Integrating Hyperspectral Imaging and Artificial Intelligence to Develop Automated Frameworks for High-throughput Phenotyping in Wheat

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE
UNIVERSITY OF MINNESOTA
BY

Ali Moghimi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ce Yang, Advisor
Peter M. Marchetto, Co-Advisor

February 2019

Acknowledgments

The path to this dissertation has been challenging yet incredible, and it would not have been possible without the collaboration and tremendous support of many wonderful individuals. I would like to express my sincere appreciation to my advisor, Dr. Ce Yang, for providing me this great opportunity and for her support and motivation during my PhD research. I would also like to thank my co-advisor, Dr. Peter Marchetto, for his encouragement and great support. I would like to acknowledge the rest of my committee members, Dr. James Anderson, Dr. Jonathan Chaplin, and Dr. Volkan Isler for providing valuable advice and guidance throughout this work.

My sincere thanks also go to my collaborators from various research labs, Dr. Shahryar Kianian and Dr. Marisa Miller from USDA-Cereal Disease Laboratory, and Dr. James Anderson and Ms. Susan Raynolds from Department of Agronomy and Plant Genetics.

I gratefully acknowledge the funding from the United States Department of Agriculture-Agricultural Research Service, the National Science Foundation, Minnesota Agricultural Experiment Station, and MnDRIVE. I would like to appreciate the graduate student funding provided by the MnDRIVE Initiative as well as the department of Bioproducts and Biosystems.

I would thank the current and past members of the Agricultural Robotics and Instrumentation Labs for their help. In particular, I should thank my great friend, Mr. Tyler Nigon for giving me endless help. I also thank my friends Mehdi, Fazel, Mina, Amir Hossain, Alireza, Daniel, Hamed, Nassir, and Mohammad. I am fortunate to have these wonderful friends far away from my hometown.

I am immensely appreciative of my parents, Ezzat and Reza, for their unconditional love, and my sisters, Mina and Marziyeh, for their tremendous support.

Last but certainly not least, I should thank my wife, Parisa, who has always believed in me and kindly offered her wholehearted support and genuine love.

Dedication

To my parents, ***Reza*** *and* ***Ezzat****,*

to my lovely wife, ***Parisa****,*

and to my adorable son, ***Parsa****.*

Abstract

The present dissertation was motivated by the need to apply innovative technologies, automation, and artificial intelligence to agriculture in order to promote crop production while protecting our environment. The main objective of this dissertation was to develop sensor-based, automated frameworks for high-throughput phenotyping of wheat to identify advanced wheat varieties based on three desired traits, including yield potential, tolerance to salt stress (an abiotic stress), and resistance to *Fusarium* head blight disease (a biotic stress). We leveraged the advantages of hyperspectral imaging, a sophisticated sensing technology, and artificial intelligence including machine learning and deep learning algorithms. Through integrating imaging and high-resolution spectroscopy, hyperspectral imaging provides valuable insights into the internal activity of plants, leaf tissue structure, and physiological changes of plants in response to their environment. Alternatively, advanced machine learning and deep learning algorithms are uniquely suited to extract meaningful features and recognize latent patterns associated with the desired phenotyping trait, and ultimately make accurate inferences and prediction.

In the first study (Chapter 2), we focused on salt stress phenotyping of wheat in a hydroponic system. A novel method was proposed for hyperspectral image analysis to assess the salt tolerance of four wheat varieties in a quantitative, interpretable, and don-invasive manner. The results of this study demonstrated the feasibility of quantitative ranking of salt tolerance in wheat varieties only one day after applying the salt treatment.

In the second study (Chapter 3), we developed an ensemble feature selection pipeline by integrating six supervised feature selection techniques to identify the most informative spectral bands from high-dimensional hyperspectral images captured for plant phenotyping applications. First, the spectral features were ranked based on their ability to discriminate salt-stressed wheat plants from healthy plants at the earliest stages of stress. The proposed method could drastically reduce the dimension of hyperspectral images from 215 to 15 while improving the accuracy of classifying healthy and stressed vegetation pixels by 8.5%. Second, a clustering algorithm was proposed to form six broad spectral bands around

the most prominent spectral features to aid in development of a multispectral camera.

In the third study (Chapter 4), we aimed to develop a phenotyping framework for *Fusarium* head blight (FHB), a devastating disease attacking small grain crops. The most informative spectral bands were identified to detect FHB-infected spikes. The results of this study revealed that a set of two broad spectral bands (766 nm and 696 nm) returns a classification accuracy of 99% in detecting FHB-infected spikes.

In the fourth study (Chapter 5), we developed an autonomous robotic framework for high-throughput yield phenotyping of wheat in the field. The data were collected by a hyperspectral camera mounted on an unmanned aerial vehicle flying over three experimental fields containing hundreds of wheat plots during two consecutive growing seasons. A deep neural network was trained to predict the yield of wheat plots and estimate the yield variation at a sub-plot scale. The coefficient of determination for predicting the yield at sub-plot and plot scale were 0.79 and 0.41with normalized root-mean-square error of 0.24 and 0.14, respectively.

In the fifth study (Chapter 6), we focused on developing a deep autoencoder network by leveraging a large unlabeled dataset (~ 8 million pixels) to learn an optimal feature representation of hyperspectral images in a low dimensional feature space for yield prediction. The result demonstrated that the trained autoencoder could substantially reduce the dimension of hyperspectral images onto a 3-, 5-, and 10-dimenionsal feature space with a mean squared error less than 7e-5, while retaining the relevant information for yield prediction.

At a higher level, this dissertation contributes to improving economic, ecological, and social impacts by improving crop production, reducing pesticides use, and properly leveraging salt-affected farmlands. From an environmental perspective, a cultivar with high yield potential and a cultivar resistant to FHB disease both promote sustainability in crop production and environment by reducing the required fertilizer and pesticide to meet the anticipated farmers' profit. The intelligent, automated phenotyping frameworks developed in this dissertation can help plant scientists and breeders identify crop varieties with the desired traits tailored around promoting crop production and mitigating food security concerns.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGL	above ground level
CFS	correlation-based feature selection
co(CS)	*Aegilops columnaris* KU11-2 (CS)
CS	Chinese spring
CV	cross validation
DN	digital number
DNN	deep neural network
DON	deoxynivalenol
EGI	excessive green index
FHB	*Fusarium* head blight
GNDVI	green normalized difference vegetation index
HSI	hyperspectral imaging
IMU	inertial measurement unit
MDPA	minimum difference of pair assignments
ML	machine learning
MSE	mean squared error
NDPSI	normalized difference plant senescence index
NDVI	normalized difference vegetation index
NRD	normalized reflectance difference
NRMSE	normalized root mean square error
OBIA	object-based image analysis
PC	principal component
PCA	principal components analysis
P-HSC	plot hyperspectral cube
QDA	quadratic discriminant analysis
ReLU	rectified linear unit
RMSE	root mean square error

SE	standard error
SFS-forward	sequential feature selection - forward selection
SIPI	structure insensitive pigment index
SL	spikes and leaves
sp(CS)	*Ae. speltoides aucheri* KU2201B
SS	soil and shadow
std	standard deviation
SVMAX	successive volume maximization
SVM-RFE	support vector machine – recursive feature elimination
TSAVI	transformed soil-adjusted vegetation index
UAV	unmanned aerial vehicles
VSM	vector-wise similarity measurement

Citation of Published Work

Two chapters of this dissertation have been published as follows:

Chapter 2

Moghimi, A., Yang, C., Miller, M. E., Kianian, S. F., & Marchetto, P. M. (2018). A Novel Approach to Assess Salt Stress Tolerance in Wheat Using Hyperspectral Imaging. *Frontiers in Plant Science*, *9*, 1182. https://doi.org/10.3389/fpls.2018.01182

Chapter 3

Moghimi, A., Yang, C., & Marchetto, P. M. (2018). Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging. *IEEE Access*, *6*, 56870–56884. https://doi.org/10.1109/ACCESS.2018.2872801

CHAPTER 1

Introduction

1.1 Problem statement

Agriculture requires to fulfill growing expectations from various perspectives. From food security perspective, agriculture is expected to meet the increasing demand for food because of the growing population. More than 2 billion people will be born into the world by 2050 (United Nations, Department of Economic and Social Affairs, 2017) with the moral right to have access to sufficient and nutritious food. From the environmental point of view, agricultural practices are expected to be environmentally sustainable to reduce its remarkable negative impact on the environment. These challenges are compounded by various limitations including climate change, lack of productive land, water shortage, and labor unavailability.

According to the United Nations Food and Agriculture Organization (FAO), about 10.9% people in the world (~ 821 million), were suffering from chronic undernourishment in 2017 (FAO, IFAD, UNICEF, WFP, & WHO, 2018). Figure 1.1 demonstrates that the percentage and number of undernourished people have been increased since 2014, reaching an

estimated 821 million in 2017. This includes people in the developed countries such as the US. A 2018 report from the United States Department of Agriculture reported that an estimated 11.8% of U.S. households (about 15.0 million households) were food insecure in 2017, highlighting the high significance of this issue in the United States (Coleman-Jensen, Rabbitt, Gregory, Singh, & Singh, 2018).

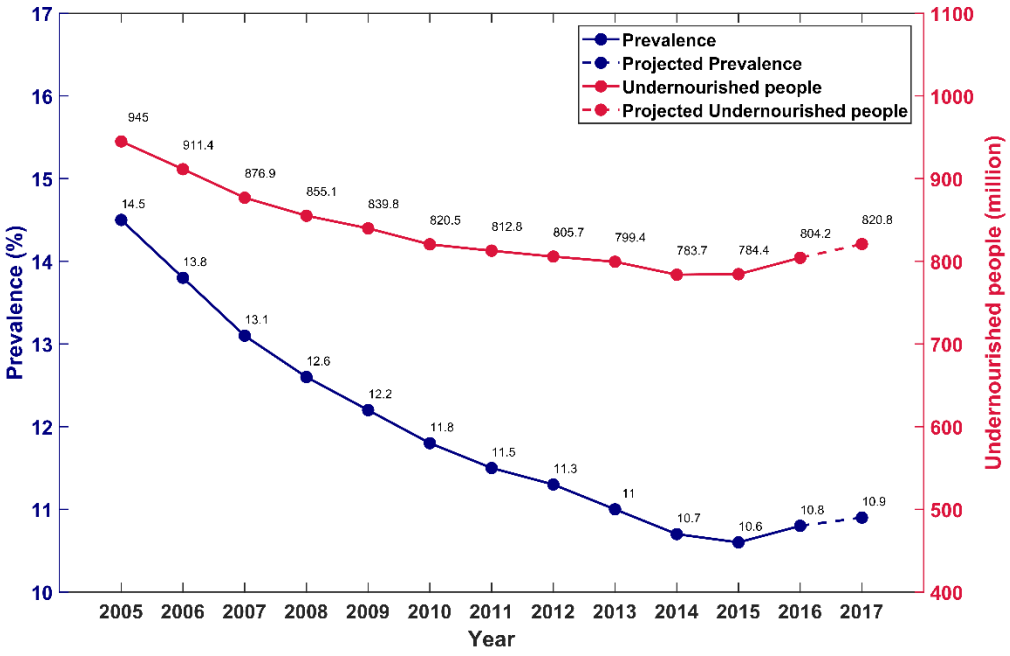

Figure 1.1. The number of undernourished people in the world from 2005 to 2016, and an estimated value for 2017. The percentage and number of undernourished people have been increased since 2014, reaching an estimated 821 million in 2017 (FAO, IFAD, UNICEF, WFP, 2018). The numbers for 2017 were estimated values and illustrated by dotted lines.

Considering the increasing world population and subsequent demand for food, crop production should double by 2050 (Tilman, Balzer, Hill, & Befort, 2011), indicating the average rate of yield increase of crops should be 2.4% annually (Ray, Mueller, West, & Foley, 2013) – the current average rate of increase is only 1.3% (Jose Luis Araus, Li, Parry, & Wang, 2014; Ray et al., 2013). In another study, it was argued that the crop production needs to increase about 59-98% between 2005 and 2050 if we are to meet the growing demand for food by 2050 (Valin et al., 2014). These statistics noticeably indicate an urgent

need for further efficiency improvement in crop production to alleviate the global concern of food security.

Agriculture is responsible for a tremendous adverse impact on the environment with significant overall economic and social costs. Pesticides have significant undesirable effects on the ecosystem by contaminating food, soil, and surface/ground water that living organisms rely on (Aktar, Sengupta, & Chowdhury, 2009). In addition, pesticides can cause a wide range of acute and chronic human health risks, from skin irritation to cancer (Centers for Disease Control and Prevention, 2014). Extensive use of chemical fertilizer is another major source of environmental pollution. Studies conducted in the United States and Europe estimated the annual cost of nitrogen pollution to be in the range of US$59–US$340 billion in the US (Sobota, Compton, McCrackin, & Singh, 2015) and €35–€230 billion in Europe (Van Grinsven et al., 2013). These costs undoubtedly demonstrate the environmental impact of intensive agricultural practices, conducted in the expectation of more yield.

Developing elite varieties of crops through breeding is one of the main solutions to further improve crop production in order to meet the projected demand for food. These elite varieties entail desirable traits such as high yield potential and being resistant to abiotic/biotic stresses. Studies have shown that advances in breeding were responsible for a substantial increase in crop production over time (De La Fuente, Frei, & Lübberstedt, 2013; Duvick, 2005; Lopes et al., 2012), indicating the importance of selecting advanced varieties.

Thanks to the advances in DNA sequencing and genotyping, development of plant genotypes is rapid and inexpensive (Peleman & van der Voort, 2003), providing a wide range of genetic information aimed at improving qualitative and quantitative traits. To enhance the genetic gain and genetic selection, accurate phenotyping[1] is required. Comprehensive high-throughput phenotyping aids breeders and geneticists to identify the underlying genetic factors and observe the consequence of genetic modification on a desirable trait. However, phenotyping has failed to keep the pace with the progress in

1. Phenotyping is a quantitative assessment of plant structure and function related to a desirable trait.

genotyping (Crain, Mondal, Rutkoski, Singh, & Poland, 2018). The conventional phenotyping approaches rely on time-consuming, labor intensive, and expensive tasks, such as repeated field measurements and visual inspection which could be subject to human errors and bias. Despite the recent attempts for developing effective high-throughput phenotyping pipelines, it remained as a bottleneck, in particular, developing phenotyping pipelines for yield potential and stress tolerance (José Luis Araus, Kefauver, Zaman-Allah, Olsen, & Cairns, 2018).

1.2 Objectives

This dissertation was motivated by the need for an automated phenotyping framework to identify advanced wheat lines[1] in an efficient, cost-effective, interpretable, and non-invasive manner. This is a multi-disciplinary study integrating remote sensing, computer vision, machine learning, and deep learning to address a challenge in plant science.

The main objective of this dissertation is to develop a sensor-based, automated framework for high-throughput phenotyping of wheat aimed at identifying advanced wheat lines based on three major traits, including yield potential, tolerance to salt stress (abiotic stress), and resistance to Fusarium head blight (biotic stress). Hyperspectral imaging was used as a sophisticated sensing technology to acquire data from wheat plants with high spatial and spectral resolution. Subsequent to the required pre-processing steps for image analysis, advanced machine learning and deep learning algorithms were deployed to extract meaningful features and recognize latent patterns associated with the desired phenotyping trait. The remainder of this dissertation is organized as follows:

Chapter 2 focuses on salt stress phenotyping of wheat in a hydroponic system. In this chapter, we propose a novel method for identifying salt tolerant wheat varieties to mitigate yield losses due to salinity, and to ultimately maintain or improve production on saline soils. The objectives of this chapter are to:

i. Rank wheat lines based on their tolerance to salt stress.

1. A line is a wild species, landrace, historical cultivar or a possible new genotype that is still being tested for its potential value in a given environmental setting.

ii. Assess the difference between the salt tolerance of various lines to attain a quantitative ranking rather than a qualitative ranking.
iii. Evaluate the feasibility of the precise ranking of wheat lines as early as one day after applying salt treatment.

Chapter 3 introduces selecting prominent spectral bands from high-dimensional hyperspectral images, tailored around practical applications in plant phenotyping. In this chapter, an ensemble feature selection method is proposed to identify the most informative spectral bands for salt stress phenotyping. The objectives of this chapter are to:

i. Develop an ensemble feature selection pipeline to aggregate the benefits of multiple feature selection approaches, therefore increasing the stability and accuracy of selecting predominant spectral features from hyperspectral images.
ii. Rank the spectral features based on their ability to discriminate salt-stressed wheat plants from healthy plants at the earliest stages of stress. The purpose of ranking spectral features was to form six broad spectral bands around the most prominent features to aid in the development of a multispectral camera.
iii. Use these multispectral bands to assess the salt tolerance of four wheat lines in the context of phenotyping and evaluate results based on findings in Chapter 2.

Chapter 4 focuses on developing a framework for *Fusarium* head blight (FHB) phenotyping. FHB is arguably among the most widespread and devastating diseases, attacking small grain crops throughout the world. In this chapter, the most informative spectral bands for classifying healthy spikes and FHB-infected spikes are identified using the ensemble feature selection method developed in Chapter 3. The objectives of this chapter are to:

i. Evaluate the performance of the developed ensemble selection method on a different image dataset collected for phenotyping a different trait.
ii. Identify a set of optimal spectral bands for FHB phenotyping.

Chapter 5 aims to develop an autonomous robotic framework for high-throughput yield phenotyping of wheat in the field. The data were collected by a hyperspectral camera mounted on a UAV flying over three experimental wheat plots during two consecutive growing seasons. In this chapter, a deep neural network was trained to predict the yield of wheat plots. Furthermore, we investigated the feasibility of estimating yield variation at the sub-plot scale by leveraging the high spatial resolution of images. The objectives of this chapter are to:

i. Predict the yield of wheat plots to identify high-yielding varieties using an automated pipeline.
ii. Estimate the yield variation with a plot to study the marginal effects on yield production. This provides further insights into the process of screening.

Chapter 6 focuses on dimensionality reduction of the input feature for yield prediction. In this chapter, a deep autoencoder network is proposed to project the hyperspectral images onto a low dimensional feature space. The objectives of this chapter are to:

i. Train a deep autoencoder to project the hyperspectral data onto lower dimensional feature spaces while retaining invaluable information for yield prediction.
ii. Compare the performance of the deep neural network in yield prediction based on the new datasets.

Chapter 7 presents the conclusion and future directions for further research.

1.3 Overview

The remaining sections of this chapter briefly overview the main components of this dissertation.

1.3.1 Wheat

Wheat is the largest cultivated cereal crops in the world as it is annually cultivated on more than 220 million hectares with 763 million metric tons production (United States Department of Agriculture, 2018c). Wheat plays a vital role both in human and animal diet. It provides about 20% of the overall dietary proteins and calories throughout the world (Shiferaw et al., 2013). Moreover, wheat is a major source of starch, energy, dietary fiber, and phytochemicals (Shewry & Hey, 2015).

In the United States, wheat is the third largest crop after corn and soybean, in terms of either value or acreage (Vocke & Ali 2013). In 2018, wheat production in Minnesota was projected to be 98.3 million bushels, which is 30 percent more than 2017 (United States Department of Agriculture, 2018b).

Nevertheless, genetic gain[1] in yield of wheat, one of the major crops, was reported to be less than 1%, far behind the necessary yield increase (i.e., 2.4%) (Crain et al., 2018; Ray et al., 2013; Reynolds et al., 2012). Other studies even claimed that wheat yields have plateaued in various regions of the world (Acreche, Briceño-Félix, Sánchez, & Slafer, 2008; José Luis Araus et al., 2018; Sadras & Lawson, 2011), indicating the importance of phenotyping for developing elite wheat varieties.

1. Genetic gain: “amount of increase in performance achieved annually per unit time through artificial selection.” (José Luis Araus et al., 2018)

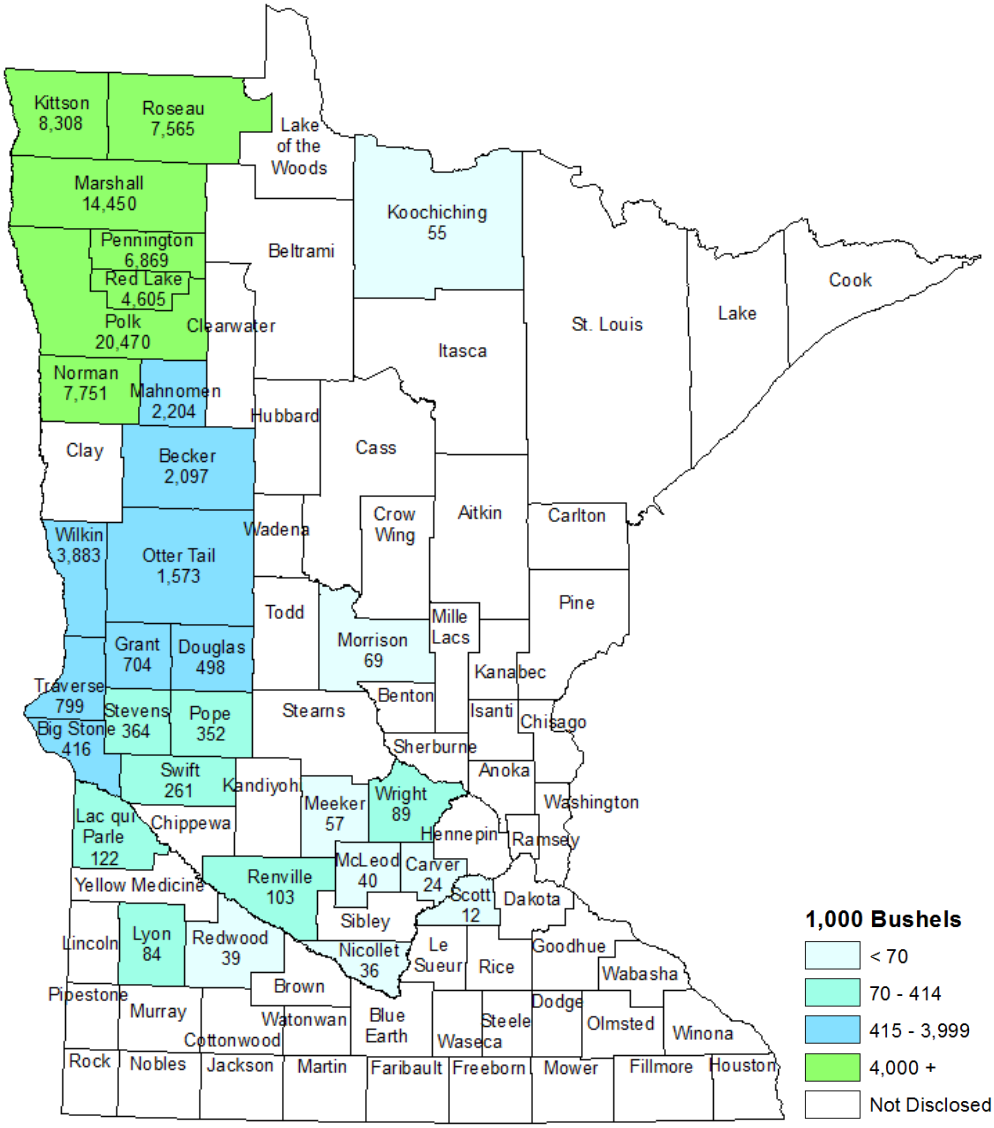

Figure 1.2. Spring wheat production in Minnesota in 2018 (United States Department of Agriculture, 2018a).

1.3.2 Hyperspectral imaging

Hyperspectral imaging (HSI) integrates imaging and high-resolution spectroscopy whereby a continuous spectrum is scanned for each pixel, typically across the visible and near infrared range of electromagnetic spectrum. In essence, a hyperspectral image is a high-dimensional data cube (Figure 1.3) in which each pixel contains both spatial (location of pixels: x- and y-dimension), and spectral information (a continuous reflectance captured across a range of electromagnetic region: z-dimension). HSI has drawn substantial attention for plant phenotyping because it provides the ability to investigate the physiological dynamics of plants caused by environmental variables (Wahabzada et al., 2016).

