolution of the UAV data (1 cm) resulted in low F1-scores for SAM classifier.

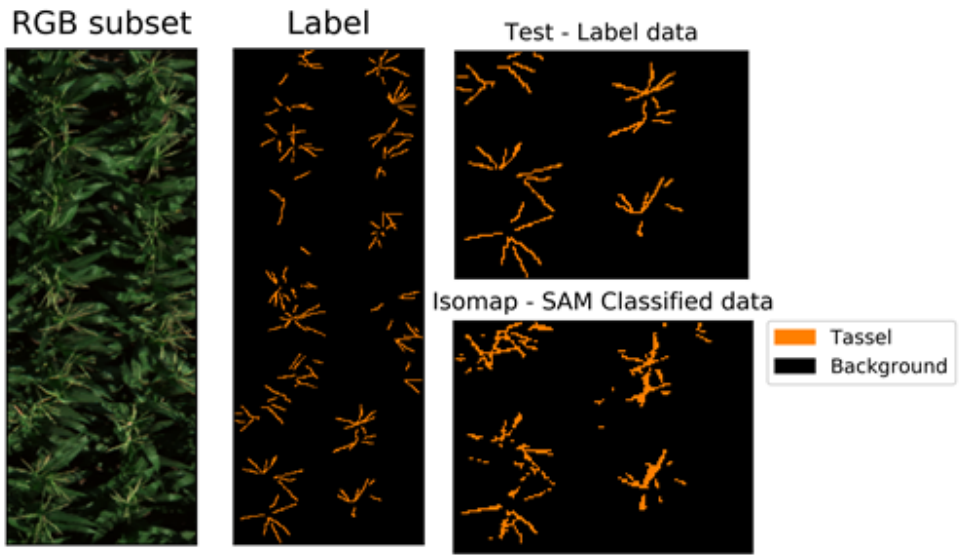

Figure 4.7. SAM classification results for UAV HSI test 1 on August 08th using Isomap for dimension reduction. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

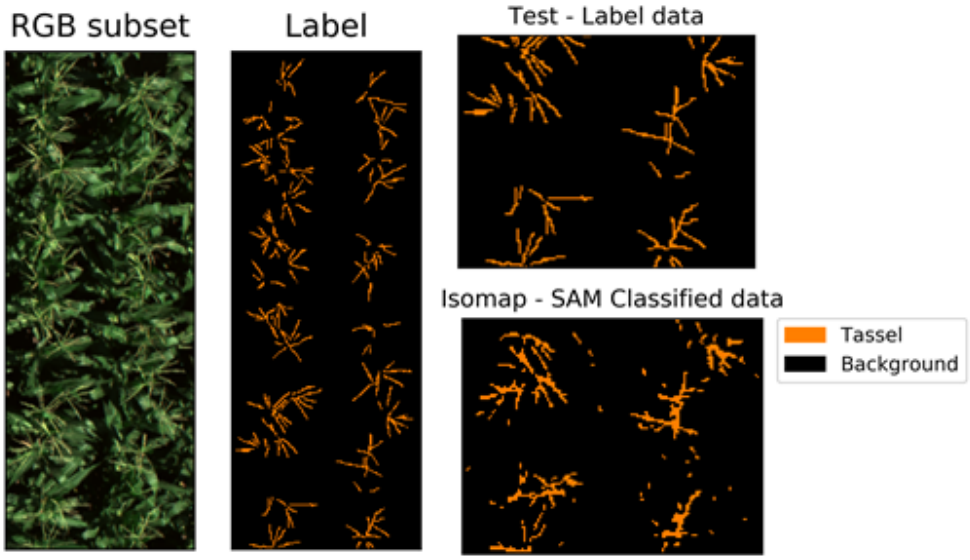

Figure 4.8. SAM classification results for UAV HSI test 1 on August 10th using Isomap as dimension reduction method. From left to right: RGB color composite of HSI, ground truth for tassel label, subset for testing, and final classified data.

4.2.2 Support Vector Machine Classification

Python libraries were used to preprocess, train, validate, and test the HSI data using SVM-RBF algorithm as a classifier. As mentioned in 3.2.4, the SVM classification algorithm is sensitive to imbalanced data, potentially providing inaccurate results. For this research, the DEC method

was used to reduce the impact of the underrepresentation of the tassels' class (5% against 95% of background pixels) in the final classification score. The SVM classification was tested on the original hyperspectral image (HSI), a hyperspectral image with PCA dimension reduction (HSI+PCA), a hyperspectral image with LLE dimension reduction (HSI+LLE), and a hyperspectral image with Isomap dimension reduction (HSI+ISO).

The results for the tests on the CEPF data (Table 4.5) show very high F1-scores and recall for all tests. In Test 1, the best performance overall was for the dataset after Isomap dimension reduction with an F1-score of 0.96 and a precision of 1.00, meaning that there were no FP in the classification process. These results are related to the structure of the tassel at the early-stage and the side view of the camera capturing the emergence of the tassel without obstruction. Test 2 and 3 yielded better performance using the original HSI, 0.97 and 0.90 respectively, and in contrast with Test 2, Isomap resulted in the lowest F1-score. This could be explained by the complexity of the images that cannot be explained by the Isomap-based features since the global manifold methods cannot exploit relationships among the local neighbors of the spectral graph. The SVM model, which is robust to the number of inputs, seems to learn the complexity of the tassels better when all the bands are included.

Table 4.5. SVM results for tassel classification with CEPF data.

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	50	2672	6	2	0.89	0.96	0.93
HSI+PCA	44	2677	1	8	0.98	0.85	0.91
HSI+LLE	23	1863	2	2	0.92	0.92	0.92
HSI+ISO	23	1865	0	2	1.00	**0.92**	**0.96**
Test 2							
HSI	845	3382	49	8	0.95	**0.99**	**0.97**
HSI+PCA	843	3381	50	10	0.94	0.99	0.97
HSI+LLE	1207	3573	54	122	0.96	0.91	0.86
HSI+ISO	1138	3592	35	191	0.97	0.85	0.86
Test 3							
HSI	558	5314	107	21	0.84	**0.96**	**0.90**
HSI+PCA	526	5258	163	53	0.76	0.91	0.83
HSI+LLE	610	3199	1935	56	0.24	0.92	0.38
HSI+ISO	641	3428	1706	25	0.27	0.70	0.43
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Figures 4.9 and 4.10 show the final classified images for Test 2 and Test 3 using SVM. For both results, the number of FP is increased by the parts of the leaves that have brighter colors due to reflection. In both images, the midrib is also misclassified as tassel

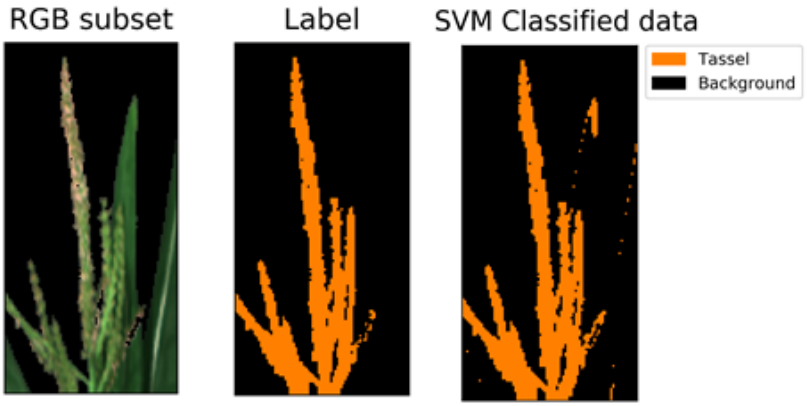

Figure 4.9. SVM classification results for CEPF HSI test 2 using the original image. From left to right: RGB color composite of HSI, ground truth for tassel label, and final classified data.