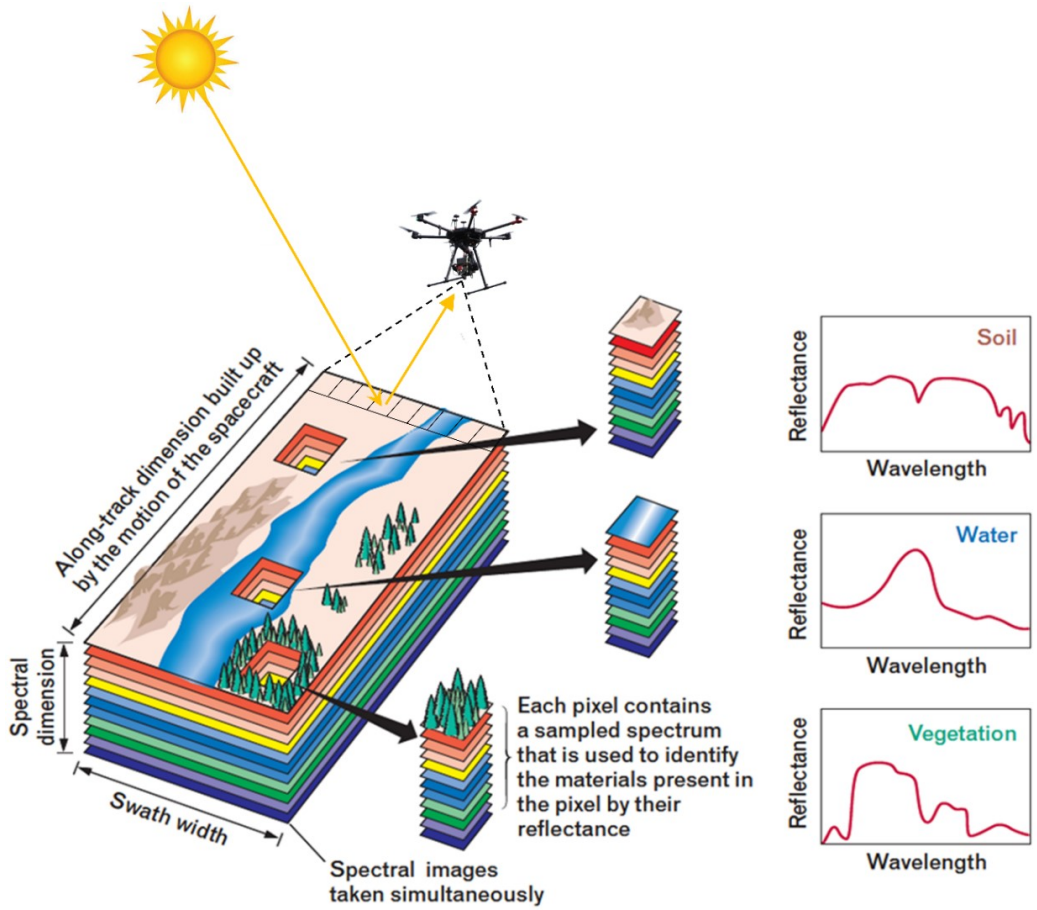

Figure 1.3. The concept of hyperspectral imaging and terminologies[1].

1. Modified from source: http://www.markelowitz.com/Hyperspectral.html

Recently, hyperspectral sensors have become more compact, more lightweight, and more available, and this has introduced new opportunities for field applications because hyperspectral images can be captured via unmanned aerial vehicles. In this dissertation, we utilize HSI because of its unique aspect among other imaging techniques. L. Li, Zhang, & Huang (2014) reviewed the application and limitation of several imaging technologies for plant phenotyping applications.

Table 1.1. Applications and limitations of popular imaging techniques for plant phenotyping.

Imaging techniques	Applications	Limitations
Visible imaging	Growth dynamics, Shoot biomass, Yield traits, Panicle traits, Root architecture, Imbibition and germination rates, leaf morphology, seedling vigor, coleoptile length and biomass at anthesis, seed morphology, root architecture	Only provides plant physiological information
Fluorescence imaging	Photosynthetic status, indirect measurement of biotic or abiotic	Difficult to analysis complicated whole-shoot of non-rosette species; pre-acclimation conditions required
Thermal imaging	Surface temperature; stomatal conductance water stress induced by biotic or abiotic factors	Imaging sensor calibration and atmospheric correction are often required; sound physics-based results interpretation needed
Spectral imaging	Water content composition parameters for seeds; leaf area index; leaf and canopy health status; panicle health status; leaf growth; coverage density	Sensor calibration required; cost, large image data sets for hyperspectral imaging, complex data interpretation

Among these techniques, hyperspectral imaging (HSI) is uniquely suited to provide insights into the internal activities of plants, leaf tissue structure, leaf pigments, and water content (A.-K. Mahlein, Steiner, Hillnhütter, Dehne, & Oerke, 2012). In this dissertation, we deploy advanced machine learning algorithms and deep learning to address these limitations associated with spectral imaging and leverage the unique advantages and full potential of HSI.

1.3.3 Machine learning

In recent years, the emergence of sophisticated sensing technologies and autonomous robotic platforms enables scientists to collect large datasets with high spectral-, spatial-, and temporal-resolution from greenhouse scale to large fields. These massive datasets, generated continuously in the context of phenotyping, require more advanced methods from pre-processing to post-processing steps, causing a paradigm shift from data collection to data analysis. Ignoring data analysis entails a tremendous loss of information that is required to perceive plant development and response to the environment and the phenotyping data.

Machine learning algorithms have shown considerable promise in developing more efficient and effective pipelines for analysis of large datasets (A. Singh, Ganapathysubramanian, Singh, & Sarkar, 2016). The stagnation of data analysis makes using advanced machine learning algorithms more urgent to improve the efficiency of breeding, and subsequently crop production. Machine learning is an application of *Artificial Intelligence* to discover underlying patterns and make inferences from a large dataset. Such patterns might help scientists make appropriate decisions and predictions for future incidents.

For the analysis of more complex and even larger datasets, deep learning has recently drawn substantial attention. Below is a definition for deep learning by LeCun, Bengio, & Hinton (2015):

"Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer."

Similar to various domains, deep learning has proved a great potential in plant phenotyping. In 2018, A. K. Singh, Ganapathysubramanian, Sarkar, & Singh (2018) reviewed the trends and future perspectives of deep learning for plant stress phenotyping. Two figures from their papers are presented here to illustrate the trend of publications related to machine

learning and deep learning (Figure 1.4) and to demonstrate a flowchart of a deep learning framework from collecting images to making decisions in the context of phenotyping (Figure 1.5).

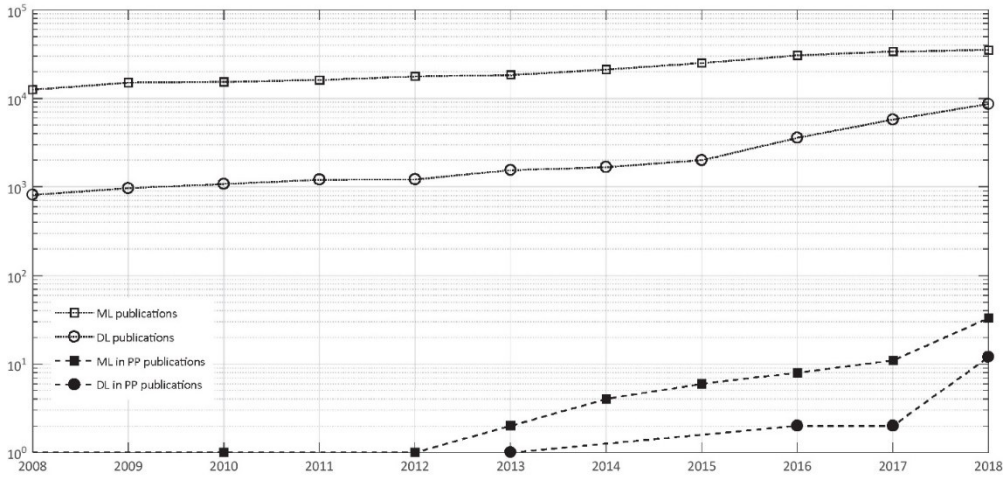

Figure 1.4. "Comparative Assessment of Publications Related to Machine Learning and Deep Learning. The open symbols indicate all topics, while the closed symbols indicate plant phenotyping topics. Notice the log scale for the y-axis. Source: Web of Science with keyword search 'Machine Learning'; 'Deep Learning'; 'Machine Learning and Plant Phenotyping'; and 'Deep Learning and Plant Phenotyping' using the 2008–2018 period. DL; deep learning; ML; machine learning; PP; plant phenotyping." (A. K. Singh et al., 2018).

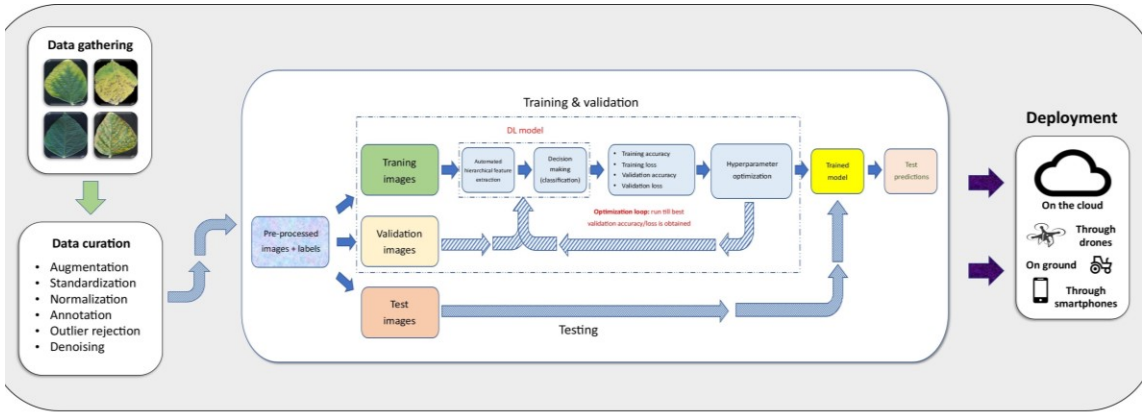

Figure 1.5. "An Illustration of a Deep Learning (DL) Tool Chain from Data Gathering to Decision Making. A key first step is to gather a large, diverse set of data. The data is then curated. Curation includes standardization, outlier rejection, some simple denoising and image preprocessing, and data augmentation. This dataset is then split into training, validation, and testing subsets. The training and validation subsets are used in training the DL architecture/network. Training the DL network essentially means optimizing the parameters (weights, biases) of the network such that the network accurately learns the mapping from the input data to the desired output label. The trained model is then finally tested on the unseen test image data subset. After successful completion of this testing, the DL network can be deployed for inference. While training the DL model can be resource intensive, deploying a trained model is relatively simple." (A. K. Singh et al., 2018).

CHAPTER 2

A novel approach to assess salt stress tolerance in wheat using hyperspectral imaging

Ali Moghimi[1], Ce Yang[1], Marisa E. Miller[2, 3], Shahryar Kianian[2, 3], Peter Marchetto[1]

[1] Department of Bioproducts and Biosystems Engineering, University of Minnesota, MN, USA

[2] Cereal Disease Laboratory, USDA-ARS, MN, USA

[3] Department of Plant Pathology, University of Minnesota, MN, USA

Abstract

Salinity stress has significant adverse effects on crop productivity and yield. The primary goal of this study was to quantitatively rank salt tolerance in wheat using hyperspectral imaging. Four wheat lines were assayed in a hydroponic system with control and salt treatments (0 and 200 mM NaCl). Hyperspectral images were captured one day after salt application when there were no visual symptoms. Subsequent to necessary preprocessing tasks, two endmembers, each representing one of the treatment, were identified in each image using successive volume maximization. To simplify image analysis and interpretation, similarity of all pixels to the salt endmember was calculated by a technique proposed in this study, referred to as vector-wise similarity measure. Using this approach allowed high-dimensional hyperspectral images to be reduced to one-dimensional gray-scale images while retaining all relevant information. Two methods were then utilized to analyze the gray-scale images: minimum difference of pair assignments and Bayesian method. The rankings of both methods were similar and consistent with the expected ranking obtained by conventional phenotyping experiments and historical evidence of salt tolerance. This research highlights the application of machine learning in hyperspectral image analysis for phenotyping of plants in a quantitative, interpretable, and non-invasive manner.

Keywords: Bayesian inference, histogram distance, hyperspectral imaging, image processing, machine learning, plant phenotyping, salt stress, wheat.

2.1 Introduction

Salinity stress is a major abiotic stress that has significant adverse effects on crop productivity and yield. These negative effects include interference of root function in absorbing water, as well as the prevention of physiological and biochemical processes such as nutrient uptake and assimilation (Carillo, Annunziata, Pontecorvo, Fuggi, & Woodrow, 2011). Unfortunately, many regions around the world are facing a rapid increase in soil salinity and sodicity. It is estimated that at least 0.3 million hectares of farmland is becoming unusable annually, and another 20 to 46 million ha are suffering decreases in production potential each year (FAO and ITPS, 2015). Nevertheless, even with lower yield potential, these salt-affected farmlands must continue to produce crops so the increasing demand for food can be met and food security concerns mitigated. The lack of new productive land threatens food security, thus the productivity of existing marginal lands must improve.

There are numerous potential solutions for mitigating salt stress, including genetic engineering of plants with salt tolerance (Agarwal, Shukla, Gupta, & Jha, 2013; Wei et al., 2017) and application of exogenous compounds such as hormones, growth regulators, or nanoparticles (Mbarki et al., 2018). Among the potential solutions, selecting plant varieties with high tolerance to salt stress appears to be one of the most promising approaches in utilizing salt-affected soil for crop production (Ondrasek, Rengal, & Veres, 2011; Sytar et al., 2017). Although some progress has been made using measurement of photosynthetic parameters as a more sensitive method to screen for salt tolerance (Kalaji et al., 2018; Sun et al., 2016), the standard process of selecting either conventionally–bred or transgenic salt-tolerant crop lines relies on laborious phenotyping to assess tolerance. Despite the emergence of innovative platforms, precise instrumentation, sophisticated sensors, and rapid development of advanced machine learning and deep learning algorithms, phenotyping is still a barrier to variety development. While DNA sequencing and plant genotyping has rapidly evolved, phenotyping still depends on conventional methods which are not as accurate or efficient. In general, these techniques can be time-consuming, destructive, subjective, and costly. In recent years, non-contact sensing technology, in particular imaging,

has been extensively deployed as a potential substitute for conventional methods for high-throughput phenotyping of plants. Thanks to the advances in developing sensors with high spatial and spectral resolution, different imaging sensors including visible, fluorescence, thermal, and spectral imaging are available, each tailored for specific applications. Each of these sensing technologies can vary in their application, as well as limitations, in the context of plant phenotyping (Li, Zhang, & Huang, 2014). Among these techniques, hyperspectral imaging (HSI) is uniquely suited to provide insights into the internal activities of plants, leaf tissue structure, leaf pigments, and water content (A. K. Mahlein, Oerke, Steiner, & Dehne, 2012). HSI also provides the ability to investigate physiological dynamics of plants caused by environmental variables (Wahabzada et al., 2016), and consequently has drawn substantial attention for plant phenotyping (Kuska et al., 2015).

Few research studies have attempted to identify salt stress in plants using hyperspectral reflectance. In a previous study, three potential indicators including blue, yellow, and red edge positions of vegetation reflectance spectrum were calculated to detect four levels of salt stress imposed on Chinese castor bean (Li, Wan, Zhou, Yang, & Qin, 2010). The authors claimed that blue and red edge positions shift to the shorter wavelength in response to salt stress and therefore could be used to detect salt stress. However, the pattern of shifting to the shorter wavelength was not consistent across all treatments and hence further research is required. In another paper, the application of HSI to identify plant tolerance to salt stress in a high throughput phenotyping system was reviewed (Sytar et al., 2017). They concluded more efficient and fully automated methods are required to analyze complex hyperspectral images.

To leverage the full potential of HSI, a large high-quality hyperspectral dataset and several preprocessing tasks are necessary (e.g., radiometric calibration, normalization, mixed pixel filtering, etc.). However, there are two major challenges that hamper the application of HSI.

The first major challenge is accounting for the variance caused by the complex interaction between incident light and leaf surfaces due to non-Lambertian reflectance properties. The direction of reflected light is a function of leaf geometry, including leaf angle and curvature. Several researchers have focused on pre-processing techniques to address the

problems related to leaf angle and curvature (Behmann et al., 2015; Makdessi et al., 2017; Wendel & Underwood, 2017). One method to resolve this problem is to generate a high-resolution 3D representation of plants by upfront geometric calibration of the hyperspectral camera (Behmann et al., 2015). However, this proposed method depends on highly intensive processing and is only suitable for close-range imaging.

The second major challenge is analyzing the complex and high-dimensional hyperspectral images in order to extract meaningful features and recognize latent patterns associated with the desired phenotyping trait in a more interpretable manner. To address this issue, machine learning (ML) and deep learning algorithms can be leveraged. Recent reviews of various ML algorithms emphasize the potential of these methods in the context of agriculture and provide guidelines for plant scientists to deploy them (Bauckhage & Kersting, 2013; Coppens, Wuyts, Inzé, & Dhondt, 2017; A. Singh et al., 2016). Singh et al. (2016) reported that ML algorithms are a promising approach to analyze large datasets generated by sophisticated imaging sensors (e.g., hyperspectral cameras) mounted to platforms that can cover large areas. Despite several studies that focus on the application of HSI for plant phenotyping, research is limited in the context of handling, processing, and analyzing hyperspectral images.

This research was motivated by the need to identify salt tolerant wheat lines to mitigate yield losses due to salinity, and to ultimately maintain or improve production on saline soils. The objectives of this study were to (i) rank wheat lines based on their tolerance to salt stress, (ii) assess the difference between the salt tolerance of lines to attain a quantitative ranking rather than a qualitative ranking, and (iii) evaluate the feasibility of precise ranking of wheat lines as early as one day after applying salt treatment. We hypothesized that the spectral response of wheat leaves experiencing salt stress would deviate from the control leaves even one day after adding the stress, and this deviation would be larger for a susceptible line compared to a salt tolerant line. To the best of our knowledge, no previous study has investigated early detection of salt tolerant plant lines using advanced phenotyping tools and approaches. This research proposes a machine learning approach to analyze hyperspectral images of wheat lines to rank their salt stress

tolerance in a quantitative, interpretable, and non-destructive manner while reducing cost, time, and labor input.

2.2 Materials and Methods

2.2.1 Sample preparation and conventional phenotyping for salt tolerance screening

To develop analytical methods for analysis of hyperspectral images, four bread wheat (*Triticum aestivum*) lines were selected with varying levels of salt tolerance. The cultivar Kharchia was included as it is historically known to maintain a stable harvest index and yield well in high salt conditions (Munns, James, & Läuchli, 2006; Schachtman, Lagudah, & Munns, 1992), and the salt-sensitive cultivar Chinese Spring (CS) was selected as well (Zhang et al., 2016). Two additional "unknown" lines were selected for screening from a set of wheat alloplasmic lines developed in Japan (Tsunewaki, Wang, & Matsuoka, 1996; Tsunewaki, Wang, & Matsuoka, 2002). Alloplasmic lines are created by substitution backcrossing to replace the cytoplasmic genomes of one species (in this case, bread wheat) with those of another (in this case, wild wheat relatives) while maintaining the original nuclear genome background, and have shown promise for improving stress tolerance and other developmental traits (Liberatore, Dukowic-schulze, Miller, Chen, & Kianian, 2016). The two alloplasmic lines selected were *Aegilops columnaris* KU11-2 (CS) (abbreviated co(CS) hereafter) and *Ae. speltoides aucheri* KU2201B (CS) (abbreviated sp(CS) hereafter) with the cytoplasmic genome type preceding the nuclear genome background, which in this case is Chinese Spring (CS).

Screening was performed in a hydroponic system in a Conviron growth chamber to ensure uniform conditions. Hydroponic systems are commonly used to screen plants for salt tolerance, including wheat. In all experiments, growth conditions in the Conviron were set at 22ºC during light conditions and 18ºC during the dark, 16h photoperiod, 375 µmol m^{-2} s^{-1} light intensity, and 50% relative humidity. Three hydroponic tanks were used per treatment (control treatment: 0 mM NaCl and salt treatment: 200 mM NaCl). Each hydroponic tank contained a grid of 16 Cone-tainers (Ray Leach brand) filled with perlite.

Within each tank, there were four genotypes each with four individual replicates (4 cone-tainers x 4 genotypes). For each treatment (salt or control), there were three replicate tanks; hence, there were a total of 48 (3 replications x 4 Cone-tainers x 4 genotypes) Cone-tainers for each treatment. The grid was placed into a tank just large enough to hold the grid, and 20L of hydroponic solution was used per tank. Genotypes were randomly assigned to positions in each cone-tainer grid using the sample and matrix functions in R (version 3.4.0). Aeration was supplied to each tank with an aquarium pump and two large airstones (at either end of the tank). Lines were transplanted into the tanks, and the lights and aeration were switched on 24 hours after transplanting. When leaf 1 emerged, ¼ strength Hoagland's solution (PhytoTech H353) was added, and the pH was adjusted to 6.5. When leaf 2 emerged, the Hoagland's was increased to ½ strength in all tanks and $CaCl_2$ was added to the tanks destined for salt treatment in a 15:1 molar ratio of NaCl to $CaCl_2$. When leaf 4 emerged, salt was added to the salt tanks gradually over 2 days to reach a final concentration of 200 mM. The water level and pH (to 6.5) were adjusted 3 times per week throughout the experiment.

To compare the salt tolerance of four wheat lines, both aerial and root biomass was harvested separately for each individual plant two weeks after salt treatment was applied. Plant matter was dried at 60-65ºC for 4 days and then weighed. Dry weight data were analyzed in R (version 3.4.0) using ANOVA (car package, version 2.1-5) and linear mixed-effect modeling (nlme package, version 3.1-120). For linear mixed-effect modeling analysis, dry weight was considered as the response in the analysis, salt level and genotypes were considered as fixed effects, and tank number was considered as a random effect. Model results were identical if tank position was considered as a nested-random effect of tank number, thus the results with tank number were used as the only random effect. To compare the response of the alloplasmics to the response of the euplasmic parents when the salt level is changed, the coefficient estimates of the lme model were examined.

2.2.2 Hyperspectral image acquisition

All tanks were transferred from the Conviron to greenhouse to take hyperspectral images under natural light conditions. To ensure that each hyperspectral image contained both salt

and control plants of a single wheat line, individual cone-tainers were removed from the randomized grid and arranged as salt and control tanks, each containing 12 cone-tainers as shown in Figure 2.1. After imaging, plants were placed back into their original randomized grid positions to avoid confounding effects from changing the tank position during the experiment.

Image acquisition was done approximately 24 hours after salt application when there were no visual symptoms. To reduce the effects of sun angle and shade, images were captured close to noon (i.e., between 11:00 and 13:00 local time). A push broom (along-track scanner) hyperspectral camera (PIKA II, Resonon, Inc., Bozeman, MT 59715, USA) was used for image acquisition, which required constant movement during image capture for two-dimensional spatial information to be accurate. A glide gear slider was used to mount the camera on a horizontal bar. A Dayton DC gearmotor (model: 2L008, Dayton electric Mfg Co. Lake Forest, IL 60045, USA) was utilized to move the slider along at a set speed, with the camera oriented to face downwards. All of this was done as per (Moghimi, Yang, Miller, Kianian, & Marchetto, 2017). The camera scanned over 240 spectral channels ranging from 400 nm to 900 nm with a spectral resolution of about 2.1 nm and captured 640 pixels in the cross-track direction (i.e., perpendicular to the direction of camera motion). The number of pixels in the along track direction was set to 2000 to assure both control and salt tanks of each line were captured in a single image. Therefore, the pixel size of each hyperspectral image, also known as hyperspectral data cube, was 2000×640×240, meaning each pixel has a 240 dimensional feature vector.

The frame rate of the camera was adjusted based on the field of view (FOV), the distance between lens and target, and the speed of the camera motion as described by Moghimi et al. (2017). The field of view (FOV) of the camera lens was 33 degrees, and the distance between the target and lens was about one meter. The speed of the camera was set to 0.025 m/s, thus the calculated frame rate was 27 frames per second to obtain square pixels (aspect ratio of 1:1). Gain and exposure time were adjusted appropriately based on light conditions to avoid over-exposure while taking advantage of the full dynamic range (12 bits).

2.2.3 Image preprocessing

2.2.3.1 Radiometric calibration

Raw images were radiometrically calibrated to account for non-uniform spatial and spectral responses of the sensor due to variability in gain and offset of each detector. Raw digital numbers (DNs) were converted to radiance ($Wm^{-2}sr^{-1}nm^{-1}$) using the radiometric calibration file provided by the camera manufacturer. Radiance was then converted to reflectance to normalize image data based on incoming solar irradiance so objects could be compared more objectively across images and across capture dates. A Spectralon panel (Labsphere, Inc., North Sutton, NH, USA) was placed in each image and was used as a reference to convert from radiance to reflectance. Spectralon reflects ~99.7% of incident light equally in all directions regardless of the illuminated light angle. Radiometric conversions were performed using Spectronon Pro software (Resonon, Inc., Bozeman, MT, USA).

2.2.3.2 Noisy band removal

Due to high noise, the first and last five bands were removed prior to any analysis. In addition, spectral bands from 753 to 766 nm and also from 813 to 827 nm were disregarded since they were noisy bands near the O_2 (~760 nm) and H_2O (820 nm) absorption regions. Following band removal, 215 of 240 bands were used for analysis. Subsequent analyses were performed using MATLAB R2017a (MathWorks, Inc., Natick, MA, USA).

2.2.4 Vegetation mask

Segmentation of the target of interest from background is a key step in image analysis. To segment vegetation pixels from background pixels, a binary mask was created by thresholding the normalized difference vegetation index (NDVI) (Rouse, Haas, Schell, & Deering, 1973) and excessive green index (EGI) (Moghimi, Aghkhani, Golzarian, Rohani, & Yang, 2015). The masks were then multiplied together element-wise to generate a primary mask for leaf segmentation (Figure 2.1 – step I). Pixels near leaf edges were likely to have spectral characteristics of mixed pixels, because they were located near the

vegetation/background boundary. To assure these mixed pixels would not pass the vegetation mask, a morphological operation (erosion with 3x3 matrix of ones as structuring element) was applied on the primary mask to check the connectivity of each pixel with its neighbors. Pixels from the primary binary mask that were connected with less than eight neighbors were excluded from the final mask (Figure 2.1 – step II). This final mask was then used to extract all vegetation pixels from the hyperspectral data cube. The masked hyperspectral data cube was converted to a 2D matrix *X* whose rows were features (i.e., wavelengths) and columns were samples (i.e., pixels) and subsequent analysis was performed on matrix *X*.

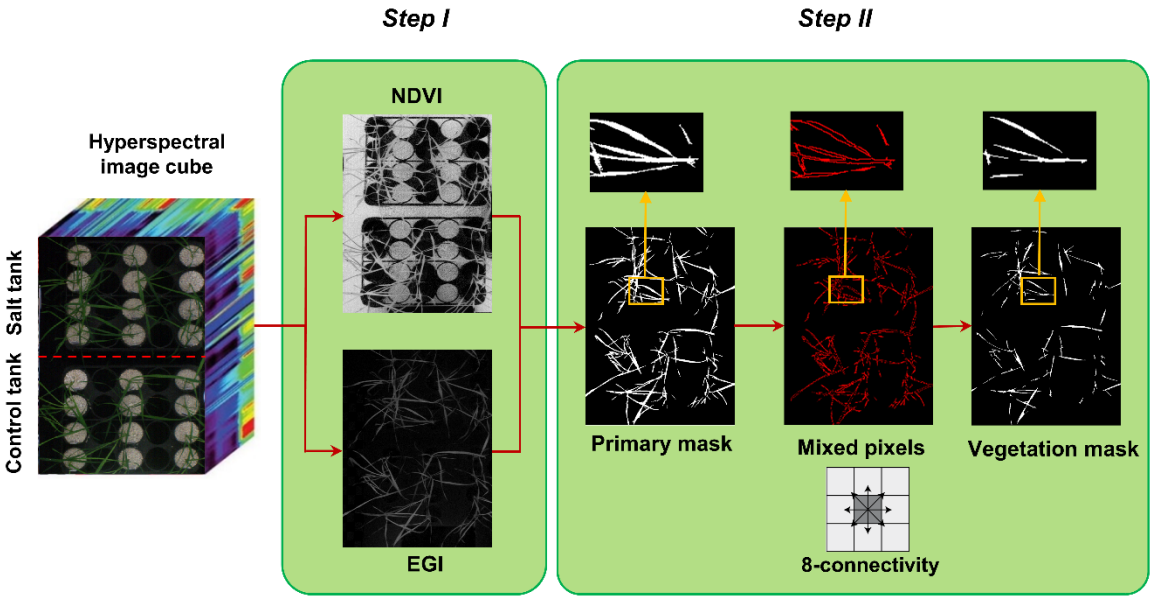

Figure 2.1. Two steps of creating vegetation binary mask, step I: segmentation of vegetation pixels from background using spectral indices (normalized difference vegetation index and excessive green index), and step II: filtering of mixed pixels at leaf edges using morphological operation (erosion with a 3 × 3 matrix of ones as structuring element).

2.2.5 Data analysis

2.2.5.1 Normalized reflectance difference

Matrix *X* was split into two matrices: matrix C $(d \times n_c)$ contained only control pixels and matrix S $(d \times n_s)$ contained only salt pixels where d denotes the number of bands, n_c and n_s represent the number of control and salt pixels from a single hyperspectral image,

respectively. The average reflectance for pixels of both treatments (control and salt) was calculated over all wavelengths for each wheat line to determine how the reflectance patterns varied due to salt stress. For this purpose, we proposed a method referred to as the normalized reflectance difference (NRD). For each wheat line, NRD was calculated as follows:

$$NRD_i = \frac{[\frac{1}{n_c}\ \sum_{k=1}^{n_c} C_{ik}\ ] - [\ \frac{1}{n_s}\ \sum_{k=1}^{n_s} S_{ik}\ ]}{[\frac{1}{n_c}\ \sum_{k=1}^{n_c} C_{ik}\ ]} \qquad i = 1,\ldots,\mathrm{d} \tag{2.1}$$

NRD_i represents the difference between the average reflectance of pixels representing control and salt treatments divided by the average reflectance of the control at band *i*.

2.2.5.2 Endmembers extraction

It should be noted that those vegetation pixels that could pass the segmentation steps might not be pure pixels because of limiting factors such as leaf angle, leaf curvature, and shadow. Therefore, to extract the spectral signatures for salt stressed and control plants, the most pure pixels for each class (salt and control class) should first be identified among the pixels passed from the segmentation steps. These pure pixels can be considered as endmembers of the two classes. Each hyperspectral image contained only a single wheat line, but included both salt and control treatments. Consequently, there were only two potential classes and subsequently two respective endmembers in each image. These endmembers are the most spectrally pure pixels in the hyperspectral image. The assumption of pure pixels existence can be correct because of the high spatial resolution we attained (~1 mm).

Based on the strategy proposed by Winter (1999), endmember pixels in a feature space are the vertices of a simplex that has the maximum volume compared to any other simplex formed by other pixels. To elaborate, consider each pixel as a point in a *d*-dimensional feature space where *d* is the number of bands. From prior assumption, there could be *n* endmembers which are pure pixels in the image. These *n* endmembers are the vertices of a (*n-1*)-simplex that has the maximum volume in a *d*-dimensional spectral feature space spanned by all pixels (i.e., this simplex contains the majority of pixels in the feature space).

Several algorithms and techniques for extracting endmembers based on this idea have been developed with the intention of improving computational time and accuracy (T. H. Chan, Ma, Ambikapathi, & Chi, 2011; Thurau, Kersting, Bauckhage, Iais, & Augustin, 2010; Winter, 1999). To find the unique set of two endmembers comprising the vertices of a 1-simplex in this study, successive volume maximization (SVMAX) was utilized. SVMAX has a modified objective function in which endmembers are identified recursively through a successive optimization problem (T. H. Chan et al., 2011). In each image, SVMAX identified two pixels that were the furthest from each other in the high-dimensional feature space, each representing one class: salt and control.

2.2.5.3 Measurement of pixels similarity to endmembers

Similarity measurement by solving a quadratic optimization problem

Once endmembers were identified, other pixels were represented as a convex combination of the endmembers. This is a factorization problem in which the matrix of data X is factorized as the product of two other matrices, W and H (Thurau, et al., 2010). Matrix W contains the endmembers extracted using the SVMAX algorithm, and matrix H is the matrix of coefficients. To find H, the Frobenius norm of $\|X - WH\|_F$ can be minimized with two constraints:

$$\begin{aligned} &min\ \|X - WH\|_F \\ &s.t. \begin{cases} 1^T.\ h_j = 1 \\ 0 \le h_{ij} \le 1 \end{cases} \end{aligned} \tag{2.2}$$

The jth column of X (x_j: 215×1) can be represented as matrix multiplication of W (215×2) and the vector of coefficients located at the jth column of H (h_j: 2×1), which can be interpreted as the abundance of corresponding endmembers in a particular pixel. These coefficients are most commonly used in spectral unmixing techniques to identify the abundance of each endmember. However, these endmember coefficients are interpreted as the '*similarity*' of a given pixel to the salt and control endmembers in the current study. Therefore, all coefficients should be non-negative and less than or equal to one. Furthermore, the summation of coefficients (each column of H) for each pixel needs to be

equal to one. Larger coefficients represent more spectral similarity between a given pixel and its corresponding endmember.

The matrix of coefficient H can be calculated by solving Eq. (2.3) as a quadratic optimization problem as follows.

$$\min_{h_j} \quad \frac{1}{2}\, h_j^T Q h_j + \, c^T h_j \,, \qquad j = 1, \dots, N$$
$$s.t. \begin{cases} 1^T .\, h_j = 1 \\ 0 \le h_{ij} \le 1 \end{cases} \tag{2.3}$$

where

$$Q = 2W^T W$$
$$c = \, -2W^T x_j \tag{2.4}$$

The algorithm (interior-point-convex) used to solve this optimization problem ran iteratively N times, where N was the number of pixels. Such an algorithm is not scalable for
high-throughput phenotyping on a large scale due to the fact that there are hundreds of images, each containing thousands of pixels, and the algorithm should iteratively run for every single pixel.

2.2.5.3.1 Vector-wise similarity measurement

In this study, a more efficient technique, referred to here as vector-wise similarity measurement (VSM), is proposed to obtain the matrix of coefficients based on the concept that the two endmembers representing the salt and control classes are the two most separated pixels in the feature space. Thus, other pixels can be projected to a line passing between these two endmembers. The distance between the projected point and each of the endmembers can be a measure of similarity, indicating how similar the spectrum of a given pixel is to salt and control endmembers. Figure 2.2 depicts the graphical illustration of the VSM method. For the purpose of visualization, Figure 2.2 illustrates the similarity of a

given pixel to either of the endmembers only in a two-dimensional feature space spanned by two bands. Note that this technique was implemented in the full dimensional feature space of hyperspectral images.

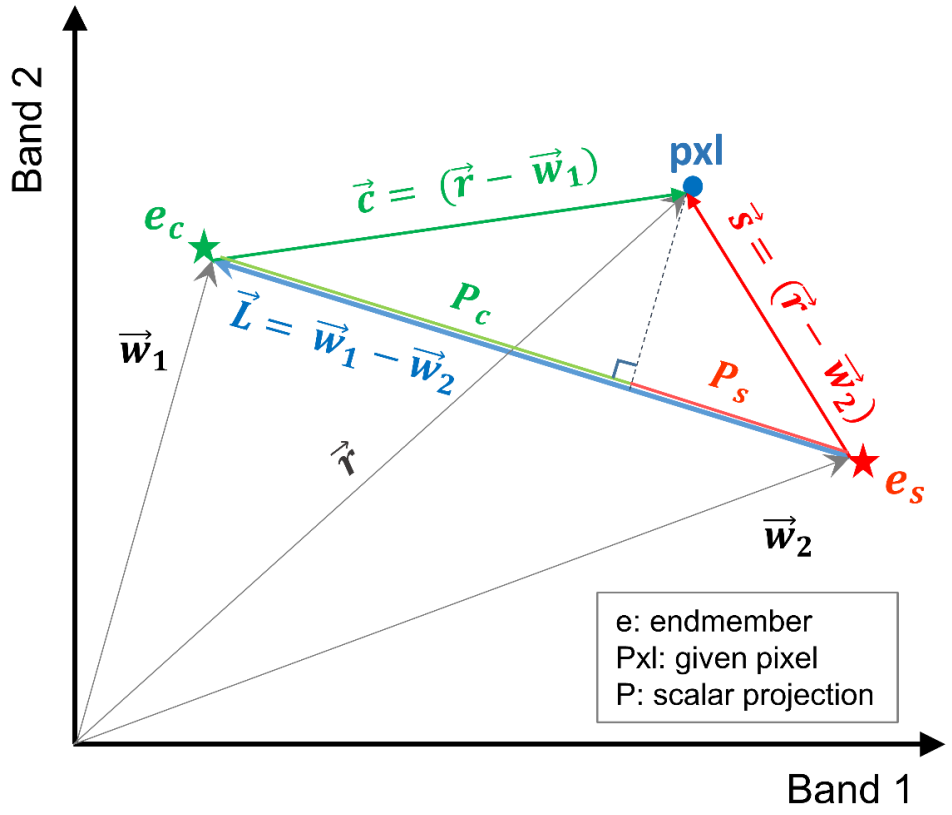

Figure 2.2. Graphical illustration of the vector-wise similarity measurement (VSM) method to determine the similarity of a given pixel to control and salt endmembers in a 2D feature space.

In Figure 2.2, $\vec{w}_1$ and $\vec{w}_2$ are the vectors of control and salt endmembers, respectively, and $\vec{r}$ is a vector representing a given pixel. The distance between a given pixel and the two endmembers is calculated as follows:

$$P_c = \frac{|\langle c, L \rangle|}{\|L\|} \tag{2.5}$$

$$P_s = \frac{|\langle s, L\rangle|}{\|L\|}$$

where $\vec{c} = (\vec{r} - \vec{w}_1)$, $\vec{s} = (\vec{r} - \vec{w}_2)$, $\vec{L} = (\vec{w}_1 - \vec{w}_2)$, P_c and P_s are the absolute values of the scalar projection of $\vec{c}$ and $\vec{s}$ on $\vec{L}$, respectively. In order to impose the constraints similar to Equation (2.2), P_c and P_s were normalized based on the distance between the two endmembers ($\vec{L}$).

$$D_c = \frac{P_c}{\|L\|} = \frac{|\langle c, L\rangle|}{\|L\|^2}$$
$$D_s = \frac{P_s}{\|L\|} = \frac{|\langle s, L\rangle|}{\|L\|^2} \quad (2.6)$$

where D_c and D_S are the normalized distance between the projected point of a given pixel to the control and salt endmember, respectively; their summation is equal to one ($D_c + D_s = 1$). Note that similarity *S* can be defined as *S = 1 – D* (i.e., similarity decreases as distance increases).

The processing time for VSM (~ 70 seconds) was approximately 3500 times faster than the quadratic optimization algorithm (~ 0.02 seconds) using the same dataset (~15000 pixels) and processing power. It was substantially faster because the similarity of all pixels can be calculated through matrix multiplication rather than going through an iterative loop. In addition, no constraint was needed for VSM, whereas two constraints had to be met in each optimization loop of the quadratic optimization approach.

Since the two distances calculated for pixels complement each other (i.e., they sum up to one), each pixel can be represented with a single value that represents the similarity of a pixel to either of the treatments. Representing pixels in this manner allows dimensionality to be reduced from 215-D to 1-D and more notably, allows for quantification of pixel similarity to either endmember. In this study, the similarity of individual pixels to the salt endmember was extracted as a row vector from the coefficient matrix obtained by VSM for further analyses. The equivalent row and column subscript values corresponding to a single index of similarity vector were determined to assign similarity values to the

corresponding non-zero pixels that could pass the mask in a given image. Figure 2.3 shows the result of mapping from similarity vector of the CS line to a gray scale image. In this image non-zero pixels represent the similarity of control and salt pixels to the CS salt endmember and pixels equal to zero represent the masked background. To illustrate the structure of similarity values in a proper manner, the gray scale image was transferred to a colormap image (Figure 2.3-b).

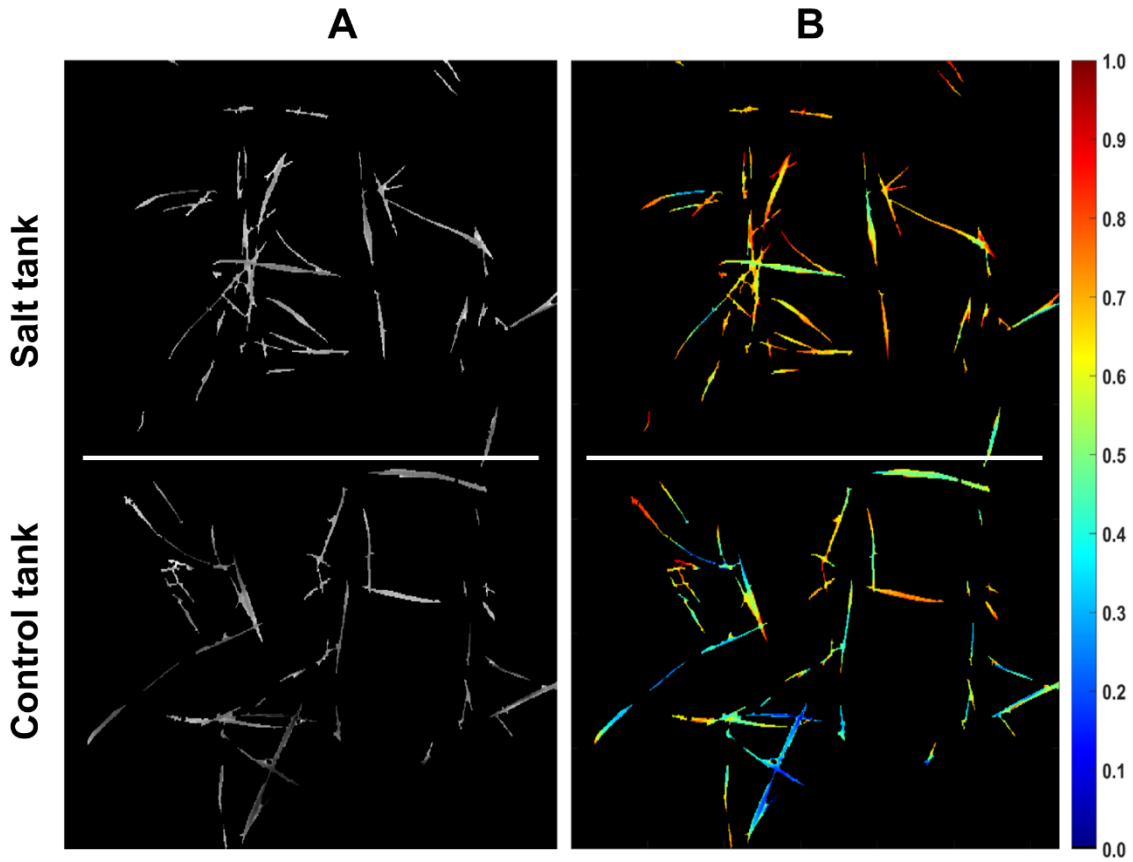

Figure 2.3. Similarity of pixels to the salt endmember for CS line. (A) Similarity of pixels to salt endmember represented as a gray scale image (bright colors denote more similarity). (B) Similarity of pixels to salt endmember represented as a colormap (larger values in colorbar denote more similarity).

2.2.5.4 Analysis of similarity vector obtained by VSM

Two methods were used to analyze the similarity vector/gray-scale image obtained by VSM for each line.

2.2.5.4.1 Histogram distance

The similarity vector obtained by VSM can be represented as a univariate histogram per each class to investigate the distribution of two classes for each line. To assess the tolerance of a wheat line to salt stress, histogram distance between the control and salt class was measured. The histogram distance indicates how much the imposed salt treatment caused a shift in the similarity distribution between the control and salt classes.

A metric distance measure called the minimum difference of pair assignments (MDPA) was used to measure the histograms distance (Cha & Srihari, 2002). MDPA was selected because the type of similarity histogram was ordinal in which order of bins matters. Moreover, similarity/difference between non-overlapping parts of control and salt histograms should be considered in measuring histograms distance. According to Cha & Srihari (2002), MDPA accounts for the similarities of the entire histograms, including overlapping and non-overlapping bins, while other methods such as Bayes error only consider the intersection of histograms for distance calculation. In addition, commonly used methods such as Euclidean distance and Bhattacharyya distance will remain unchanged if we permute the histogram bins. This *'shuffling invariant'* property is not suitable for calculating the similarity of two ordinal type histograms.

MDPA was calculated based on the equation proposed by Cha & Srihari (2002) as follows:

$$D(H(C), H(S)) = \sum_{i=0}^{b-1} \left| \sum_{j=0}^{i} (H_j(C) - H_j(S)) \right| \tag{2.7}$$

where $H(C)$ and $H(S)$ are the similarity histogram of control and salt classes to salt endmember, respectively; and b is the number of bins. In this study, the number of bins was set to 100 and therefore, the width of each bin was 0.01 because the similarity ranged from zero and one.

According to the MDPA equation, the distance between control and salt histograms is the minimum required sample replacements among bins such that the salt histogram becomes identical to the control histogram. This requires that the number of samples (i.e., pixels) be

identical for both control and salt class. However, it is rare that the number of samples/pixels belonging to the control and salt classes be equal after initial image processing. To account for the problems associated with an imbalanced dataset, the larger dataset was sub-sampled using a stratified random sampling method in which each bin was randomly sub-sampled based on the ratio between samples of the smaller and larger datasets. The stratified random sampling assured that the distribution of the sub-sampled dataset remained unchanged, which was required for MDPA to achieve proper results. In addition, MDPA, obtained from Equation (2.7) for each wheat line, was divided by the number of pixels belonging to one of its histograms (either salt or control histogram since they have equal number of pixels after subsampling) to account for the difference between the numbers of pixels among the wheat lines.

2.2.5.4.2 Bayesian inference

After sub-sampling, the gray-scale image obtained by VSM per each wheat line had equal number of salt and control pixels, each representing the similarity to the salt endmember. To make inferences about the tolerance of each wheat line using these gray-scale images, the posterior probability of salt class was calculated using Bayesian inference. The framework of the Bayesian inference is Bayes' rule, which was written in this study as:

$$P(salt|x) = \frac{P(salt) \times P(x|salt)}{P(x)} = \frac{P(salt) \times P(x|salt)}{P(salt) \times P(x|salt) + P(control) \times P(x|control)} \quad (2.8)$$

where $P(salt|x)$ is the posterior probability of salt class given an observation x. $P(x|salt)$ is the class-conditional probability which is the conditional probability of observation x given salt class, and $P(x)$ is the evidence representing the occurrence probability of observation x. $P(salt)$ is the prior probability, and it was calculated as the ratio of the number of salt pixels to the total number of vegetation pixels in a single image. The posterior would have been biased toward the larger class if the dataset were imbalanced, and this is why stratified random sampling was performed in an earlier step. After sub-sampling, each class had an equal number of pixels, and therefore, an equal prior ($P(salt) = P(control) = 0.5$). As a result, Eq. (2.8) can be simplified by canceling out the equal prior

of salt and control classes from the numerator and denominator.