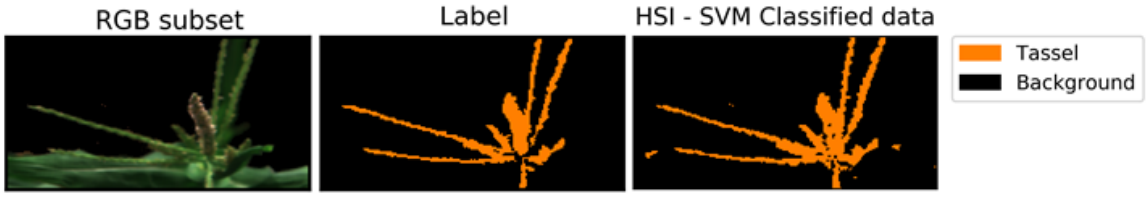

Figure 4.10. SVM classification results for CEPF HSI test 3 using the original image. From left to right: RGB color composite of HSI, ground truth for tassel label, and final classified data.

Table 4.6 presents the results for PhenoRover experiments. The F1-scores are considerably lower compared to the results presented by the CEPF data, which could be explained by the complexity of the in field data and the imbalance in the classes. The pixels for the tassel class are about 5 % of the total number of pixels in the image. The HSI without dimension reduction resulted in the best model, with a higher F1-score and a higher ability to classify pixels belonging to tassels.

Table 4.6. SVM results for tassel classification with PhenoRover data

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	394	14832	328	286	0.55	**0.58**	**0.56**
HSI+PCA	384	14797	363	296	0.51	0.56	0.54
HSI+LLE	138	16742	51	549	0.73	0.20	0.32
HSI+ISO	125	16738	55	562	0.69	0.18	0.29

Test 2							
HSI	654	19592	371	263	0.64	**0.71**	**0.67**
HSI+PCA	642	19588	375	275	0.63	0.70	0.66
HSI+LLE	390	18125	129	556	0.75	0.41	0.53
HSI+ISO	367	18146	108	579	0.77	0.39	0.52
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Figure 4.11 presents the final image classified for Test 2. The experiment using the original HSI for SVM classification had more TP than the other models, but also misclassified many pixels resulting in Type I and II errors

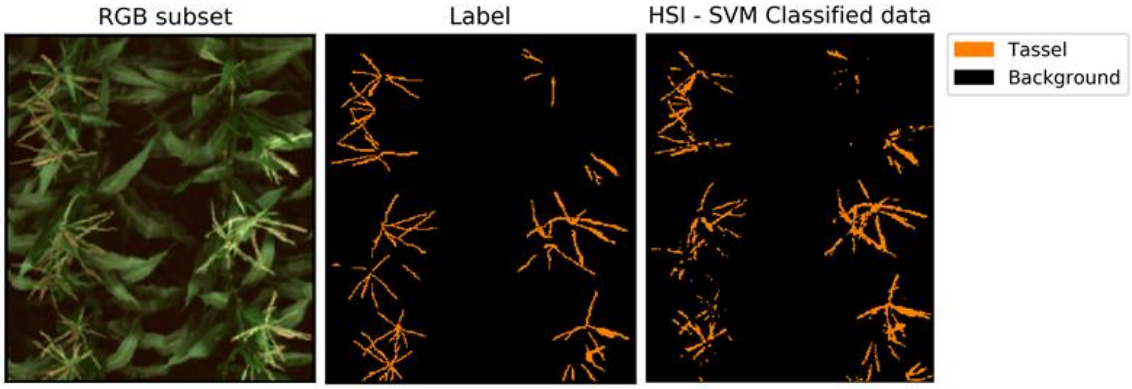

Figure 4.11. SVM classification results for PhenoRover HSI test 2 using the original image. From left to right: RGB color composite of HSI, ground truth for tassel label, and final classified data.

Tables 4.7 and 4.8 show the SVM classification results for the UAV data collected on August 8th and 10th, respectively. For both dates, the highest F1-scores correspond to the model using the original HSI without any dimension reduction.