In this study, the similarity of pixels to the salt endmember, denoted as *s,* was considered as the observation *x*. To compute the class-conditional probability for salt and control classes, the similarity values of each class, which were continuous, were discretized into two ordinal bins by specifying 0.5 as the split point. Pixels with more similarity to the salt endmember ($s > 0.5$) were categorized in one bin and pixels with more similarity to the control endmember ($s \leq 0.5$) were categorized in the other bin. Afterward, the posterior probability of salt class for pixels with more similarity to the salt endmember (i.e., $P(salt|s > 0.5)$) was calculated for each of the wheat lines as follows:

$$P(salt|s > 0.5) = \frac{P(s > 0.5|salt)}{P(s > 0.5|salt) + P(s > 0.5|control)} \tag{2.9}$$

Eq. (2.8) was derived from Eq. (2.8) by canceling out the equal prior of salt and control classes and considering $s > 0.5$ as the observation. The numerator of Eq. (2.8) is the fraction of salt pixels with more similarity to the salt endmember to the total number salt pixels, and the denominator represents the fraction of all vegetation pixels with $s > 0.5$ to the total number of vegetation pixels within a single image.

For a susceptible line, the spectral response of pixels representing the salt stressed plants become more district from that of the control plants because of physiological and metabolic alteration in stressed plants. As a result, the salt pixels shift more toward the salt endmember and become more distinct from the control endmember; and hence, the number of similar pixels to the salt endmember among salt treatment pixels ($P(s > 0.5|salt)$) is much larger than the number of pixels similar to the salt endmember among control pixels ($P(s > 0.5|control)$). Therefore, it can be inferred that a susceptible line should have a larger posterior probability of salt class ($P(salt|s > 0.5)$) than a tolerant line in which a given pixel with more similarity to the salt endmember can be assigned to the salt class with a lower confidence as its posterior probability of salt class is relatively lower.

2.3 Results

2.3.1 Conventional biomass measurement to assess salt tolerance

The primary objective of this research was to quantitatively rank wheat lines based on salt tolerance using HSI. As a case study, four *Triticum aestivum* bread wheat lines (see Methods for full description) were selected for assessment of salt tolerance with destructive biomass measurements in parallel with HSI.

Many different molecular, physiological, and growth parameters can be used to assay salt tolerance differences between genotypes, including Na+ uptake, the ratio of K/Na+, photosynthetic activity (Nxele, Klein, & Ndimba, 2017), gene expression (Agarwal et al., 2013), and aerial and root biomass in salt versus control conditions over extended growing periods (Munns & James, 2003). To assess salt tolerance, salt treatments were applied for two weeks, and then root and aerial biomass were measured on a dry weight basis.

According to previous studies, Kharchia is a salt tolerant line since it maintains a stable harvest index and yields well in high salt conditions (Munns et al., 2006; Schachtman et al., 1992), while CS is a salt-sensitive cultivar (Zhang et al., 2016). Therefore, the main objective of performing conventional phenotyping was to identify the tolerance of the two unknown additional alloplasmic lines, co(CS) and sp(CS). The results of biomass measurements for these two lines were compared with CS since they contain the exact same nuclear background as CS, which allowed for a direct comparison of biomass to CS.

The biomass measurements revealed that both CS and sp(CS), unlike co(CS), showed a reduction in both aerial and root biomass after growth in the presence of 200 mM NaCl (Figure 2.4). The analysis of variance found significant interactions at all levels for aerial and root biomass, including between salt level and genotype (Supplemental Table 2.1). A closer examination of effect sizes using linear mixed modeling showed significantly less change in aerial and root biomass from 0 to 200 mM in co(CS) when compared to CS (Supplemental Table 2.2), indicating that the alloplasmic line co(CS) is more salt tolerant than the nuclear donor CS in terms of salt effect on biomass. However, overall growth rate

may be impacted in co(CS), as biomass in the absence of salt is less than that of CS. A possible explanation for this observation is that altered nuclear-cytoplasmic communication in this line could lead to improper expression of organellar (or nuclear) transcripts involved in stress tolerance, therefore "priming" the alloplasmic for stress and reducing sensitivity to salt stress (as measured by the difference in biomass between 0 and 200 mM).

The response of the other alloplasmic line (i.e., sp(CS)) was not significantly different when compared to CS (Supplemental Table 2.2); however, it trended towards less change in response to salt compared to CS for aerial biomass. The change in root biomass was almost identical to that of CS. Based on these observations, it can be inferred that sp(CS) is slightly more tolerant than CS.

Although it is historically known that Kharchia is more tolerant than CS, the result of biomass measurements of Kharchia was also compared with CS to examine if the conventional biomass measurement could capture the difference between these two lines with dissimilar genome backgrounds. Intriguingly, the magnitude of aerial biomass change between the control and treatment in the highly salt-tolerant Kharchia cultivar was not significantly different when compared with salt-sensitive CS (Figure 2.4 and Supplemental Table 2.2). However, similar to sp(CS), the trend was also towards a smaller change in response to salt than CS. This indicates that biomass measurement, although a convenient parameter to measure in a lab environment, may not always reflect the actual field performance in desirable traits such as harvest index or yield. This is consistent with previous results that showed a substantial biomass decrease for Kharchia in the presence of salt, yet also a high relative yield and harvest index (Schachtman et al., 1992). Without the substantial historical knowledge of how Kharchia was derived from Indian landraces adapted to sodic soils (Munns et al., 2006), the assessment of salt tolerance with hydroponic screening and biomass measurement for this study may have missed this highly valuable source of germplasm.

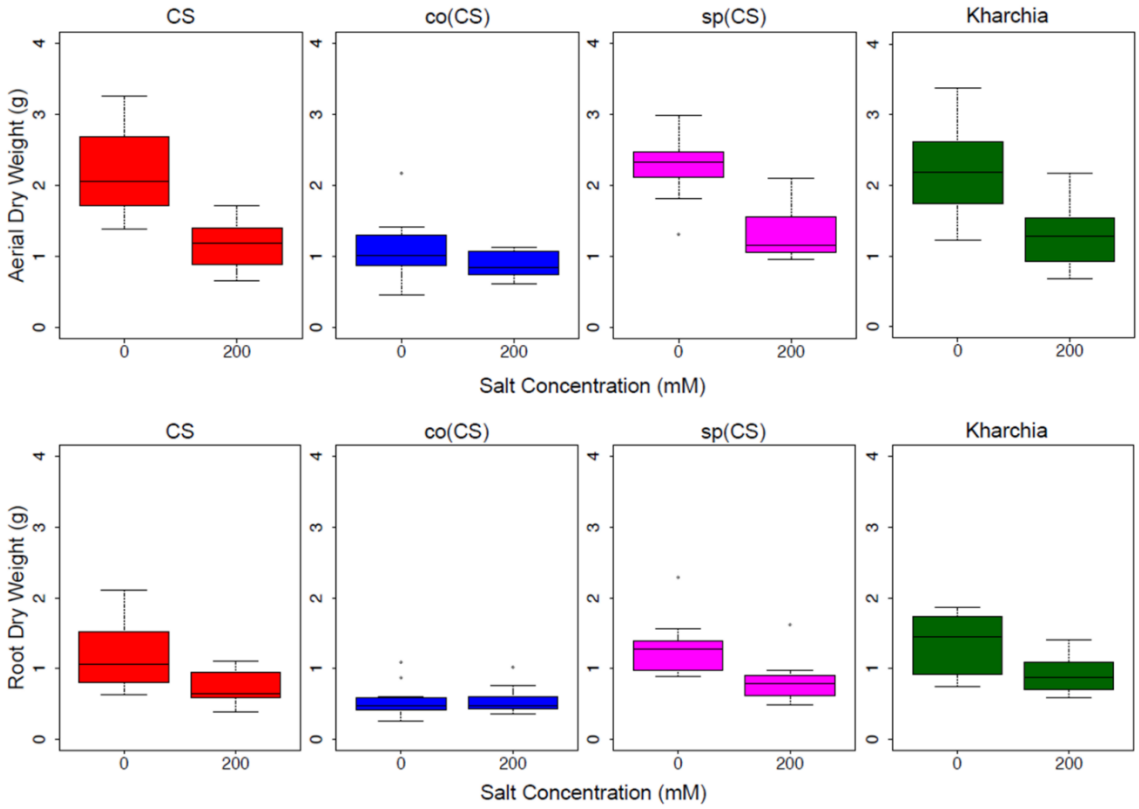

Figure 2.4. Aerial and root biomass in control (left bars) and salt (right bars) conditions. The upper panel shows aerial biomass and the lower panel shows root biomass for Chinese Spring (CS), the two alloplasmic lines Ae. columnaris(CS) (co(CS)) and Ae. speltoides(CS) (sp(CS)), and Kharchia.

Based on the results of our conventional salt tolerance and historical knowledge, we can conclude that Kharchia and co(CS) are more salt-tolerant than sp(CS) and CS. In addition, the time-consuming and laborious process of conducing the conventional biomass measurement for salt tolerance assessment underscored the need for more informative and quantitatively precise screening techniques to rapidly and non-destructively assess salt tolerance, particularly when comparing cultivars with drastically different genetic backgrounds and growth regimes, such as Kharchia and CS that differ in vernalization requirements and photoperiod sensitivity (Koebner, Martin, Orford, Miller, & Lupton, 1996). However, since co(CS) and sp(CS) have identical nuclear backgrounds to CS but only differ in their organellar genomes, direct comparisons of biomass are more valid.

2.3.2 Normalized reflectance difference

To gain an overall view of how the four wheat lines differed in response to salt-stress, normalized reflectance difference (NRD) was examined. NRD in Eq. (2.1) indicates how much the reflectance of the salt class changed compared to the control class in response to the imposed stress (Figure 2.5). According to the NRD, Kharchia is the most tolerant among the four lines, which is consistent with previous studies and historical knowledge (Munns et al., 2006; Schachtman et al., 1992) but inconsistent with assessment of salt tolerance with biomass measurement. On the other hand, CS appears to be the most susceptible due to the larger difference between average reflectance of control and salt pixels. This observation was consistent with both previous studies (Zhang et al., 2016) and biomass measurement in this study. Based on NRD, the two alloplasmic lines, co(CS) and sp(CS), are somewhere in between Kharchia and CS. The NRD method alone provides a rapid and brief insight to identify the tolerant lines, but not in a quantitative manner. A quantitative ranking can be generated by calculating the area under the NRD curve as the summation area of all trapezoids formed between two successive bands and the NRD curve. To give more physical meaning to the area under NRD, wavelength (nm) of the x-axis was converted to energy (J) using the Planck equation. Consequently, area under the curve (AUC) can be presented in the unit of energy. The results of AUC for the wheat lines are shown in Figure 2.6. The order of lines in terms of their tolerance to salt based on AUC is similar to the results achieved by conventional biomass assessment, except that we could detect the strong salt tolerance of Kharchia more readily with AUC. It is also notable that both alloplasmic lines show smaller AUC than the CS nuclear donor.

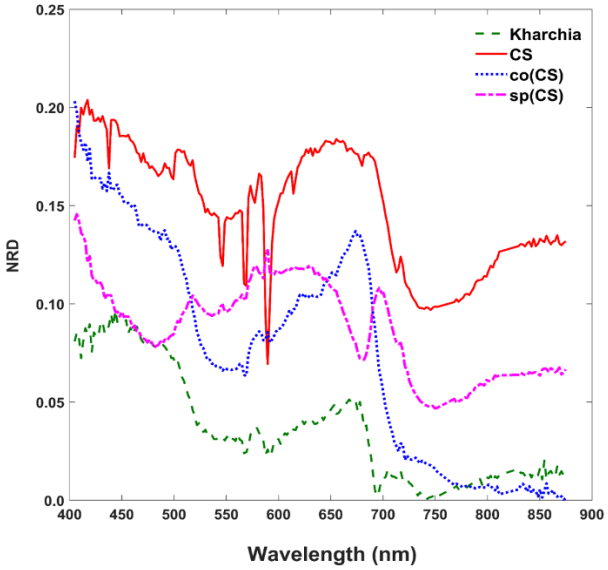

Figure 2.5. Normalized reflectance difference (NRD) of wheat lines, indicating how much the averaged reflectance of the salt pixels changed compared to the control pixels in response to the imposed salt stress on each wheat line.

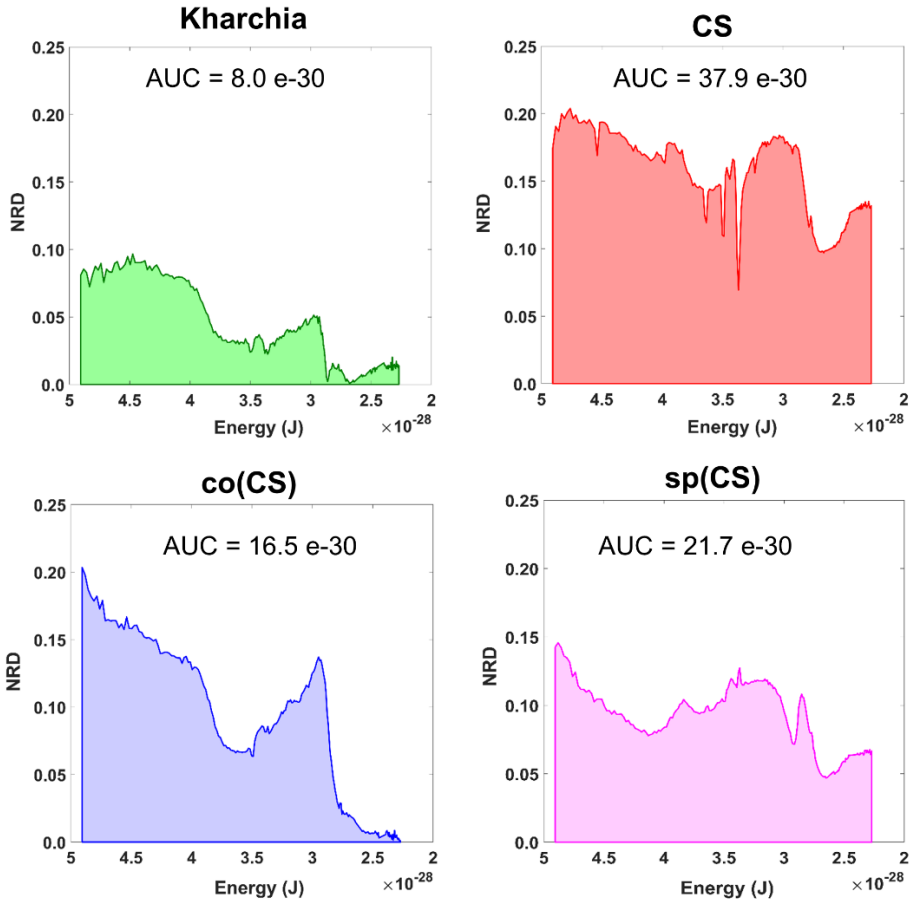

Figure 2.6. Area under the normalized reflectance difference curve calculated as the summation area of all trapezoids formed between two successive bands and the NRD curve (wavelength (nm) on x-axis was converted to energy (J) using the Planck equation).

2.3.3 Endmembers extraction using SVMAX

Figure 2.7 presents the salt and control endmember locations of CS with respect to other pixels where all of the pixels extracted to find endmembers have been projected onto the first three components of principal components analysis for visualization purposes. It is evident that the endmembers identified by SVMAX are the pixels located at the extremes in the space spanned by the first three principal components.

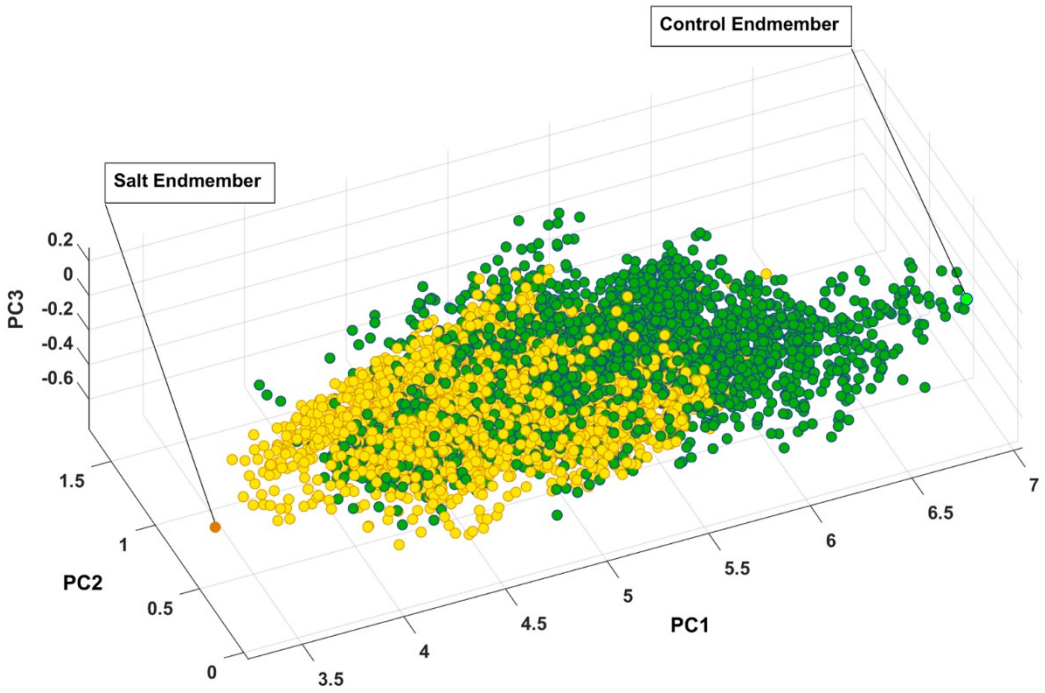

Figure 2.7. Control and salt endmembers identified by SVMAX for CS line as the furthest away pixels form each other illustrated in a space spanned by the first three principal components (i.e., PC1, PC2, and PC3). Green and yellow circles represent control and salt pixels, respectively.

For this study, the endmembers are assumed to be the pixels that represent either the control or salt treatments. Afterwards, the reflectance of other pixels can be represented as the similarity to either of these endmembers while reducing the dimensionality from 215 to one.

Figure 2.7 illustrates how pixels belonging to each of the two classes are intertwined (i.e., many occupy similar feature space) while a slight shift toward the lower values of PC1 can

be observed. One explanation for this intertwinement is that reflectance of a plant is not only a function of health status but it also depends on other parameters such as leaf angle, geometry of leaf, and configuration of sensor, leaf surface, and light source. As a result, we focused on the distribution of similarity to salt endmember for control and salt pixels rather than a pixel-by-pixel analysis.

2.3.4 Histogram distance measure

The distributions of pixel similarity to the salt endmember for control or salt treatments are illustrated via histogram (Figure 2.8). CS had the largest mean shift of salt pixels toward the salt endmember from $m_c = 0.50$ to $m_s = 0.66$ indicating that it is the least tolerant line. It is also notable that the standard deviation decreased with salt induction causing the pixels to aggregate towards the salt endmember. In the similarity distribution of the salt stress class of CS, there were no pixels similar to the control endmember.

For Kharchia, the mean of the salt pixels' similarity to the salt endmember remained approximately unchanged compared to the control pixels ($m_c \cong m_s$), and the standard deviation changed slightly (Figure 2.8). This suggests that the distribution of control and salt pixels between the two endmembers was similar for both conditions.

The mean of salt pixels' similarity to the salt endmember for co(CS) and sp(CS) shifted to the larger value indicating that salt stress caused the spectral response of pixels to change. However, this shift was lower compared to CS line in which salt stress made a large difference between the control and salt distribution.

Except for Kharchia (the most tolerant wheat line), salt stress caused the similarity distribution of pixels belonging to the salt class to shift toward the salt endmember and caused the standard deviation to decrease.

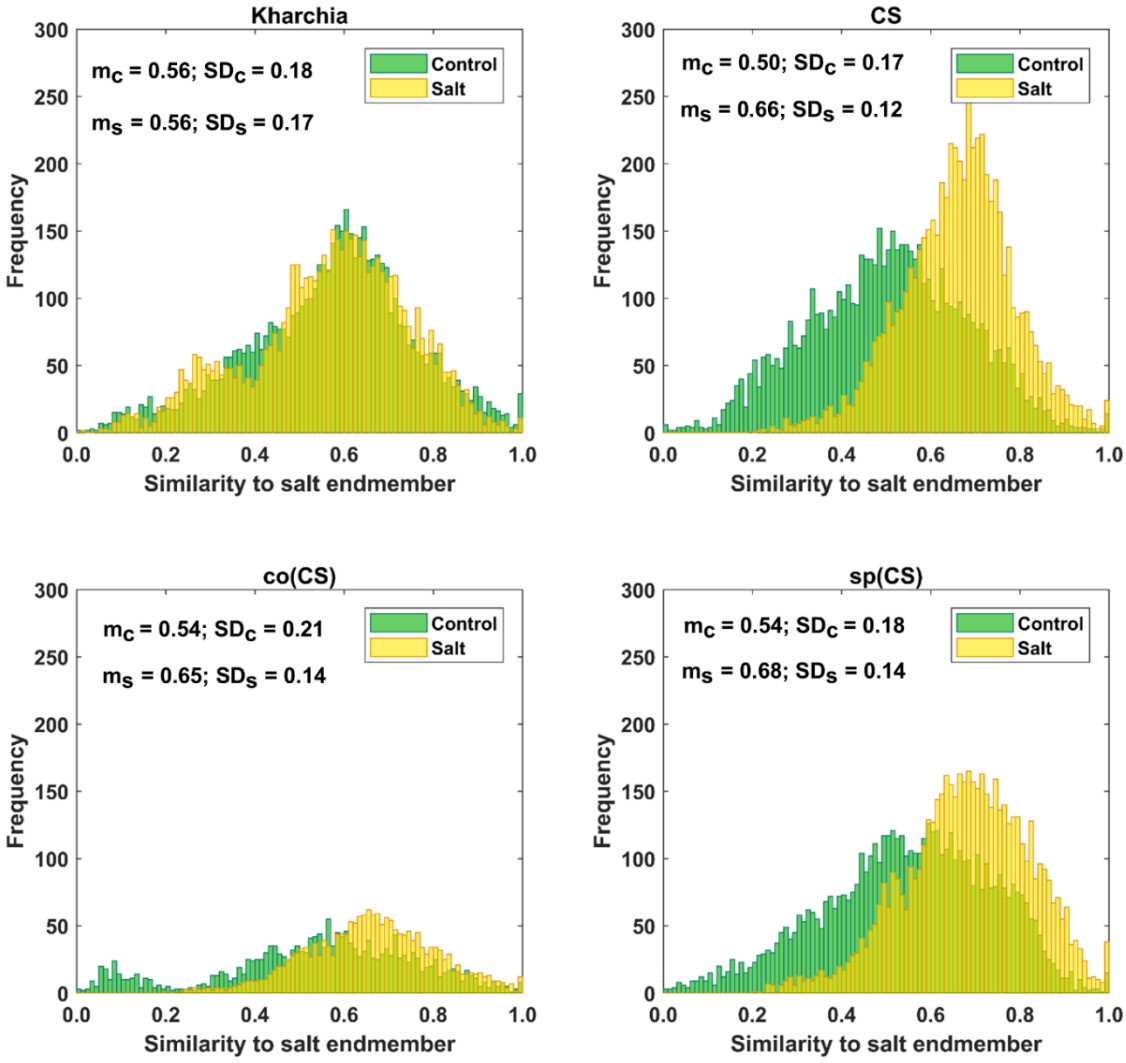

Figure 2.8. Histogram of similarity to salt endmember for control and salt pixels of the four wheat lines. Kharchia, as a tolerant line, and CS, as a susceptible line, have the smallest and largest mean shift of salt pixels (mc) toward the salt endmember, respectively.

To quantify the histogram shift due to the imposed salt stress, MDPA was calculated based on Eq. (2.7). Based on MDPA, CS has the largest distance between its salt and control classes, whereas Kharchia has the lowest distance (Figure 2.9). Except for Kharchia, that has been discussed ad nauseam, the results of MDPA for ranking wheat lines are similar to the results of conventional biomass measurement.