Table 4.7. SVM results for tassel classification with UAV data 20190808

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	449	17481	237	193	0.65	**0.70**	**0.68**
HSI+PCA	419	17436	282	223	0.60	0.65	0.62
HSI+LLE	283	17437	72	415	0.80	0.41	0.54
HSI+ISO	320	17414	95	378	0.77	0.46	0.58
Test 2							
HSI	433	16090	227	155	0.66	**0.74**	**0.69**
HSI+PCA	384	16042	275	204	0.58	0.65	0.62
HSI+LLE	136	16274	120	375	0.53	0.27	0.35
HSI+ISO	150	16282	112	361	0.57	0.29	0.39
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Table 4.8. SVM results for tassel classification with UAV data 20190810

	TP	TN	FP	FN	Pr	Rc	F1
Test 1							
HSI	664	17013	301	626	0.69	**0.72**	**0.70**
HSI+PCA	611	16920	394	315	0.61	0.66	0.63
HSI+LLE	414	17014	105	555	0.80	0.43	0.56
HSI+ISO	437	17023	96	532	0.82	0.45	0.58
Test 2							
HSI	318	19896	237	318	0.57	**0.50**	**0.53**
HSI+PCA	252	19877	256	384	0.50	0.40	0.44
HSI+LLE	6	19959	2	641	0.75	0.01	0.02
HSI+ISO	19	19947	14	628	0.58	0.03	0.06
TP: True positive, TN: True negative, FP: False positive, FN: False negative, Pr: Precision, Rc: Recall, F1: F1 Score							

Figure 4.12 shows the result for Test 1 on August 10th corresponding to the highest F1-score for the UAV data. The SVM model produced better results for this experiment using the original HSI without dimension reduction.

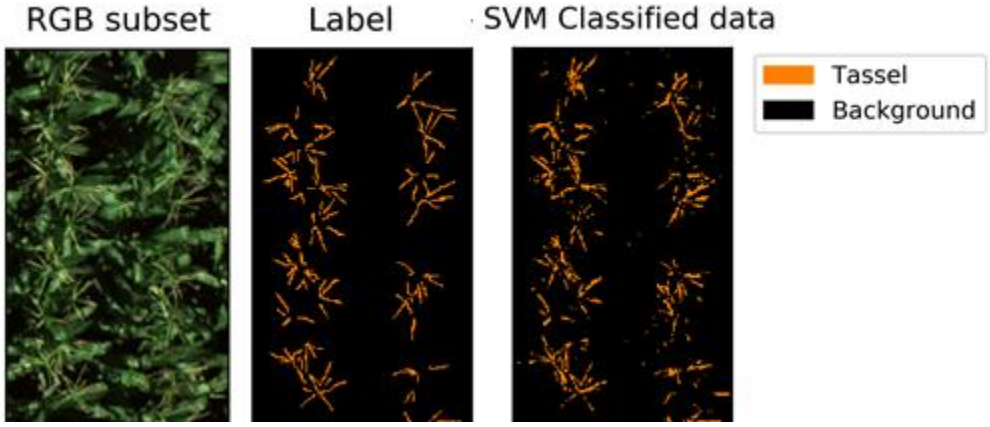

Figure 4.12. SVM classification results for UAV HSI test 2 using the original image collected on August 10th. From left to right: RGB color composite of HSI, ground truth for tassel label, and final classified data.

5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

The main goal of this research was to evaluate classification models for detecting tassels in maize crops using hyperspectral information. For this purpose, hyperspectral imagery were collected from a controlled facility and from crops in real environment. Three different systems were tested, 1) the conveyor imaging system from the Controlled Environment Phenotyping Facility (CEPF) with top view and side view images of the plants with a spatial resolution of 2.08 mm/pixel, 2) the PhenoRover, a wheel-based platform with sensor above canopy collecting data with 5 mm/pixel, and 3) the UAV multi-rotor system collecting data above canopy with 10 mm/pixel. Two classification models for tassels were successfully tested, SAM and SVM, using the original HSI and after feature reduction methods such as PCA, LLE, and Isomap.

By performing the experiments at different locations, we concluded that the controlled facility is an appropriate location to investigate the development of the tassels from the early-stage when the spectral signature varies the most from the other parts of the plant, to the late-stage when the anthers have completely extruded and the signature becomes similar to the midrib and ends of the leaves. The top view of the CEPF data foresaw some of the challenges that are inherent with the above canopy data collected from the platforms in the field, such as the similarity of the tassels with the midribs when anthers are extruding and the complexity of the tassels' shapes.