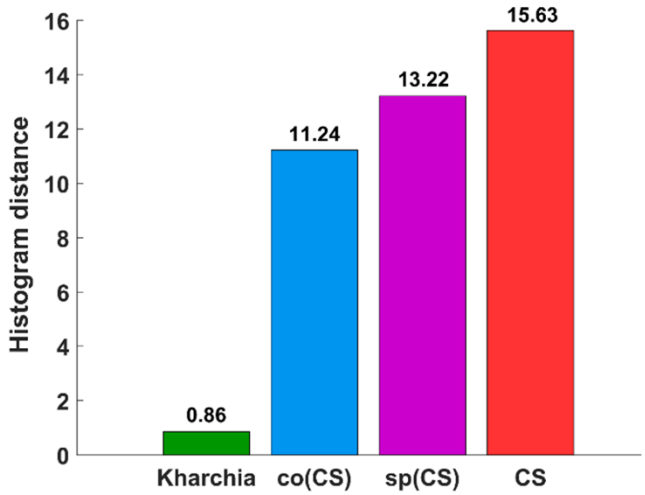

Figure 2.9. Results of the minimum difference of pair assignments (MDPA) for calculating histogram distance between distribution of similarity to salt endmember for control and salt pixels.

2.3.5 Posterior probability

To provide an additional method to quantify salt tolerance of lines, we also calculated the posterior probability of salt class for each line if a given pixel is more similar to the salt endmember (i.e., $P(salt|s > 0.5)$. Table 2.1 presents the posterior probability, prior, class-conditional probability, and evidence for each wheat line. As expected, CS had the largest posterior probability. This indicated that if a pixel of CS was more similar to the salt endmember, then this pixel could be classified as salt class with a higher confidence in comparison with other lines. The relatively large posterior probability of salt class for CS also implied that the spectral response of salt treated plants became more distinct than control plants even after a short time (only one day after salt application) such that the salt pixel were more similar to the salt endmember and more distinct from the control pixels. This can be attributed to the susceptibility of CS to salt stress. However, no such inference could be made for Kharchia because pixels with more similarity to the salt endmember could not be classified with high confidence to salt class. In addition, the subtle difference of posterior and prior probabilities of Kharchia implied that the imposed salt stress did not have a significant impact on the spectral response, and hence, the salt and control pixels of Kharchia were alike in terms of similarity to the salt endmember.

The quantified ranking of wheat lines based on Bayes' rule is consistent with the MDPA ranking, and similar to the results of biomass measurement. Again, we could readily detect the robust salt tolerance of Kharchia compared to CS, which was missed with conventional phenotyping. We could also quantify differences between the nuclear donor CS and the two alloplasmic lines, co(CS) and sp(CS). The result of posterior probability indicates that HSI could be useful to measure subtle changes which are less noticeable with conventional phenotyping.

Table 2.1. Bayes' rule components (prior, class-conditional probability, evidence, and posterior) for wheat lines

	P(salt)	P(s>0.5|salt)	P(s>0.5)	P(salt|s>0.5)
Kharchia	50.00%	68.61%	68.32%	50.21%
CS	50.00%	89.77%	70.88%	63.33%
co(CS)	50.00%	86.21%	74.84%	57.60%
sp(CS)	50.00%	89.32%	75.70%	59.00%

The similarity values were also discretized into 10 ordinal bins to observe how the posterior probability of both salt and control class vary over the variation range of similarity to the salt endmember (i.e., $posterior_i = \{P(class_i|s_j)\ s.t.\ s_j \in bin_j, j = 1, \dots, 10\}, \quad and\ i \in \{control, salt\}$) (Figure 2.10). The distribution of posterior probability of the two classes for Kharchia, the most tolerant line, represented a high degree of uncertainty in making inference because posterior probabilities of the two classes fluctuated throughout the similarity bins. This fluctuation can be associated with the absence of significant difference in spectral response of control and salt pixels due to the high degree of tolerance to salt stress in Kharchia. For the other wheat lines, the posteriors of two classes intersect at one point, which can be considered as the threshold of decision for making inferences. For co(CS), the second most tolerant line, posteriors fluctuated after the threshold point, such that the posterior of salt class is highly likely indicating the moderate tolerance of co(CS) line to salt stress. However, for CS and sp(CS), the two most susceptible lines, posterior probability of control and salt classes diverged after the decision point. In general, $P(salt|s_j)$ of these two susceptible lines monotonically increased as similarity to the salt

endmember increased. The trend of posteriors for these two wheat lines indicated how salt pixels became more distinct than control pixels in response to the imposed salt stress.

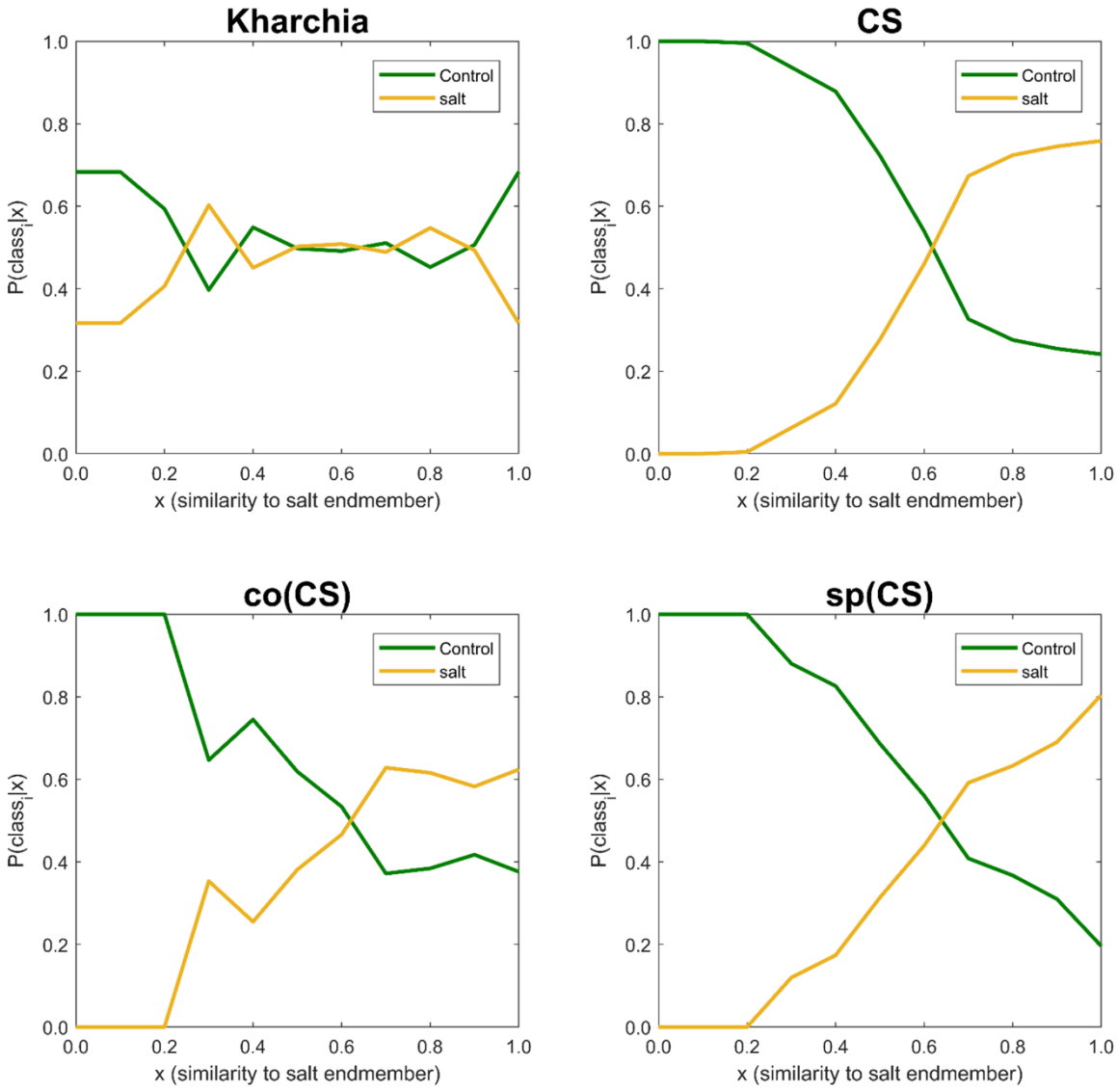

Figure 2.10. Posterior probability of belonging to control and salt classes given similarity to salt endmember. Posterior probability was calculated for each individual 10 bins of similarity to salt endmember.

To illustrate the result of posterior probability more intuitively, $P(salt|s_j)$ for all pixels in each wheat line image is shown as a colormap in which a distinct color was used to highlight the difference between posterior probability of pixels in control and salt regions

(Figure 2.11). According to Figure 2.11, the control and salt pixels of CS, the most susceptible line, can be visually distinguished, whereas the majority of Kharchia pixels, regardless of belonging to either class, have a similar posterior probability.

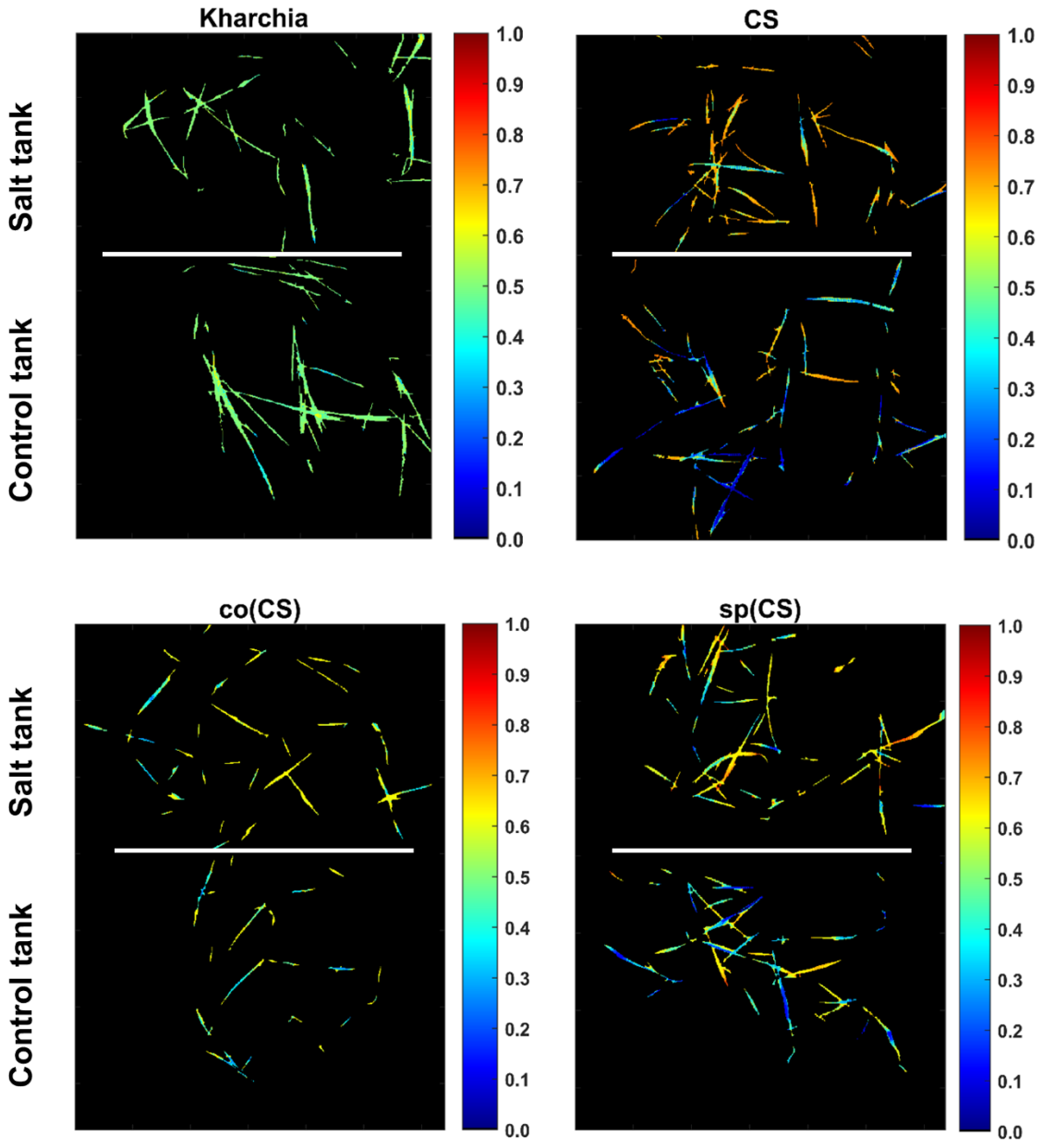

Figure 2.11. Posterior probability of belonging to salt class given similarity to salt endmember for control and salt pixels of each wheat line [i.e., P(salt|s)].

2.4 Discussion

Three methods including area under the NRD curve, MDPA, and posterior probability were utilized in this study to analyze hyperspectral images of wheat lines. In all methods, salt stress treatment of each line was compared to its control treatment because differences in spectral responses of a given line may not necessarily be related to the imposed stress, but rather to differences in inherent characteristics such as having waxy and/or darker leaves. The order of ranking of the examined wheat lines was similar for all of these methods. Kharchia was ranked as the most tolerant line followed by co(CS) and sp(CS). CS was identified as the most susceptible line by all three methods. In addition to ranking the wheat lines, more inferences could be made from calculating the posterior probability compared to the other two methods. For instance, it could provide the ability to observe the variations of posterior probability over all similarity bins (Figure 2.10). This observation can be used to define a threshold of making a decision for classification purposes if the classification of salt and healthy plants is of interest.

In this section, results and achievements of this research study are discussed.

2.4.1 Quantitative ranking

Our findings revealed that conventional phenotyping methods to identify salt tolerant wheat lines could be replaced by the fast and non-invasive methodology proposed in this study. It was surprising to find that the conventional assessment of salt tolerance with biomass measurement was not consistent with the anecdotal and historical evidence of salt tolerance for the Indian landrace Kharchia. However, biomass measurements of Kharchia were indeed consistent with previous studies that documented a significant biomass decrease, yet stable harvest index and yield in response to salt stress (Schachtman et al., 1992). Kharchia does show a significantly reduced level of Na+ uptake in the 5^{th} leaf when compared to salt-sensitive cultivars, which likely explains the stable harvest index of this line (Schachtman et al., 1992). Based on these contradictory findings, it appears that the most informative way to conventionally assess salt tolerance is to measure multiple growth and physiological parameters simultaneously. However, this is laborious, expensive, and

not feasible for many lines or populations simultaneously as is required for breeding programs. In contrast, our method could be used to quantitatively and objectively rank salt tolerance of individual wheat lines in a non-destructive and cost-effective manner. Moreover, the proposed method could be successfully used to detect subtle differences between lines, such as between alloplasmics and euplasmics.

2.4.1.1 Early detection

Detection of tolerant wheat lines was achieved as early as one day after the salt treatment when no visual symptoms were observed, and physiological and growth measurements were not yet possible. Early detection enables faster screening cycles and reduces the energy and costs needed to maintain plants in a controlled environment. The findings of this experiment provide evidence that breeders and plant geneticists would be able to properly manage time, energy, cost, and space in greenhouses while maintain accuracy and improve precision by implementing HSI and the proposed analytical methods. Faster assessment of stress tolerance is a major advantage to breeding programs and basic research alike.

2.4.1.2 Analysis of complex and high dimensional hyperspectral images

Appropriate preprocessing tasks and machine learning techniques were utilized to leverage the full potential of HSI in a phenotyping context. The achievements of this study in the context of hyperspectral image analysis are discussed in the following sub-sections.

2.4.1.2.1 Elimination of mixed pixels

A mixed pixel contains spectral information derived from a combination of objects that inherently have varying spectral characteristics. Mixed pixels are often located along the edges of objects in images, especially when the spatial resolution of an image is finer than the individual objects contained in the image (e.g., a leaf). In this study, mixed pixels along the edges of leaves were eliminated from the binary vegetation mask using morphological image processing. Without masking, these mixed pixels would have led to reduced accuracy through the endmember selection and subsequently in computing the similarity of pixels to the salt endmember on which further analyses were based. Prior to using morphological operations, filtering based on vegetation indices allowed us to appropriately

mask pixels that were covered by shadows from other leaves or objects.

2.4.1.2.2 Interaction between incident light and leaf surface

The interaction between incident light and the leaf surface is complex due to leaf angle and curvature. The intensity of the reflected light from leaves is both a function of physiology and leaf geometry, and tends to cause a significant change in the spectral response of vegetation. Behmann et al. (2015) reported 60% of the spectral response of sugar beets could be influenced by the geometry of plant leaves. This provides explanation for observing several control and salt pixels having similar spectral properties (pixels were not perfectly distributed in two well-separated classes). Although we did not explicitly address the issue of leaf geometry, we could mitigate its negative effects on the results by considering the similarity distribution of pixels to the salt endmember. By deploying a fuzzy concept (i.e., where each pixel can be partially similar to either endmember) rather than a hard classification (i.e., binary classification where each pixel is a member of only a single class), downstream processing techniques were possible, such as posterior probability and histogram distance.

2.4.1.2.3 Complex interpretation of hyperspectral images

In this research, the curse of dimensionality in analysis of hyperspectral images was addressed by projecting from a high dimension (215-D) feature space onto one dimension. The one-dimensional feature space, which can be represented as a gray-scale image, denoted the similarity of each pixel to the salt endmember while maintaining the required information for making further inferences. As a result, interpretation of hyperspectral images could be performed in a much more efficient manner with more meaningful information. By reducing dimensionality, results are simpler and are easier to comprehend. Because agricultural research groups are oftentimes comprised of many scientific disciplines, this directness is imperative. In general, the proposed strategy helps to interpret complex hyperspectral images by discovering the underlying hidden features caused by salt stress. This strategy will also provide more flexibility in selecting ML methods to analyze images. For instance, MDPA was utilized to measure the distance between the similarity distributions of control and salt pixels to the salt endmember. Another benefit of

conversion to a one-dimensional feature space is that visualization of relevant spectral information is possible in a more meaningful manner while involving all bands in similarity measurement and maintaining spatial integrity.

The other important advantage is that the required computational time and data storage space for analyzing and storing hyperspectral data was significantly reduced due to converting from a high-dimensional to one-dimensional space. For instance, the average size of a hyperspectral image (2000×640×215) was approximately 400MB while the size of obtained gray scale image (2000×640×1) was less than 1MB. It is also noteworthy that the VSM method used for computing the coefficient matrix proposed herein significantly reduced the processing time compared to the quadratic optimization problem that is commonly used.

2.5 Conclusion

In this study, we investigated the application of hyperspectral imaging in assessment of salt stress in four wheat lines one day after salt treatment. A novel pipeline was proposed for hyperspectral image analysis to leverage the full potential of HSI in a salt stress phenotyping context. The proposed pipeline can be applied for other plant phenotyping traits. The results of this study demonstrated the feasibility of quantitative ranking of wheat lines based on their salt tolerance by integrating HSI and novel analytical approaches. Quantitative and objective ranking methods are much needed, and provide invaluable information that can accelerate genomics research, such as GWAS or QTL mapping, used in breeding programs. The quantitative ranking of salt stress tolerance helps breeders integrate salt tolerance results with other desired traits (e.g., grain yield), which ultimately accelerates the development of new plant varieties. In future work, we will focus on identifying sensitive wavelengths to aid in development of a multispectral camera for salt stress phenotyping, and consequently reducing the cost of the sensor as well as the complexity of data collection associated with a hyperspectral line scanner.

2.6 Author Contributions

AM conducted the hyperspectral image collection, preprocessing of images, and developed the pipeline for analysis of images. MEM performed hydroponic screening assays and biomass measurements, and SK helped oversee the experimental design of screening. CY and PM provided technical assistance in analysis of images. AM wrote the paper and MEM co-wrote the biology parts of the paper. CY, PM, and SK edited the paper.

2.7 Funding

This research was financially supported by the United States Department of Agriculture-Agricultural Research Service, the National Science Foundation (IOS 1025881 and IOS 1361554), and Minnesota Agricultural Experiment Station.

2.8 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.9 Acknowledgments

The authors gratefully acknowledge the funding from the United States Department of Agriculture-Agricultural Research Service, the National Science Foundation (IOS 1025881 and IOS 1361554), and Minnesota Agricultural Experiment Station. Financial support from the MnDRIVE Initiative provided as graduate student funding is also much appreciated. We thank Dr. Shuhei Nasuda from Kyoto University and the National BioResource Project of Japan for providing the seeds of wheat alloplasmic lines. We also would like to thank Mr. Alireza Mahdavian, Mr. Chuanqi Xie, and Mrs. Parisa Kafash for their help in setting up the imaging system and collecting data, and special thanks to Mr. Tyler Nigon for his valuable comments in the manuscript revision.

2.10 Data Availability Statement

The datasets generated and analyzed during the current study are available in the Data Repository for University of Minnesota at https://doi.org/10.13020/D69Q3K.

2.11 Supplementary Material

Supplemental Table 2.1. Analysis of Deviance for Conventional Biomass Measurements

Analysis of Deviance Table for Aerial Biomass (Type II tests)

	Chisq	**Df**	**Pr(>Chisq)**
Salt Level	27.31	1	1.733e-07 ***
Genotype	48.511	3	1.657e-10 ***
Salt Level:Genotype	12.635	3	0.005497 **

Analysis of Deviance Table for Root Biomass (Type II tests)

	Chisq	**Df**	**Pr(>Chisq)**
Salt Level	23.9625	1	9.823e-07 ***
Genotype	41.9167	2	4.179e-09 ***
Salt Level:Genotype	7.6443	2	0.05396 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

Supplemental Table 2.2. Linear Mixed Effects Modeling for Conventional Biomass Measurements

Linear Mixed Effects Modeling of Aerial Biomass*					
Akaike Information Criterion (AIC)	188.6683				
Bayesian Information Criterion (BIC)	213.0949				
Log-Likelihood	-84.33417				
	Value	Standard Error	DF	t-value	p-value
Intercept[1]	2.205	0.15205739	81	14.501104	0
salt_level200[2]	-0.0052375	0.00107521	4	-4.87115	**0.0082**
co(CS)[3]	-1.1057504	0.18508003	81	-5.974445	**0**
sp(CS)[3]	0.0833572	0.18508003	81	0.450385	0.6536
Kharchia[3]	-0.0408333	0.18089641	81	-0.225728	0.822
salt_level200:co(CS)[4]	0.0041496	0.00129401	81	3.206773	**0.0019**
salt_level200:sp(CS)[4]	0.0003749	0.00129401	81	0.289705	0.7728
salt_level200:Kharchia[4]	0.0009923	0.001294	81	0.766813	0.4454

Linear Mixed Effects Modeling of Root Biomass*					
Akaike Information Criterion (AIC)	137.2866				
Bayesian Information Criterion (BIC)	161.7131				
Log-Likelihood	-58.64329				
	Value	Standard Error	DF	t-value	p-value
Intercept[1]	1.195	0.09669393	81	12.358583	0
salt_level200[2]	-0.0022625	0.00068373	4	-3.309058	**0.0297**
co(CS)[3]	-0.6513636	0.13981919	81	-4.658614	**0**
sp(CS)[3]	0.0777273	0.13981919	81	0.555913	0.5798
Kharchia[3]	0.1641667	0.13674587	81	1.200524	0.2334
salt_level200:co(CS)[4]	0.002236	0.00097787	81	2.286597	**0.0248**
salt_level200:sp(CS)[4]	-0.0000511	0.00097787	81	-0.052294	0.9584
salt_level200:Kharchia[4]	0.0000848	0.00097787	81	0.086769	0.9311

*Both models fit by Residual Maximum Likelihood (REML)

[1]Chinese Spring at 0 mM salt

[2]The change (coefficient) when salt level changes from 0 to 200 mM for Chinese Spring

[3]The change in dry weight from Chinese Spring to the indicated genotype when the salt level is 0

[4]The coefficient changes from Chinese Spring to the indicated genotype when the genotype changes from 0 to 200 mM salt

CHAPTER 3

Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging

Ali Moghimi, Ce Yang, Peter M. Marchetto

Department of Bioproducts and Biosystems Engineering, University of Minnesota, MN, USA

Abstract

Hyperspectral imaging is becoming an increasingly popular tool for high-throughput plant phenotyping because it provides remarkable insights about the health status of plants. Feature selection is a key component in hyperspectral image analysis, largely because a significant portion of spectral features is redundant and/or irrelevant, depending on the desired application. This paper presents an ensemble feature selection method to identify the most informative spectral features for practical applications in plant phenotyping. The hyperspectral dataset contained images of four wheat lines, each with a control and a salt (NaCl) treatment. To rank spectral features, six feature selection methods were used as the base for the ensemble: correlation-based feature selection, ReliefF, sequential feature selection, support vector machine-recursive feature elimination (SVM-RFE), LASSO logistic regression, and random forest. The best results were achieved by the ensemble of ReliefF, SVM-RFE, and random forest, which drastically reduced the dimension of the hyperspectral dataset from 215 to 15 features, while improved the accuracy in classifying the salt-treated vegetation pixels from the control pixels by 8.5%. To transform the hyperspectral dataset into a multispectral dataset, six wavelengths as the center of broad multispectral bands around the most prominent features were determined by a clustering algorithm. The result of salt tolerance assessment of the four wheat lines using the derived multispectral dataset was similar to that of the hyperspectral dataset. This demonstrates that the proposed feature selection pipeline can be utilized for determining the most informative features, and can be a valuable tool in the development of tailored multispectral cameras.