The difference of the spatial resolution was an important characteristic, the higher spatial resolution, the better results of the classifier. Nevertheless, capturing UAV data with spatial resolution higher than 10 mm/pixel is a challenging, time consuming process, and generates big amounts of data to store and to process. The spatial resolution is also limited by the altitude that the platform can be flown without impacting the tassels. Additionally, the spatial resolution is related to the purity of the pixels, that plays an important role in the classification problem, since the tassels' lateral spikelet flowers are very narrow and a spectral mixture could occur within pixels, reducing the ability to the algorithms to classify them correctly when the spatial resolution is too coarse.

By performing the experiments in the fields conditions using different systems and different spatial resolutions, we faced the challenges caused by the weather, such as variations in

sun angle and shadows, and interception from other plants due to tall plants in adjacent plots. Also, the tassels have a complex geometry, variation in size and shapes, and disparity in the window of flowering within plots (e.g. not all plants flower at the same time).

Regarding to the annotation required for the training labels in a pixel classification problem, collecting the labels is time consuming task and the shape and size of the tassels could lead to inaccurate labels that will impact the performance of the algorithms Also, the number of pixels for the tassel class is very small compared to the rest of the image, resulting in an unbalanced dataset for all images collected with the three different systems. The classification methods implemented in this study using reflectance could handle the imbalance between the classes very well, but the overall performance did not overpass the 70% accuracy in the field conditions.

By performing experiments with different feature extraction methods, it is concluded that the computation intensity for the non-linear dimension reduction methods depend on the number of neighbors and components to extract, and the dimension of the data to process (hyperspectral cube). It seems to be necessary to incur the computational expense when using SAM as the classifier, since the classification performance is better when using Isomap, for which there are computationally advantageous implementations. For SVM, the dimension reduction is not needed and yields good performance classifying tassels using hyperspectral data. Lastly, we envision that hyperspectral remote sensed data classification can ultimately contribute to a fast, accurate, and non-destructive method of monitoring tassels in maize crops.

5.2 Future Work

Extensions of this exploratory study in multiple areas could be fruitful:

- Further studies are needed to investigate the performance of the classification when using additional features from other technologies, such as LiDAR and adding spatial features. Additionally, specific information from the genetics could be introduced.
- The time and tremendously intensive computational overhead required for Isomap can be improved using Landmark Isomap, where selected points known as landmarks are used to compute the Isomap and embedding the lower dimensionality to the rest of the image.

- Annotation tools must be explored to extract reference data required for training and to evaluate the performance of the models from different labeling systems.
- Methods based on deep learning, and particularly convolutional neural networks that can exploit both the spatial and spectral domains of the hyperspectral data are potentially promising.

REFERENCES

[1] W. Rosa, Ed., “Transforming Our World: The 2030 Agenda for Sustainable Development,” in A New Era in Global Health, New York, NY: Springer Publishing Company, 2017.

[2] R. Bongiovanni and J. Lowenberg-DeBoer, “Precision Agriculture and Sustainability,” Precis. Agric., vol. 5, pp. 359–387, Aug. 2004, doi: 10.1023/B:PRAG.0000040806.39604.aa.

[3] J. W. White et al., “Field-based phenomics for plant genetics research,” Field Crops Res., vol. 133, pp. 101–112, Jul. 2012, doi: 10.1016/j.fcr.2012.04.003.

[4] N. Yu, L. Li, N. Schmitz, L. F. Tian, J. A. Greenberg, and B. W. Diers, “Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform,” Remote Sens. Environ., vol. 187, pp. 91–101, Dec. 2016, doi: 10.1016/j.rse.2016.10.005.

[5] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, Remote sensing and image interpretation. 2015.

[6] P. J. Pinter, Jr. et al., “Remote Sensing for Crop Management,” Photogramm. Eng. Remote Sens., vol. 69, no. 6, pp. 647–664, Jun. 2003, doi: 10.14358/PERS.69.6.647.

[7] S. Manfreda et al., “On the Use of Unmanned Aerial Systems for Environmental Monitoring,” Remote Sens., vol. 10, no. 4, p. 641, Apr. 2018, doi: 10.3390/rs10040641.

[8] P. S. Thenkabail, J. G. Lyon, and A. Huete, Hyperspectral indices and image classifications for agriculture and vegetation, vol. II. London: CRC Press, 2019.