Keywords: Band selection, classification, ensemble feature selection, hyperspectral imaging, machine learning, multispectral imaging, plant phenotyping, salt stress, wheat.

3.1 Introduction

Hyperspectral imaging (HSI) integrates imaging and high resolution spectroscopy whereby a continuous spectrum is scanned for each pixel, typically across the visible and near infrared range of electromagnetic spectrum. Because of this unique aspect, HSI has been utilized as a fascinating and intriguing research tool in various domains including agriculture (Mohd Asaari et al., 2018; Thomas et al., 2018), environmental surveillance (Schmid et al., 2016; Veraverbeke et al., 2018), mineralogy (Aslett, Taranik, & Riley, 2018; Bedini & Rasmussen, 2018), biotechnology (Feng et al., 2018; Moghimi et al., 2017), medical diagnoses (Beaulieu et al., 2018; Liu, Wang, & Li, 2012; G. Lu & Fei, 2014), and pharmaceuticals (Masia, Pope, Watson, Langbein, & Borri, 2018; Pavurala, Xu, & Krishnaiah, 2017). HSI has recently drawn substantial attention in plant biology because it enables extensive investigation on the internal activities, physiological dynamics, and cell structure of plants (A. K. Mahlein et al., 2012; Moghimi, Yang, Miller, Kianian, & Marchetto, 2018; Wahabzada et al., 2016). Recently, hyperspectral sensors have become more compact, more lightweight, and more available, and this has introduced new opportunities for field applications because hyperspectral images can be captured via unmanned aerial vehicles.

Despite impressive progress of hyperspectral sensor development and autonomous platforms, there are still several challenges in the post-processing and analysis portion of the workflow, hindering the comprehensive application of HSI. Analysis of hyperspectral images is oftentimes challenging, mainly due to the complexity and high-dimensionality of the hyperspectral data. To find the correlation between the spectral response of plants and the desired traits, the complex and high-dimensional hyperspectral images must be interpreted appropriately. Among the hundreds of wavelengths scanned by the hyperspectral imager, a small set of wavelengths may be related to the desired traits, and the remaining are oftentimes either irrelevant or redundant. These redundant and irrelevant features make interpretation complex, and increase risk of overfitting, computational cost, and required storage space. In addition, when applying classification algorithms to a dataset with a high dimension of features a larger number of annotated samples (e.g., pixels) are

required for each class (referred to as 'the curse of dimensionality'). Another drawback of high-dimensional feature space is that the performance of subsequent modeling (e.g., classification, clustering, regression, etc.) might be deteriorated due to the presence of redundant and irrelevant features that merely expand the dimension of feature space without offering new information. Therefore, dimensionality reduction is a beneficial pre-processing step in implementing machine learning algorithms for hyperspectral image analysis.

Two main approaches exist for dimensionality reduction of a high-dimensional dataset: feature extraction and feature selection. In the feature extraction approach, all original features are combined in a transformation process to map from a high-dimension to a low-dimension feature space. For example, principal component analysis and linear discriminant analysis (two popular feature extraction methods) require all features to project to a lower-dimensional feature space. Alternatively, in feature selection, a subset containing discriminative features is selected such that it captures the most valuable information with minimum redundancy. The original semantics of features do not change using the feature selection approach, and this is advantageous because the subset is interpretable by a domain expert (Saeys, Inza, & Larranaga, 2007).

There are two main reasons that feature selection is preferred over feature extraction for hyperspectral image analysis. First, feature selection maintains the physical meaning associated with the measured reflectance per each wavelength as they are the functions of the object characteristics such as plants' internal/physiological activities, whereas with feature extraction, the physical information is lost through transforming to a new feature space. Secondly, all features must still be measured in feature extraction approach because each feature in the new space is a combination of all the original features. Alternatively, if a subset of the original wavelengths can be identified for a specific type of application using the feature selection approach, then a multispectral camera can be designed for that particular application to leverage the advantages incorporated with multispectral imaging. Multispectral cameras are less expensive and complex compared to hyperspectral cameras, and thus enable wider application of spectral imaging for trait analysis among plant

breeders and plant scientists.

Feature selection algorithms can be classified in three general subsets: filter, wrapper, and embedded methods, each of which has advantages and drawbacks (Saeys et al., 2007).

Filter methods rank features based on a statistical criteria, such as distance, correlation, and information, without involving classification/regression models (i.e., no learning is involved). Filter methods are either *univariate* or *multivariate.* In univariate methods, each single feature is ranked independently, whereas multivariate methods incorporate feature dependency (Saeys et al., 2007), inferring the ability of multivariate methods in handling feature redundancy (Tang, Alelyani, & Liu, 2014). Filter methods are fast and computationally efficient, and are therefore scalable to large datasets. The major drawback of filter methods is that the selected feature subset may not be the best feature subset for classification purposes since they were selected regardless to their effects on the performance of classifiers (Saeys, Abeel, & de Peer, 2008). Some of the widely used filter methods are correlation-based feature selection (Chutia, Bhattacharyya, Sarma, & Raju, 2017), ReliefF (Chutia et al., 2017), and information gain (Koonsanit, Jaruskulchai, & Eiumnoh, 2012).

Wrapper methods include a feature subsets search algorithm wrapped around a classification algorithm (Kohavi & John, 1997). The search algorithm explores the space of all possible feature subsets and generates various feature subsets. The performance of the classifier is evaluated using the generated feature subsets, and subsequently an optimal feature subset is identified based on the classification performance. Wrapper methods tend to offer improved results compared to filter methods because the best feature subset is selected by incorporating the performance of a predefined classifier algorithm as the selection criteria. Moreover, wrappers have the ability to handle feature dependencies. These advantages of wrapper methods are achieved at the cost of losing computational efficiency and being biased towards the classifier algorithm (Jovic, Brkic, & Bogunovic, 2015). Examples of wrappers methods include sequential feature selection (Rady, Ekramirad, Adedeji, Li, & Alimardani, 2017), genetic algorithms (Nagasubramanian et al., 2017; Thorp, Wang, Bronson, Badaruddin, & Mon, 2017), and support vector machine-

recursive feature elimination (SVM-RFE) (Archibald & Fann, 2007)

In embedded methods, selecting an optimal feature subset is embedded in the construction of a classifier. Embedded methods offer a high accuracy due to the search process being guided by the learning process. However, in contrast to wrapper methods, embedded methods tend to have less computational cost because feature selection is conducted during the classifier construction (i.e., no need to iteratively run the classifier for all feature subsets). Similar to wrappers, feature selection in embedded methods inevitably depends on the classifier that was utilized, meaning that the selected feature subset may not necessarily be effective for other classifiers (Hira & Gillies, 2015). Random forest (J. C. W. Chan & Paelinckx, 2008) and LASSO (Qian, Ye, & Zhou, 2013) are among the most representative algorithms of embedded methods.

Each of the feature selection approaches has advantages and disadvantages, and tend to be tailored for specific applications. Despite the availability of several feature selection approaches and a substantial number of feature selection algorithms, defining an efficient benchmark to select predominant spectral features has remained a challenge in hyperspectral image analysis (Damodaran, Courty, & Lefevre, 2017).

This study was motivated by the need to reduce the complexity and high-dimensionality of hyperspectral images in high-throughput plant phenotyping where HSI has recently drawn a substantial attention and introduced new opportunities. The first objective of this research was to develop an ensemble feature selection pipeline to aggregate the benefits of multiple feature selection approaches, therefore increasing the stability and accuracy of selecting predominant spectral features from hyperspectral images. The second objective was to rank the spectral features based on their ability to discriminate salt-stressed wheat plants from healthy plants at the earliest stages of stress. The purpose of ranking spectral features was to form six broad spectral bands around the most prominent features to aid in development of a multispectral camera. The third objective was to use these multispectral bands to assess the salt tolerance of four wheat lines in the context of phenotyping and evaluate results based on findings in previous studies.

The remainder of the paper is organized as follows. Section II describes the hyperspectral dataset containing the images of four wheat lines with various levels of salt tolerance. Section III introduces the ensemble feature selection method proposed in this study for spectral feature selection. Section IV reports the ranking of spectral features and discusses the experimental results of salt stress phenotyping in wheat using the selected multispectral bands. Section V draws the conclusion and presents the idea of developing a digital spectral library for plant diseases and stresses in future work.

3.2 Hyperspectral dataset

A hyperspectral dataset from a phenotyping study (Moghimi, Yang, Miller, et al., 2018), freely accessible to the public at (Moghimi & Yang, 2018), was used in this study. They investigated the salt stress tolerance in four wheat lines using HSI. The hyperspectral dataset for each wheat line contained the spectral reflectance of leaves for a null control and a salt treatment (as sodium chloride; Table 3.1). The spatial resolution was about 1 mm and the wavelengths range was from 400 nm to 900 nm with a spectral resolution of about 2.1 nm. In this study, the dataset of the most susceptible line (Chinese Spring) was selected among the four wheat lines to find sensitive spectral features for salt stress phenotyping of wheat. Once the spectral features were identified, all four hyperspectral datasets of wheat lines were utilized to evaluate the feasibility of the selected bands in ranking the wheat lines based on their tolerance to salt stress. The matrix of the data was initially normalized to make each spectral feature have zero-mean and unit-variance.

Table 3.1. Hyperspectral image dataset of four wheat lines to assess salt stress (Moghimi & Yang, 2018)

Wheat lines	Abbreviation	Number of spectral features	Number of pixels	class size	
				C0 (control)	C1 (salt)
Chinese Spring	CS	215	12896	7060	5836
speltoides aucheri KU2201B (CS)	sp(CS)	215	11665	5590	6075
Aegilops columnaris KU11-2 (CS)	co(CS)	215	5228	1878	3350
Kharchia	-	215	14652	9013	5639

3.3 Proposed ensemble feature selection method

3.3.1 Feature Selection Methods

The hyperspectral dataset of CS was split into training (70% of dataset) and test (30% of dataset) datasets (Figure 3.1-step I). Training dataset was used by feature selection techniques, referred to as rankers in this study, and the test dataset was employed to evaluate all potential ensembles of rankers. The following six rankers were the individual base rankers and used to construct an ensemble feature selection pipeline (Figure 3.1-step II).

3.3.1.1 Correlation-based feature selection (CFS)

CFS is a univariate filter method that measures the Pearson's correlation coefficient between features and the output variable (Hall, 2000). The spectral features are ranked based on their correlation to the class. In this study, the number of components in the feature subset was considered to be one to identify the rank of each individual feature. Therefore, CFS ignored features redundancy and correlation between features since the features were treated independently.

3.3.1.2 ReliefF

ReliefF is also a univariate filter method that assigns weight to features through an iterative process (Kira & Rendell, 1992; Kononenko, Šimec, & Robnik-Šikonja, 1997). In each iteration, an instance is randomly selected and the weight of features is updated based on the distance between the randomly selected instance and the *k* closest instances from each class. The weight of a given feature increases at each iteration if it makes the selected instance be more similar to its neighbors in the same class and be more distinct from its neighbors from the other class. In this study, the iteration was performed for all samples and the number of neighbors, *k,* was set to 10.

3.3.1.3 Sequential feature selection - forward selection (SFS-forward)

SFS-forward is a wrapper method to find an optimal feature subset using a greedy search

algorithm. It starts with an empty subset and adds one feature through sequential iteration to the subset based on the performance of a predefined classifier. Quadratic discriminant analysis (QDA) was used in this study as the classifier. The order in which all features added to the subset denoted the ranking of features.

3.3.1.4 Support vector machine-recursive feature elimination (SVM-RFE)

SVM-RFE is another wrapper method in which support vector machine is used as the classifier (Guyon, Weston, Barnhill, & Vapnik, 2002). Liner SVM classifier is trained using the training dataset containing all features in the first iteration. Then SVM-RFE eliminates one single feature with the lowest weight among all weights required to define the best hyperplane in separating the classes. These weights are assigned by the SVM classifier to features. This procedure was recursively executed until only one feature remained in the feature dataset to rank all features based on their elimination order through the recursive process.

3.3.1.5 LASSO Logistic Regression

The idea of adding L1 regularization into the objective function of linear least-squares regression, so-called *LASSO,* was first proposed by (Tibshirani, 1996) to shrink coefficients and control the model complexity. The regularization term defines an upper bound on the sum of the absolute values of the model coefficients to prevent overfitting and make the coefficient of irrelevant features equal to zero. Therefore, the coefficient vector of features becomes sparse. For classification purpose, the LASSO estimator can be added to Logistic Regression model, and coefficients (β) can be obtained as follows:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\|Y - \beta X\|_2^2 + \lambda \|\beta\|_1 \right) \tag{3.1}$$

where λ is the tuning parameter which controls the balance between bias and variance. The larger it is the larger the number of coefficients are penalized to be equal to zero, and hence the variance decreases while bias increases. A number of 100 different values (between 0 and 1) for λ were tested to obtain the optimized values for coefficients ($\hat{\beta}$). Since the spectral features were first centered and scaled, the absolute value of non-zero coefficients

implies the importance of features, while irrelevant features have coefficients of zero. LASSO is an embedded method for feature selection.

3.3.1.6 Random Forest

Random forest is another embedded method building an ensemble of de-correlated decision trees (Breiman, 2001), each trained on a training set created by bootstrap aggregation in which samples are randomly drawn from initial training set with replacement. To split each node in each decision tree, a randomly selected subset of features is evaluated. Based on the defined split-criterion, which was entropy in this study, a feature with the best strength of classification is selected among the randomly selected feature for each node to divide the training samples into smaller and more pure subsets. The importance of features is calculated based on how much the splitting feature can decrease the weighted impurity at that node, and it is then averaged over all the trees in the forest. Number of trees was 1000 in this study.

3.3.2 Ensemble feature selection method

Each ranker has its own drawback such that the final selected feature subset may not be the optimal feature subsets in the feature space. To aggregate the benefits of each feature selection approach and increase the robustness and accuracy of the final selected feature subset, ensemble feature selection approach was used (originally proposed by (Saeys et al., 2008)). Ensemble feature selection method reduce the risk of capturing local optima where a single ranker may be trapped. Various methods have been used to combine the feature ranking obtained by each base ranker (Abeel, Helleputte, Van de Peer, Dupont, & Saeys, 2009; Ben Brahim & Limam, 2017). In this study, the complete linear aggregation method was used to average the ranking *f* assigned to feature *i* by ranker *j* as presented in Eq. (3.2) , and the ranking values of features was then sorted to list features based on their importance.

$$Rank_i = \frac{1}{n}\sum_{j=1}^{n} f_{ij} \quad i = 1,2,\ldots,215, \text{ and } n = 6 \tag{3.2}$$

In this equation, the rankers had the same weight, which is equal to one over the number of rankers. To ensure the optimal feature subset is achieved, all base rankers used for aggregation should be accurate and diverse (Kuncheva, 2014). However, based on the dataset and the nature of the problem, a given ranker may not be accurate or may contribute to redundancy in the ensemble process. Therefore, a more intelligent approach was designed to evaluate which combination of these six rankers would result in the selection of the best feature subset based on the classification of vegetation pixels into control and salt classes.

In this study, the decision on exclusion of a ranker was made through a recursive elimination process. In this greedy search, a ranker was removed at each step if its elimination resulted in the best classification performance among all choices for elimination in that step. For instance, at step 1, there were six iterations, at each of which one ranker was eliminated. At the end of step 1, the ranker whose elimination resulted in the best classification performance among the six options was eliminated for the next step. This process continued until only one ranker remained. Algorithm 1 presents a summary of the recursive elimination of rankers. Besides the ensemble of all six rankers (Figure 3.1-step III), there were five other ensembles of rankers each of which was obtained at each step (Figure 3.1-step IV). The combination of rankers that offered the best classification performance was selected among these six sets. This approach assures that the presence of a ranker would not deteriorate the classification result.

Two criteria related to the classification performance were defined to compare all potential ensembles of rankers at each step.

The first criterion was the size of feature subset to attain the minimum classification error of a quadratic discriminant analysis (QDA) classifier. This process was performed on the training dataset, containing 70% of the whole dataset, with 10-fold cross validation. To identify the best feature subset, the process of training and testing with QDA was conducted in an iterative process in which the ensemble feature list, obtained at iteration k by Eq. (3.2), was imported to QDA using a stepwise approach until all features were exhausted. Therefore, at the ith iteration, QDA was trained and tested 10 times using only the first ith

features. For instance, at the first iteration, only the first feature at the top of the ensemble list was used, and all 215 features were employed at the 215th iteration.

The mean of cross validation error (CV error) and standard deviation of error over 10 folds were recorded for each iteration. The standard deviation of error indicates the error rate uncertainty; thus, one-standard error rule was used as a conservative metric to determine feature subset size (Hastie, Tibshirani, & Friedman, 2009). Using this approach, the most parsimonious feature subset within one standard error (1SE) of the minimum error was selected as the best feature subset. In other words, the mean of error for this parsimonious subset should not be more than 1SE above the minimum error.

The second criteria was the classification accuracy of a QDA algorithm evaluated on the test dataset (the remaining 30% of the dataset), which has not been used during any of the feature selection and parameter tuning process of rankers. For this purpose, the QDA algorithm was trained on the whole training dataset using only the features obtained by the 1SE rule. The performance of the trained QDA was evaluated on the test dataset. The metric used for classification accuracy was F1-mean which was defined as a mean between F1 measures of both control ($F1_{C0}$) and salt ($F1_{C1}$) classes.

The procedure of eliminating a ranker per iteration was considered as a multiple criteria decision-making problem in which the performance of feasible ensemble of rankers, obtained at iteration k, was evaluated based on the defined criteria: feature subset size and F1-mean. To allocate weight to the alternatives, we utilized the technique for order preference by similarity to an ideal solution (TOPSIS) proposed by (Hwang & Yoon, 1981) as a simple and useful method for multiple criteria decision making problems. TOPSIS dispensed the weight among feasible ensembles at each iteration k based on the Euclidian distance of each ensemble to the negative ideal solution divided by the summation of distance to the positive and negative ideal solutions. The positive ideal solution was a solution capturing the maximum benefit criteria (i.e., F1-mean) and the minimum cost criteria (i.e., minimum required number of features or feature subset size), and vice versa for the negative ideal solution. The best ensemble is the one that has the maximum weight, indicating it has the maximum normalized distance to the negative ideal solutions. TOPSIS

facilitates decision-making on eliminating rankers and ultimately determining the best ensemble of rankers. Figure 3.1 shows the flowchart of the proposed pipeline for ensemble feature selection.

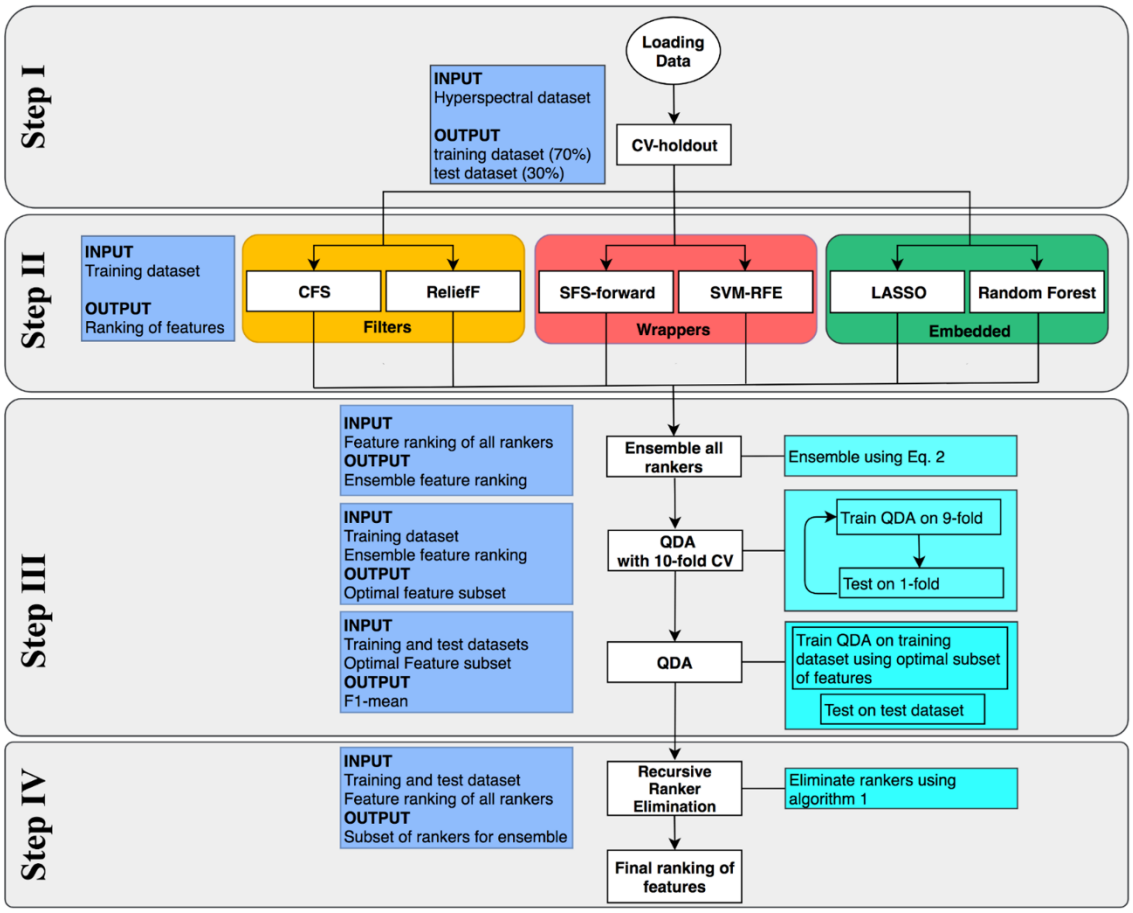

Figure 3.1. Flowchart of the proposed pipeline for ensemble feature selection. Step I) Splitting the data into training set (70%) and test set (30%) using cross validation – holdout method. Step II) Ranking the spectral features using six base rankers: correlation-based feature selection, ReliefF, sequential feature selection - forward selection, support vector machine-recursive feature elimination, and random forest. Step III) Ensemble of all six rankers. In this step, the optimal feature subset size and F1-mean for the ensemble of all six rankers are calculated. Step IV) Eliminating rankers through a recursive process presented in Algorithm 1 to identify the best ensemble of the rankers.

Algorithm 1. Recursive ranker elimination to identify the best ensemble of the rankers

Let $\mathcal{X}_{train} = \{(x_{ij}, y_i) | \ i = 1, 2, \dots, n_{trian}\ ,\ j = 1, 2, \dots\ ,\ 215\}$ be the training dataset
Let $\mathcal{X}_{test} = \{(x_{ij}, y_i) | \ i = 1, 2, \dots, n_{test}\ ,\ j = 1, 2, \dots\ ,\ 215\}$ be the test dataset
Let $\mathcal{F} = \{[f_{ij}] \mid i = 1, \dots, 215, j = 1, 2, \dots, n\}$, a matrix containing ranking of features for all *n* ranker
For step = 1 to (n – 1) **do**
    For k = 1 to (number of columns in $\mathcal{F}$) **do**
        $\hat{\mathcal{F}}_k = \{[f_{ij}] \mid i = 1, \dots, 215, j = \{1,2, \dots,6\} \backslash k\}$, eliminate the kth column of $\mathcal{F}$
        E: ensemble $\hat{\mathcal{F}}_k$ using Eq 2
        R: sort features based on their ensemble ranking E
        For i = 1 to 215 **do**
            r = R[1:i], the first i features at the top of the list
            Train and test QDA on $\mathcal{X}_{train}$ with 10 folds cross validation using features $\in r$
            e_i: mean of error for 10 folds
            std_i: standard deviation of e_i
        End For
        Err $= \min(e_i)$
        $SE = \frac{std_{argmin(e_i)}}{\sqrt{10}}$
        m: Find the size of the most parsimonious feature subset whose error is less than (Err + SE)
        Train QDA on $\mathcal{X}_{train}$ using only the first m feature in R
        Test the trained QDA classifier on $\mathcal{X}_{test}$
        Calculate $F1$ measure for both C0 and C1 classes
        F1-mean: mean of $F1_{C1}$ and $F1_{C0}$
        $C_k = [m, F1]$, two criteria used by TOPSIS to allocate weights
    End For
    W_k: weights assigned by TOPSIS to each ensemble
    $ind = argmax(W_k)$, denotes the index of the rankers to be eliminated
    $M_{step} = C_{ind}$, the performance criteria of the best ensemble obtained at this iteration
    $\mathcal{F}[:, ind] = [\ ]$, eliminate the ranking list whose elimination resulted in maximum TOPSIS weight
End For
Return M_{step}, presented in Table 3.3

3.3.3 Clustering of top-ranked features to develop broader multispectral bands

After determining the optimal ensemble and the subsequent ranking of features, Algorithm 2 was developed to cluster the most informative features. The clustering algorithm was designed to form six broad spectral bands around the most prominent features to mimic the spectral bands of a typical multispectral camera (six bands were chosen because multispectral cameras rarely possess more than six spectral channels). The six most informative bands are composed of multiple nearby spectral features at the top of the ensemble feature list. The clustering algorithm was initialized with a set of six cluster

centers with zero values, then it clustered the features through an iterative loop starting from the first feature at the top of the ensemble ranking list. At each iteration, one feature was assigned to a cluster if it is within 10 nm of the center of cluster, otherwise was allocated to the next empty cluster with zero center. At the end of each iteration, the cluster centers were updated. This process of clustering continued until the next feature could not be assigned to any of the six clusters. The input of Algorithm 2 was the ranking of features obtained by Algorithm 1, and the output was the center of broad bands to select an appropriate set of optical band-pass filters for developing a multispectral sensor.

Algorithm 2. Clustering of the top features to develop broad multispectral bands

Let $\mathcal{F} = \{ f_i \mid i = 1,\ 2, \ldots, 215\}$ be the features ranking obtained by Algorithm 1
Let $\mathcal{C} = \{ C_j = 0 \mid j = 1, 2, \ldots, m_c\}$ be the cluster centers initialized with zero values
Let m_c denote the number of desired clusters, here m_c = 6
Let n_c denote the number of filled clusters, initially n_c = 0
for *i* = 1:215 **do**
    $D_j = |f_i - C_j|$, compute the distance of $f_i \in \mathcal{F}$ to cluster centers
    $M_d = \min(D_j)$, find the minimum of distance
    if M_d < 10 nm **then**
        assign f_i to C_j, $j = \text{argmin}(D_j)$
    else
        $n_c = n_c + 1$
        if $n_c <= m_c$ **then**
            assign f_i to the first empty cluster
        Else
            break
        end if
    end if
    update the cluster center $\mathcal{C} = \left\{ C_j = \frac{1}{n}\sum_{k=1}^{n} f_k \mid f_k \in C_j, n = |C_j| \right\}$
end for
Return $\mathcal{C}$, containing six centers of clustered features

3.3.4 Phenotyping the salt tolerance of wheat lines using selected multispectral bands

The ultimate goal of clustering spectral features was to identify the most discriminative broad spectral bands for the salt stress phenotyping of wheat. To test the potential of the selected broad bands for salt stress phenotyping in wheat lines, the hyperspectral dataset for each of the four wheat lines was transformed to multispectral dataset. Each multispectral band was obtained by averaging the adjacent hyperspectral bands within ±5

nm of the cluster centers calculated by Algorithm 2. To identify how many of these six bands might be sufficient for salt tolerance assessment, the multispectral bands were ranked in a recursive elimination process described in Algorithm 3 (similar approach to Algorithm 1). In Algorithm 3, a band is removed at each iteration if its elimination led to the least classification error averaged across all wheat lines (Algorithm 3). The classification error referred to the summation of the mean and standard deviation of error rate obtained for 10-fold CV. During the last iteration, Algorithm 3 saved the most discriminative band that had the lowest classification error at that iteration across all four lines. In essence, Algorithm 3 ranked the six bands based on the mean of classification error across four wheat lines obtained by removing a band at each iteration.

After determining and ranking the most informative multispectral bands, then the tolerance of wheat lines to salt stress were assessed using techniques proposed by . These techniques included the minimum difference of pair assignments (MDPA) and Bayesian inference, the latter of which returned the posterior probability (i.e., P(salt|x)) of belonging to salt class given an observation *x*, which was defined as the similarity of pixel to the salt endmember. To compare differences between the values obtained by using multispectral and hyperspectral datasets for each of these techniques, the normalized root mean square error (NRMSE) was calculated as follows:

$$NRMSE = \frac{\sqrt{(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2)/n}}{(y_{max} - y_{min})} \tag{3.3}$$

where y denotes the value calculated for one of the techniques (posterior probability or MDPA) using hyperspectral dataset, $\hat{y}$ refers to the value obtained for the same technique by using multispectral dataset, and n is equal to the number of wheat lines.

Algorithm 3. Ranking of six broad multispectral bands

Let $\mathcal{X}_L = \{(x_{ij}, y_i) | \ i = 1, 2, \dots, N, \ j = 1, 2, \dots, \ m, m = 6\}$ be the multispectral dataset for the Lth wheat line
n_bands = 6, initial number of bands before the elimination process
for step = 1: (m -1) **do**
    for L = 1:4 **do**, for all four wheat lines
        for b = 1:n_bands **do**
            $\widehat{\mathcal{X}}_L = \mathcal{X}_L$
            $\widehat{\mathcal{X}}_L = \{(x_{ij}, y_i) | \ i = 1, 2, \dots, N, \ j = \{1, 2, \dots, \ n\_bands\} \backslash b\}$, eliminate the bth column from dataset
            Train and test QDA on $\widehat{\mathcal{X}}_L$ with 10-fold CV
            Record mean of error and standard deviation of error for 10-folds CV
            $e_{Lb} = mean(error) + std(error)$, error for Lth wheat line by removing bth band
        end for
    end for
    $ME_b = \frac{1}{4}\sum_{L=1}^{4} e_{Lb}$, b = 1, 2, …, n_bands; mean e_{Lb} across four wheat lines
    $ind = argmin(ME_b)$
    $\mathcal{X}_L[:, ind] = [\,]$, update the dataset of all four wheat lines by removing the band at column *ind*
    n_bands = n_bands – 1
end for
Return ranking of bands based on the order in which they were removed

3.4 Experimental Results and discussion

3.4.1 Band pair correlation

Adjacent spectral bands tend to be correlated in hyperspectral imaging. Figure 3.2 shows the correlation between all pairs of bands, which is a symmetric matrix represented as a colormap image. Among highly correlated spectral bands (~505 to ~645 nm), a few bands (567, 569, 589, 590, and 593 nm) had less correlation with their adjacent bands. These un-correlated bands are responsible for the two stripes in Figure 3.2 denoted by the arrows. In general, bands between ~505 to ~645 nm were highly correlated to each other and also correlated to bands between ~690 to ~715nm, indicating that some of these spectral features are redundant and do not contribute valuable information.

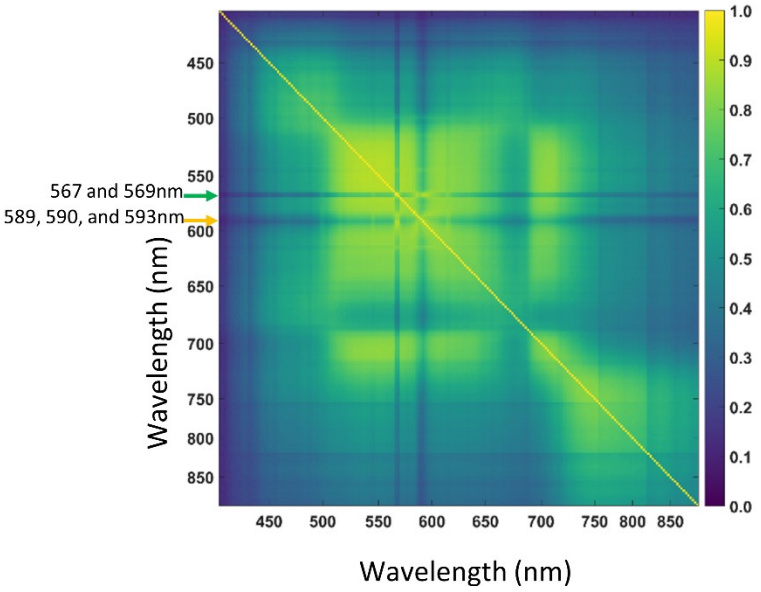

Figure 3.2. Correlation between all pairs of spectral features (i.e. 215 spectral bands from 400 nm to 900 nm) presented as a colormap image.

3.4.2 Ranking of spectral features by six rankers

The high correlation between features emphasized the importance of feature selection to eliminate the redundant features while retaining the relevant information. Each ranker utilized in this study ranked the spectral features based on their importance (Figure 3.3). Four of the six rankers (i.e., random forest, LASSO, ReliefF, and SVM-RFE) identified 589 nm as the most informative spectral feature among all 215 bands. Intriguingly, 589 nm is among the Fraunhofer lines (i.e. spectral absorption lines of elements) and associated with sodium (*Na*). It is a promising result considering sodium chloride (NaCl) was the applied salt treatment to the wheat plants. This result suggests that these four rankers were able to capture the fundamental physical meaning concealed in the spectral response of wheat plants treated with NaCl. It is remarkable that the hyperspectral images were captured only one day after the salt treatment was applied. SFS-forward ranked 589 nm as the third most important feature and ranked 751 nm and 532 nm as the first two important spectral features. CFS ranked 589 nm as the 210th most informative feature and 583 nm as the first most informative feature despite the fairly close correlation between these two bands (i.e., 0.77). Except SFS-forward, all rankers had their first two dominant features between 581 nm and 589 nm, indicating the importance of this region of the

electromagnetic spectrum for determining salt stress in wheat.

The other inference from Figure 3.3 is that the spectral features in blue (around 450 nm) and red (650 nm) regions, known as the chlorophyll absorption bands, had the least importance (i.e., largest ranking value) among all the spectral features.

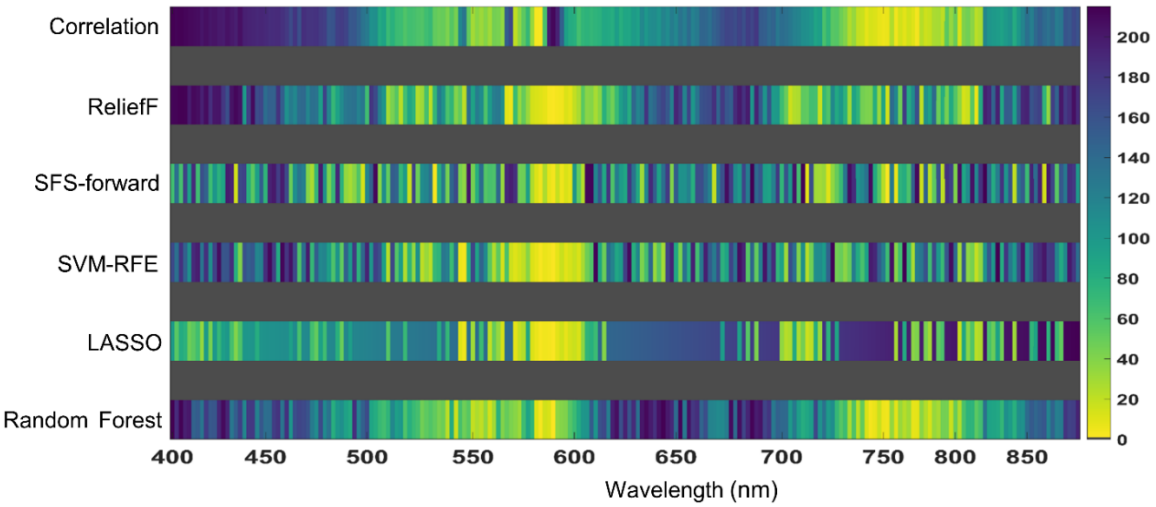

Figure 3.3. Ranking of the 215 spectral features for each of the six rankers. Top-ranked spectral features are represented with lighter color and low-ranked features are represented with darker color. Dark blue refers to the least important features with large ranking values and light yellow refers to the most important features with low ranking values. Gray color (zero value in colorbar) was used to separate the results of rankers.

Feature subset size and F1-mean were computed for each ranker to compare the rankers with each other and further compare them with the results obtained by the ensemble of rankers (Table 3.2). The QDA classifier required a different number of features from the feature set ranked by each ranker to attain a CV error with 1SE of the minimum error. The lowest CV error on the training dataset and the maximum F1-mean on the test dataset were both achieved when QDA used the feature set ranked by SFS-forward. However, this performance was attained at the cost of larger feature size (28 features) and higher standard deviation of CV error. Alternatively, the lowest feature subset size and standard deviation of error was achieved using random forest. The F1-mean for random forest dropped only 1.63% compared to that of SFS-forward, while the feature subset size for random forest

(12 features) was significantly less than SFS-forward (28 features). Among all rankers, CFS resulted in the lowest classification accuracy as indicated by the lowest $F1_{C0}$ and $F1_{C1}$ values.

Table 3.2. Classification performance of QDA on the training and test datasets for individual rankers (for training dataset: mean and standard deviation of cross validation error and feature subset size obtained by the one standard error rule; for test dataset: F1 measure of control (C0) and salt (C1) classes obtained by training QDA on the training dataset using only features in feature subset and testing on unseen dataset)

	Training dataset with 10-fold CV			Test dataset		
	Feature subset size	Error (%)	StD (%)	$F1_{C0}$ (%)	$F1_{C1}$ (%)	F1-mean (%)
CFS	21	27.33	1.30	71.74	71.76	71.75
ReliefF	19	21.57	1.47	78.41	75.85	77.13
SFS-forward	28	20.36	1.78	78.87	76.55	77.71
SVM-RFE	19	21.86	1.55	78.27	74.91	76.59
LASSO	18	21.39	1.68	78.23	75.20	76.71
Random Forest	16	22.58	1.29	77.29	74.87	76.08

3.4.3 Ensemble of feature selection techniques

The ranking assigned to the features by the six rankers were aggregated together (Figure 3.4 (a)) using Eq. (3.2) (Figure 3.1-Step III). As expected from the ranking results of the base rankers, spectral features near the 589 nm band were dominant at the top of the ranking list of features.

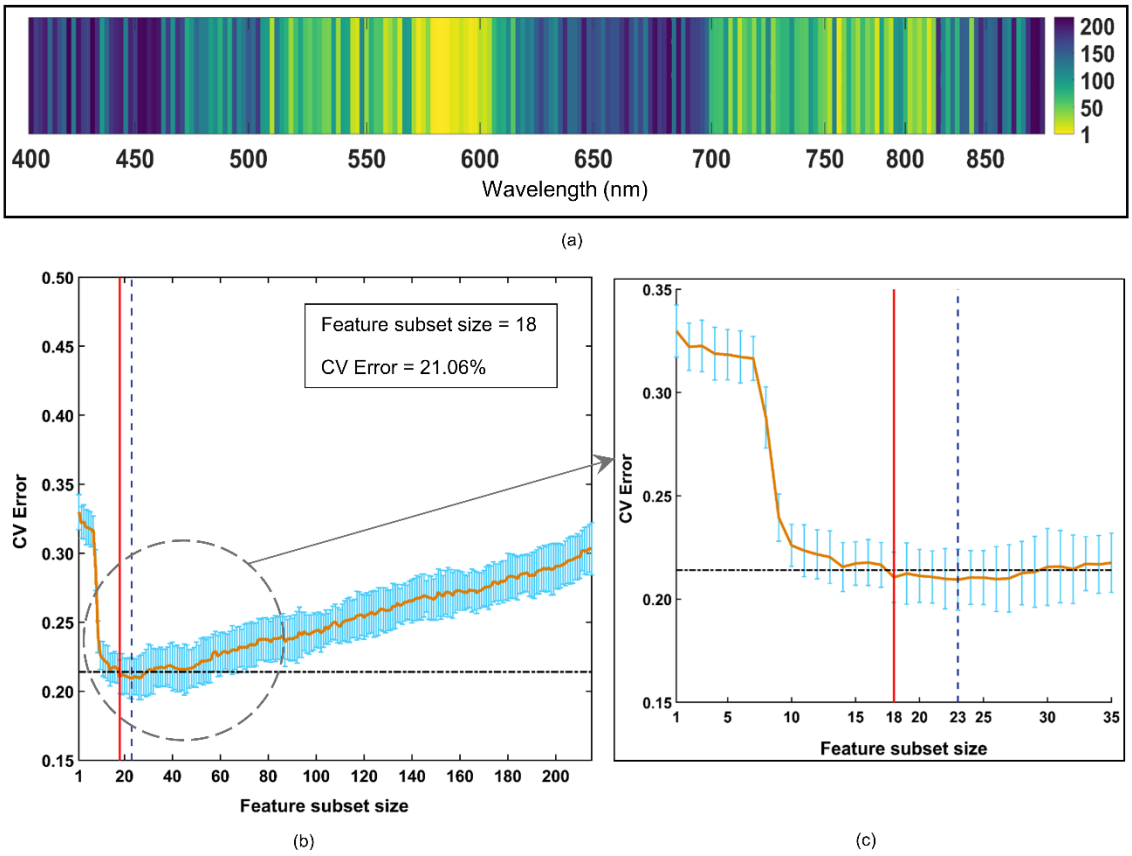

Figure 3.4. (a) Ranking of features obtained by the ensemble of all six rankers: correlation-based feature selection, ReliefF, sequential feature selection - forward selection, support vector machine-recursive feature elimination, and random forest. (b) Cross validation error (CV error) of QDA on the unseen fold of the 10-fold cross validation for each feature subset size. The feature subset size required to obtain the minimum CV error is shown by the vertical dashed line, the lowest error plus one standard error is presented by the horizontal dashed line, and the feature subset size of the most parsimonious model whose error was within 1SE of the minimum error is illustrated by the vertical solid line. (c) CV error for the first few feature subset size to show with more detail how the most parsimonious feature subset size is selected.

From the ensemble ranking list, a subset of features was selected by QDA based on the minimum CV error and 1SE rule (Figure 3.4 (b)). Figure 3.4 (b) illustrates the mean and standard deviation of the 10-fold CV error for all 215 feature subsets, in which the leftmost subset contains only the top ranked feature and the rightmost subset contains all features (the model complexity increases across the horizontal axis). CV error decreased with an increasing size of feature subset until feature subset contained 23 features whereby CV error began to increase, which underscore the need for feature selection. Because of the uncertainty around CV error of each feature subset (shown in Figure 3.4 (b)), it was

appropriate to determine the most parsimonious feature subset within one standard error of the minimum CV error for determining the optimum number of features. QDA classifier needed the first 18 features at top of the feature list ranked by the ensemble of all rankers to achieve an error rate within 1SE of the minimum CV error. To achieve a similar CV error obtained by using all the 215 features (CV error = 0.30), QDA required the top seven features from the feature list.

3.4.4 Recursive elimination of rankers

A base ranker was disregarded from the ensemble process if its elimination resulted in an improved classification performance among all options at a given step. The F1-mean and number of features in the subset were used by TOPSIS to determine which rankers should be eliminated from the ensemble process at each step (results are presented in Table 3.3). The highest F1-mean was achieved at step 1 using 19 features obtained by the ensemble of all rankers except CFS. However, elimination of LASSO and SFS-forward at the next two iterations resulted in four less features while only reducing F1-mean by less than 0.5% compared to step 1 that offered the largest F1-mean. Although the results obtained at step 2 and 3 were similar, the ensemble of three rankers at step 3 is preferred due to less complexity and less computation compared to the ensemble of four rankers at step 2. Therefore, the final ensemble feature subset was obtained by aggregating the ranking of ReliefF, SVM-RFE, and random forest, each of which belongs to a different feature selection category: filter, wrapper, and embedded, respectively.

The size of feature subset obtained by the ensemble of these three rankers (Table 3.3; Figure 3.5) dropped to 15 features, three features less than the ensemble of all six rankers. Furthermore, the difference between classification accuracy of the ensemble of three rankers (F1-mean = 77.07) and the maximum accuracy of individual rankers achieved by SFS-forward (F1-mean = 77.71) was less than 1%, while the model complexity of SFS-forward (28 features) was approximately twice as that of the three rankers ensemble (15 features). This demonstrates the benefit of using ensemble feature selection.

Table 3.3. Feature subset size and F1-mean after each step of the recursive elimination of rankers (Algorithm 3).

Recursive step	Removed technique	Feature subset size	F1-mean (%)
Full dimensional hyperspectral dataset	-	215	68.49
Step 0- Ensemble 6 techniques	-	18	77.32
Step 1- Ensemble 5 techniques	CFS	19	77.47
Step 2- Ensemble 4 techniques	LASSO	15	77.09
Step 3- Ensemble 3 techniques	SFS-forward	15	77.07
Step 4- Ensemble 2 techniques	ReliefF	17	76.89
Step 5- 1 technique	SVM-RFE	16	76.08

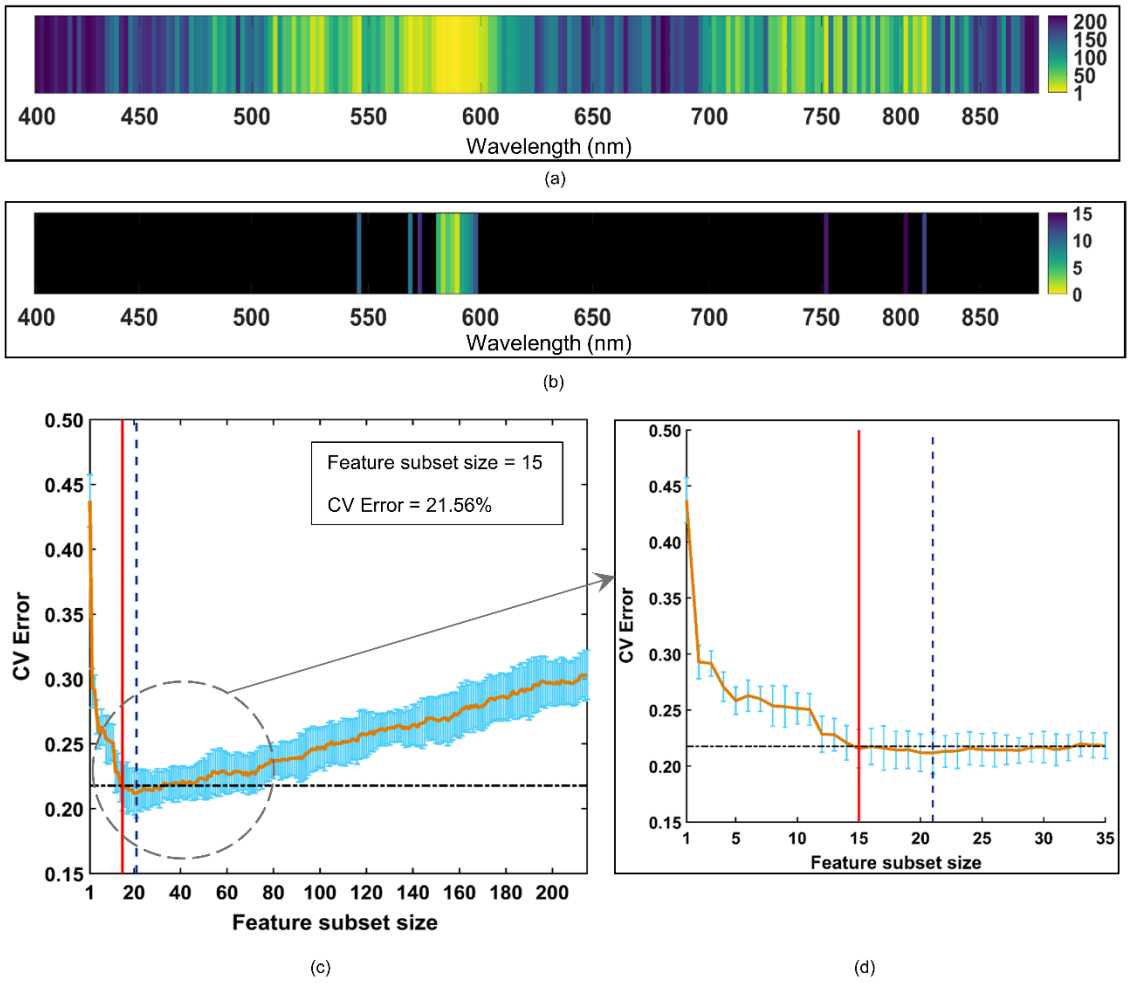

Figure 3.5. (a) Ranking of features obtained by the ensemble of selected rankers: ReliefF, SVM-RFE, and random forest. (b) The 15 features selected among the 215 spectral features scanned from 400 nm to 900 nm. (c) Cross validation error (CV error) of QDA on the unseen fold of the 10-fold cross validation for each feature subset size. The feature subset size required to obtain the minimum CV error is shown by the vertical dashed line, the lowest error plus one standard error is presented by the horizontal dashed line, and the feature subset size of the most parsimonious model whose error was within 1SE of the minimum error is illustrated by the vertical solid line. (d) CV error for the first few feature subset size to show with more detail how the most parsimonious feature subset size is selected.