[9] S. Sankaran et al., “Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review,” Eur. J. Agron., vol. 70, pp. 112–123, Oct. 2015, doi: 10.1016/j.eja.2015.07.004.

[10] L. Han, G. Yang, H. Yang, B. Xu, Z. Li, and X. Yang, “Clustering Field-Based Maize Phenotyping of Plant-Height Growth and Canopy Spectral Dynamics Using a UAV Remote-Sensing Approach,” Front. Plant Sci., vol. 9, Nov. 2018, doi: 10.3389/fpls.2018.01638.

[11] J. L. Hatfield and C. L. Walthall, “Meeting Global Food Needs: Realizing the Potential via Genetics × Environment × Management Interactions,” Agron. J., vol. 107, no. 4, pp. 1215–1226, 2015, doi: 10.2134/agronj15.0076.

[12] W. H. Maes and K. Steppe, “Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture,” Trends Plant Sci., vol. 24, no. 2, pp. 152–164, Feb. 2019, doi: 10.1016/j.tplants.2018.11.007.

[13] F. Liebisch, N. Kirchgessner, D. Schneider, A. Walter, and A. Hund, “Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach,” Plant Methods, vol. 11, no. 1, p. 9, Feb. 2015, doi: 10.1186/s13007-015-0048-8.

[14] “Crop Production 2018 Summary 02/08/2019,” Crop Prod., p. 132, 2018.

[15] P.-C. Cheng and D. R. Pareddy, “Morphology and Development of the Tassel and Ear,” in The Maize Handbook, M. Freeling and V. Walbot, Eds. New York, NY: Springer, 1994, pp. 37–47.

[16] R. L. (Bob) Nielsen, “Tassel Emergence & Pollen Shed - Corny News Network (Purdue University).” https://www.agry.purdue.edu/ext/corn/news/timeless/Tassels.html (accessed Apr. 16, 2020).

[17] A. E. Fonseca, M. E. Westgate, L. Grass, and D. L. Dornbos, “Tassel Morphology as an Indicator of Potential Pollen Production in Maize,” Crop Manag., vol. 2, no. 1, pp. 1–15, 2003, doi: 10.1094/CM-2003-0804-01-RS.

[18] A. L. Kaleita, B. L. Steward, R. P. Ewing, D. A. Ashlock, M. E. Westgate, and J. L. Hatfield, “Novel Analysis of Hyperspectral Reflectance Data for Detecting Onset of Pollen Shed in Maize,” Trans. ASABE, vol. 49, no. 6, pp. 1947–1954, 2006, doi: 10.13031/2013.22274.

[19] S. Srinivasan, S. V. Mirnezami, B. Ganapathysubramanian, P. Schnable, and T. Gonsalves, “An automated tassel detection and trait extraction pipeline to support high-throughput field imaging of maize,” presented at the ICVGIP, Dec. 2018.

[20] S. Lu, S. Inoue, H. Shibaike, S. Kawashima, S. Yonemura, and M. Du, “Detection potential of maize pollen release stage by using vegetation indices and red edge obtained from canopy reflectance in visible and NIR region,” J. Agric. Meteorol., vol. 71, no. 2, pp. 153–160, 2015, doi: 10.2480/agrmet.D-14-00035.

[21] F. Kurtulmuş and İ. Kavdir, “Detecting corn tassels using computer vision and support vector machines,” Expert Syst. Appl., vol. 41, no. 16, pp. 7390–7397, Nov. 2014, doi: 10.1016/j.eswa.2014.06.013.

[22] J. Xue and B. Su, “Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications,” Journal of Sensors, 2017. https://www.hindawi.com/journals/js/2017/1353691/ (accessed Apr. 16, 2020).

[23] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” IEEE Trans. Inf. Theory, vol. 14, no. 1, pp. 55–63, Jan. 1968, doi: 10.1109/TIT.1968.1054102.

[24] J. A. Benediktsson and P. Ghamisi, Spectral-Spatial Classification of Hyperspectral Remote Sensing Images. Artech House, 2015.

[25] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-Learning-Based Feature Extraction for Classification of Hyperspectral Data: A Review of Advances in Manifold Learning,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 55–66, Jan. 2014, doi: 10.1109/MSP.2013.2279894.

[26] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, “Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture,” Remote Sens. Environ., vol. 90, no. 3, pp. 337–352, Apr. 2004, doi: 10.1016/j.rse.2003.12.013.

[27] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. B. de Colstoun, and J. E. McMurtrey, “Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance,” Remote Sens. Environ., vol. 74, no. 2, pp. 229–239, Nov. 2000, doi: 10.1016/S0034-4257(00)00113-9.

[28] S. Liang, X. Li, and J. Wang, Advanced Remote Sensing: Terrestrial Information Extraction and Applications. Academic Press, 2012.

[29] C. Rodarmel and J. Shan, “Principal Component Analysis for Hyperspectral Image Classification,” Surv. Land Inf. Syst., vol. 62, Jan. 2002.

[30] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000, doi: 10.1126/science.290.5500.2319.

[31] J. B. Tenenbaum, “Mapping a manifold of perceptual observations,” in Proceedings of the 1997 conference on Advances in neural information processing systems 10, Denver, Colorado, USA, Jul. 1998, pp. 682–688, Accessed: Apr. 16, 2020. [Online].

[32] E. Pasolli, S. Prasad, M. M. Crawford, and J. C. Tilton, Advances in Hyperspectral Image Classification Methods for Vegetation and Agricultural Cropland Studies. Routledge Handbooks Online, 2018.

[33] S. T. Roweis and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000, doi: 10.1126/science.290.5500.2323.

[34] A. E. Maxwell, T. A. Warner, and F. Fang, “Implementation of machine-learning classification in remote sensing: an applied review,” Int. J. Remote Sens., vol. 39, no. 9, pp. 2784–2817, May 2018, doi: 10.1080/01431161.2018.1433343.

[35] N. Lee, Y. S. Chung, S. Srinivasan, P. Schnable, and B. Ganapathysubramanian, “Fast, automated identification of tassels: Bag-of-features, graph algorithms and high throughput computing,” presented at the International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, Nov. 2016, Accessed: Apr. 16, 2020. [Online]. Available: http://videolectures.net/kdd2016_lee_automated_identification/.

[36] W. Tang, Y. Zhang, D. Zhang, W. Yang, and M. Li, “Corn tassel detection based on image processing,” in International Workshop on Image Processing and Optical Engineering, 2011, doi: 10.1117/12.917672.

[37] F. A. Kruse et al., “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data,” Remote Sens. Environ., vol. 44, no. 2, pp. 145–163, May 1993, doi: 10.1016/0034-4257(93)90013-N.

[38] S. Rashmi, S. Addamani, and A. Ravikiran, “Spectral Angle Mapper Algorithm for Remote Sensing Image Classification,” 2014.

[39] J. L. Gage, N. D. Miller, E. P. Spalding, S. M. Kaeppler, and N. de Leon, “TIPS: a system for automated image-based phenotyping of maize tassels,” Plant Methods, vol. 13, no. 1, p. 21, Mar. 2017, doi: 10.1186/s13007-017-0172-8.

[40] “Purdue University Controlled Environment Phenotyping Facility.” https://ag.purdue.edu/cepf/ (accessed Apr. 16, 2020).

[41] G. Polder, G. W. A. M. van der Heijden, L. C. P. Keizer, and I. T. Young, “Calibration and Characterisation of Imaging Spectrographs,” J. Infrared Spectrosc., vol. 11, no. 3, pp. 193–210, Jun. 2003.

[42] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, “A transformation for ordering multispectral data in terms of image quality with implications for noise removal,” IEEE Trans. Geosci. Remote Sens., vol. 26, no. 1, pp. 65–74, Jan. 1988, doi: 10.1109/36.3001.

[43] H. He and Y. Ma, “Class Imbalance Learning Methods for Support Vector Machines,” in Imbalanced Learning: Foundations, Algorithms, and Applications, IEEE, 2013, pp. 83–99.

[44] K. Veropoulos, I. C. G. Campbell, and N. Cristianini, “Controlling the Sensitivity of Support Vector Machines,” in Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 1999, pp. 55–60.

[45] A. Tharwat, “Classification assessment methods,” Appl. Comput. Inform., Aug. 2018, doi: 10.1016/j.aci.2018.08.003.