Figure 3.5 (a) depicts the feature ranking achieved by the ensemble of ReliefF, SVM-RFE, and random forest, and Figure 3.5 (b) demonstrates the ranking and distribution of the 15 selected features over the scanned electromagnetic wavelengths. The band nearest to the region of sodium absorption (589 nm) and the adjacent bands were among the most informative and were consequently the top-ranked features.

Similar to the ensemble of all six rankers, the ensemble of the top three rankers resulted in a decreasing CV error of QDA to a point (21 features), followed by an increasing CV error as feature subset size increased. However, the CV error obtained by the ensemble of three rankers had a sharper decrease within the first few feature subsets compared to the ensemble of six rankers. The two top-ranked features (589 nm and 583 nm) from the ensemble of three rankers were able to achieve a similar CV error obtained by using all 215 features (CV error = 0.30), whereas the top seven features from the ensemble of six rankers were required to achieve a comparable CV error.

3.4.5 Clustering of top-ranked features

The ensemble feature list derived from the top three techniques was utilized in Algorithm 2 to find the first six clusters of spectral features whose mean values were at least 10 nm apart. The first 18 features were needed to form the six clusters (Table 3.4). Cluster 1 had the largest size (nine members), whereas cluster 3 and 5 had the smallest size (one member).

Table 3.4. Clustering of features ranked by the ensemble of three rankers: ReliefF, SVM-RFE, and random forest. Algorithm 1 was used for clustering (only integer part of wavelengths are presented; superscripts denote feature rank).

	Cluster center	Members of each cluster								
Cluster 1	589	589^{1}	583^{2}	587^{3}	585^{4}	581^{5}	591^{6}	593^{7}	595^{8}	597^{11}
Cluster 2	573	569^{9}	573^{13}	579^{17}						
Cluster 3	546	546^{10}								
Cluster 4	805	809^{12}	801^{15}							
Cluster 5	751	751^{14}								
Cluster 6	528	525^{16}	530^{18}							

3.4.6 Importance of multispectral bands

The spectral bands within ±5 nm of the six cluster centers (Table 3.4) from the hyperspectral dataset were averaged to imitate a multispectral dataset for each of the four wheat lines (Table 3.1). Algorithm 3 ranked these six bands via a recursive elimination process as

$$\boldsymbol{\mathcal{R}} = \{\mathbf{528}, \mathbf{805}, \mathbf{589}, \mathbf{573}, \mathbf{751}, \mathbf{546}\}$$

where 528 nm is the center of the most informative band that persisted after the last iteration and 546 nm is the center of the least important band eliminated during the first iteration. Among all wheat lines, Kharchia had the largest CV error across the hyperspectral and all multispectral datasets (Figure 3.6). This result can be explained by the fact that Kharchia is the most salt-tolerant line among the four lines (Moghimi, Yang, Miller, et al., 2018), indicating the spectral response of plants under the salt treatment were less likely to be affected by the imposed stress. Consequently, the classification algorithm failed to accurately discriminate between pixels that represented plants under each of the control and salt treatments for Kharchia. Alternatively, the lowest CV error was found for CS, which was the most susceptible line among the others. The lowest CV error for CS was achieved by using the six multispectral broad bands and it increased as the number of bands decreased.

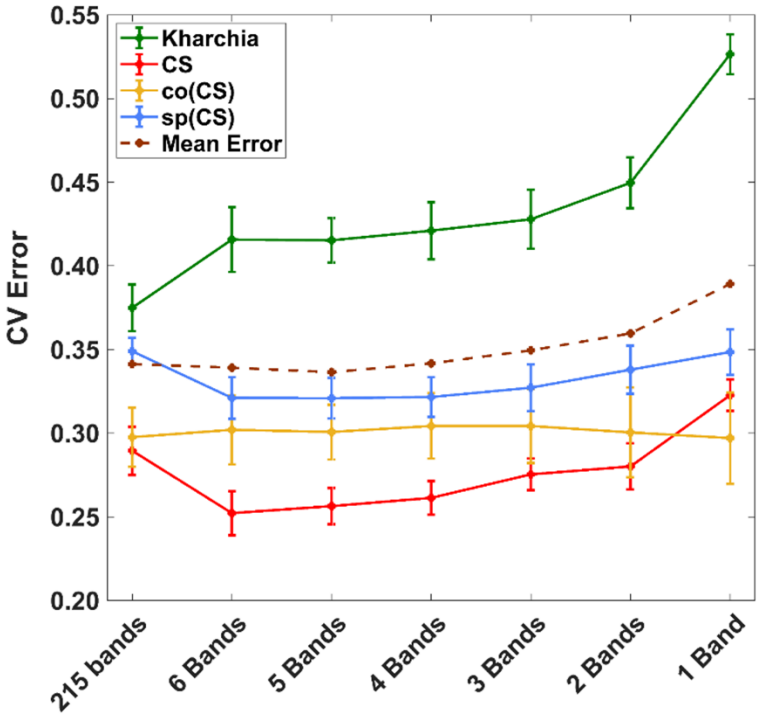

Figure 3.6. Mean of error across four wheat lines (dash line) and CV error of four wheat lines for hyperspectral (n = 215) and multispectral datasets using 10-fold cross validation. Bands were selected from the $\mathcal{R}$ set in order (i.e., *5 Bands* contains the first five bands in $\mathcal{R}$ and *1 Band* contains only the first band in $\mathcal{R}$).

3.4.7 Salt stress phenotyping of wheat lines using multispectral bands

The performance of multispectral bands in ranking wheat lines based on their salt tolerance was determined using methods used by (Moghimi, Yang, Miller, et al., 2018) (i.e., MDPA and posterior probability), and results were compared with those derived from the hyperspectral dataset (Figure 3.7). Using hyperspectral dataset and MDPA and Bayesian approaches, the ranking of salt tolerance for the evaluated wheat lines (from most tolerant to most susceptible) is: 1) Kharchia, 2) co(CS), 3) sp(CS), and 4) CS, which is consistent with the results of conventional phenotyping and historical evidence (Moghimi, Yang, Miller, et al., 2018; Munns et al., 2006; Zhang et al., 2016). The results presented here show that the same ranking of wheat lines is obtained using six multispectral bands developed in this study (Figure 3.7). Bands were selected from the $\mathcal{R}$ set, which contains the bands in the order of their importance.

For both MDPA and posterior probability, lower values indicate a higher tolerance to salt stress. Use of only the top two multispectral bands in $\mathcal{R}$ set (i.e., 528 nm and 805 nm) resulted in the same ranking of wheat lines as was achieved using more multispectral bands or the complete hyperspectral dataset. However, using fewer bands may lead to more uncertainty because as the number of bands decreases, the values deviate more from the base values obtained by using the hyperspectral bands (Figure 3.7). To quantify the difference between the values achieved by using various multispectral band combinations with base values obtained by all bands, NRMSE was computed for both techniques (Table 3.5). NRMSE could also be used to compare the performance of MDPA and Bayesian inference in ranking wheat lines as the number of multispectral bands decreased, since NRMSE calculated for both techniques has the same scale due to normalization based on the range of base values for each technique.

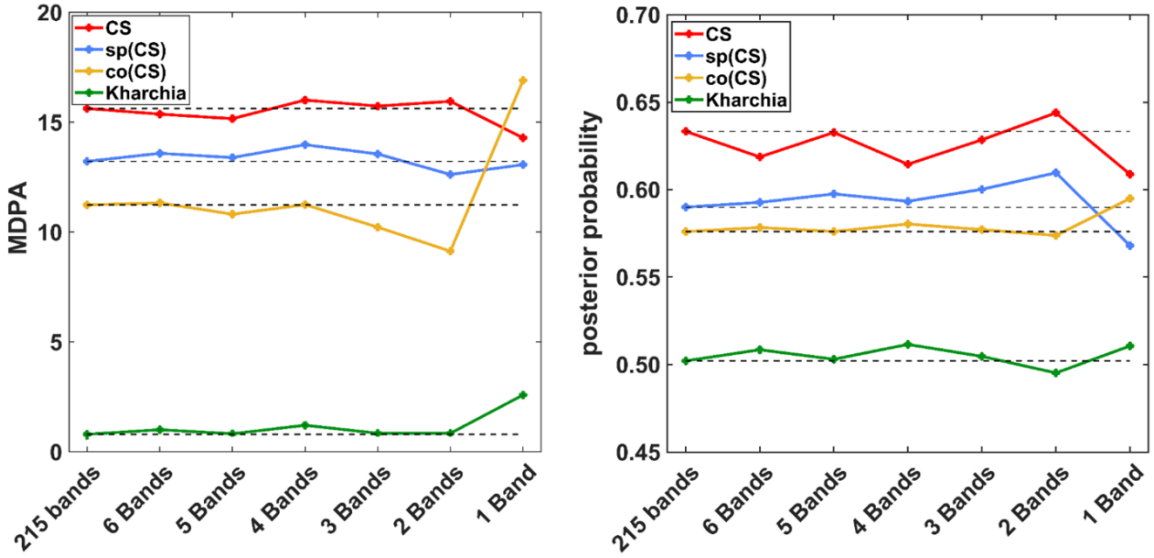

Figure 3.7. Performance comparison between the selected multispectral bands and the full hyperspectral dataset (reported by (Moghimi, Yang, Miller, et al., 2018) for ranking wheat lines based on their salt stress tolerance using (a) Minimum difference of pair assignments (MDPA) and (b) the Posterior probability P(salt|x). For both MDPA and posterior probability, lower values indicate a higher tolerance to salt stress.

Table 3.5. Normalized root mean square error (NRMSE) of minimum difference of pair assignments (MDPA) and Bayesian inference approach calculated for the selected six multispectral bands.

	NRMSE (%)					
	6 Bands	5 Bands	4 Bands	3 Bands	2 Bands	1 Band
MDPA	1.71	2.20	3.18	3.63	7.46	20.55
Bayesian inference	6.23	2.90	8.29	4.42	8.94	14.83

By decreasing the number of multispectral bands, NRMSE fluctuated for the Bayesian inference approach, whereas it consistently increased for the MDPA approach. By using only the first three bands in $\mathcal{R}$ (528, 805, and 589 nm), NRMSE for MDPA was less than 4%, indicating a multispectral camera with three broad bands centered at these bands would be sufficient for salt stress phenotyping.

3.5 Discussion

The present study proposed a pipeline to ensemble multiple rankers to rank spectral features of a high-dimensional hyperspectral dataset. The proposed pipeline offered a novel ensemble of base rankers to aggregate the benefits associated with the base rankers while disregarding those whose presence reduced the performance of classification. The algorithm was able to determine the inclusion/exclusion of a ranker for the final ensemble based on two criteria: feature subset size and F1-mean.

The ensemble feature selection drastically reduced the dimension of the hyperspectral dataset (from 215 to 15 spectral features). While 200 features were eliminated, the F1-mean improved by 8.5% from 68.49% to 77.09% using only 15 features of the feature ranking list obtained by the ReliefF, SVM-RFE, and random forest rankers. This reveals that a small subset of spectral features are able to capture a large portion of the most valuable information, while the remaining features tend to be redundant and/or contribute to noise. Moreover, ranking 589 nm, a known *Na* absorption band, as the top-ranked feature indicated that the proposed ensemble method was able to discover a physical meaning concealed in the high-dimensional spectral data of NaCl-treated wheat plants under salt stress.

In addition to the ensemble feature selection pipeline, the other contribution of this study was finding the centers of broad multispectral bands and ordering them based on their importance in discriminating salt stress leaves from healthy ones. This discovery can be a foundation for developing a custom-designed multispectral camera that aims to leverage the advantages of a multispectral versus hyperspectral sensing (e.g., simplicity, reducing cost, avoiding complications associated with line scanning hyperspectral cameras, etc.). Furthermore, by using the multispectral bands identified in this study, the salt stress tolerance of four wheat lines was assessed in a more interpretable and efficient manner. The ranking of wheat lines using the multispectral dataset was consistent with the analysis of the hyperspectral dataset evaluated by (Moghimi, Yang, Miller, et al., 2018).

In this study, we initiated the idea of establishing a library for spectral features correlated to the plant diseases and stresses, each of which adversely affecting the plant physiology. Consequently, the spectral response of plant would change based on the type of disease and stress. This underscores the importance of hyperspectral imaging as a promising research tool to find meaningful wavelengths for each type of applications at the first place. Once important features are identified, then multispectral sensors can be designed accordingly for that specific application. In general, the benefits of identifying appropriate multispectral bands to reduce the dimension of a hyperspectral dataset can be summarized as (i) reducing the complexity of interpretation as complexity is a function of input data dimensions, (ii) reducing the risk of overfitting by selecting a less complex model, (iii) reducing the running time, computational cost, and storage required for processing and storing large hyperspectral datasets, (iv) reducing the required number of samples/pixels, and (v) reducing the cost of data collection (e.g. a multispectral sensor versus a hyperspectral sensor). Furthermore, a simpler model with less redundant/irrelevant features is more robust.

3.6 Conclusion

The proposed ensemble feature selection pipeline ranked the spectral features of a hyperspectral dataset from the most to the least important based on the accuracy of a classifier in distinguishing between healthy and stressed vegetation pixels. The top 15 features in the feature ranking list achieved the best classification results, and the top two features were able to achieve a similar CV error as that of all 215 spectral features. The findings of this study suggest that *feature selection* should be embedded as one of the main pre-processing steps in hyperspectral image analysis. Feature selection mitigates the challenges in post-processing (e.g., complex interpretation and expensive computational cost) while improving classification accuracy. We were able to determine six multispectral bands that can substantially contribute in the development of a custom-designed multispectral camera for wheat salt stress phenotyping. A multispectral camera designed for a specific application can benefit scientists, farmers, agricultural practitioners, and field consultants across a wide range of use cases. Some benefits include avoiding: (i) the high cost of hyperspectral sensors, (ii) difficulties with data collection, especially as it relates to properly using a line-scanning camera, and (iii) the required expertise for analyzing high-dimensional hyperspectral images. The proposed feature selection pipeline can be utilized for feature selection and, subsequently, dimensionality reduction of other types of high-dimensional datasets (e.g., microarray datasets). In future work, we will employ this feature selection pipeline to select informative spectral features in phenotyping for other abiotic/biotic stress tolerance in plants, and work to extend the library of spectral features associated with various plant diseases and stresses.

3.7 Funding

This work was supported by the Minnesota's Discovery, Research, and InnoVation Economy (MnDRIVE) program through the research area of Robotics, Sensors, and Advanced Manufacturing.

CHAPTER 4

Selecting informative spectral bands using machine learning techniques to detect *Fusarium* head blight in wheat

4.1 Introduction

Fusarium head blight (FHB, also known as scab), is among the most widespread and devastating disease, attacking small grain crops throughout the world. FHB is caused largely by the fungal pathogen *Fusarium graminearum*, which can substantially reduce the yield and grain quality of wheat and other small grain cereals, and subsequently, have a significant impact on agriculture, economy, and society. Moreover, the FHB pathogens produce mycotoxins, such as deoxynivalenol (DON), that contaminate the grain, making it unsuitable for human and livestock consumption (O'Donnell, Kistler, Cigelnik, Ploetz, & Taylor, 1998).

FHB is very difficult to manage completely. It is best to use an integrated strategy that includes management practices (e.g., tillage and crop rotation), the application of fungicides, and the deployment of resistant cultivars. Management practices alone will not completely address the problem. With respect to fungicide treatments, they cause additional costs to producers who are already on a tight profit margin in their operations. Moreover, fungicides can have adverse impact on environment (such as contaminating soil, and surface/ground water), and some even contribute to higher mycotoxin levels in grains. Development of FHB resistant cultivars is the most effective solution in response to FHB threat (Rawat et al., 2016), and it is an essential approach to mitigate the yield and quality losses caused by FHB (Anderson et al., 2001). Consequently, many cereal improvement programs currently dedicate a significant portion of their overall effort on developing varieties resistant to FHB.

On the St. Paul campus, several research labs are currently involved in breeding wheat and barley for resistance to FHB. Each year, thousands of experimental wheat lines are planted in field plots managed by these labs. When wheat heads have fully emerged, they are inoculated with a spore suspension of the FHB pathogen. To create optimal conditions for infection throughout the nursery, an overhead irrigation system is established in the field. About three weeks after inoculation, each line in the nursery is carefully assayed for the severity (% infection) of FHB by visual inspection. This disease assessment or "disease phenotyping" is the most demanding part of breeding for disease resistance. The sheer magnitude of this task becomes evident when considering that thousands of lines are screened each growing season through visual assessments to score/rate them based on their resistance to FHB. This process is extremely demanding, laborious, and time-consuming. Additionally, it often results in inconsistent scores because individual raters may score diseased wheat lines differently.

Over the past few years, the emergence of sophisticated imaging systems mounted on unmanned aerial/ground vehicles introduced new opportunities for high-throughput phenotyping of plants. One of the non-contact sensing technologies that has drawn a substantial attention in stress phenotyping is hyperspectral imaging (HSI) which offers the ability to investigate the spectral response of stressed plans over hundreds of narrow bands with a high spatial resolution (spatial distribution pattern). HSI provides remarkable insights about internal alternations (e.g., degradation of chlorophyll, damages of leaf tissue structure, decrease of water content, etc.) in plants caused by abiotic/biotic stresses because these changes entail deviations in spectral response of the stressed plans compared to the healthy one. Several studies focused on the detection of FHB in small grains using hyperspectral imaging (Barbedo, Tibola, & Fernandes, 2015; Elke Bauriegel & Herppich, 2014; Whetton, Hassall, Waine, & Mouazen, 2018), all emphasizing the importance of identifying the infected kernels.

Despite the unique advantages/capabilities of HSI, its application has been mainly limited to research studies, and its practical application has lagged behind the theoretical and methodological developments. One of the main challenges is the complex analysis and

interpretation of hyperspectral images to appropriately recognize patterns and identify meaningful features from the high-dimensional images.

In the context of FHB phenotyping, various methods have been utilized to reduce the dimension of hyperspectral images for detection of FHB in wheat kernels. These methods include: (i) band ratio (Delwiche & Kim, 2000; Polder, Van Der Heijden, Waalwijk, & Young, 2005), (ii) feature extraction (e.g. principal component analysis, and linear discriminant analysis) (E. Bauriegel, Giebel, Geyer, Schmidt, & Herppich, 2011; Shahin & Symons, 2011), (iii) Pearson correlation coefficient (Alisaac, Behmann, Kuska, Dehne, & Mahlein, 2018), and (iv) spectral indices (e.g. normalized difference vegetation index (NDVI) and structure insensitive pigment index (SIPI)) (Alisaac et al., 2018; E. Bauriegel et al., 2011).

Delwiche & Kim, (2000) examined 231 possible band ratio between two bands ($\frac{\rho(\lambda_1)}{\rho(\lambda_2)}$; $\lambda_1 < \lambda_2$) ranging from 458 nm to 844 nm. They reported the best ratio obtained by 568nm and 715nm, which resulted in classification error between 2 and 17% to classify 32 healthy kernels from scab-damaged kernels of three varieties. In another study, Delwiche, Kim, & Dong (2011) identified two pair of wavelengths, one pair from visible region (502 nm and 678 nm) and the other from near infrared region (1198 nm and 1496 nm), have the best performance in discrimination of FHB kernels with accuracy of about 95%.

Bauriegel et al. (2011) attempted to determine the optimal wavelengths range for discrimination of FHB kernels. They concluded that four spectral range (500-533 nm, 560-675 nm, 682-733 nm, and 927-931 nm), computed based on factor loadings of PCA, are the most appropriate wavelength range for FHB detection. They also proposed an index, called head blight index, composed of two spectral range with 10 nm width (550-560 nm and 665-675 nm). The head blight index could predict the disease degree in 84% of kernels within ±10 percent of the visual rating of disease degree.

In another investigation, Shahin & Symons (2011) selected six wavelengths (484 nm, 567 nm, 684 nm, 817 nm, 900nm and 950 nm) using PCA to detect FHB infected kernels with classification accuracy of 92%.

Bauriegel and Herppich (2014) stated that two spectral regions (550-560 nm and 665-675 nm), each with 10 nm width, are needed to detect the FHB infected spikes.

More recently, Alisaac et al. (2018) conducted a study to assess the resistant of seven wheat varieties to FHB using hyperspectral imaging over 400-2500 nm. According to their results, there is a high correlation between the severity of FHB disease and four broad spectral ranges (430-525 nm, 560-710 nm, 740-810 nm, 1115-2500 nm). They also investigated the correlation between the severity of FHB disease and fifteen vegetation indices (e.g., NDVI and SIPI), concluded that there is a high correlation between them. However, Bauriegel et al. (2011) reported that the classification results obtained by using vegetation indices (e.g., NDVI and SIPI) were not satisfactory.

It can be inferred from the results of former studies that there is a discrepancy in selecting the most informative bands to detect the FHB spikes from the healthy ones. This variation can be explained by the fact that spikes experiencing FHB exhibit premature bleaching, while healthy spikes are still green. This causes the spectral response of infected spikes deviate in significant portion of visible and near infrared region from the spectral response of healthy spikes. Consequently, various feature selection methods may recognize different spectral bands or region to be the most informative for distinguishing the FHB spikes from the healthy ones. Therefore, a more robust method is needed to identify the most discriminative spectral bands to detect the infected spikes in a more dynamic field environment at the early stages of infection when the detection can be challenging.

One possible solution is to integrate several feature selection techniques to leverage the benefits associated with each of these techniques. In this study, we utilized the ensemble method proposed by Moghimi, Yang, & Marchetto (2018) with several adjustments to identify sensitive bands for FHB detection. The objective of this study was to take advantage of high spectral resolution of HSI, as a research tool, to identify the most sensitive spectral features for assessment of FHB resistance of wheat lines. The result of this study can be used to develop a custom-designed sensor tailored around FHB phenotyping in a rapid, accurate, efficient, and cost-effective manner.

4.2 Methods

4.2.1 Plant preparation

The spring wheat variety ‘Wheaton’, which is historically known as a susceptible variety to scab, was selected for selecting the most sensitive wavelengths in discriminating healthy plants from infected spikes. Wheaton seeds were planted in 10 pots (four seeds per pot), five pots as control and five pots for FHB treatment. After the heading time, the spikes in the pots designated for scab were inoculated by injecting the FHB inoculum (Figure 4.1). The process of spike inoculation was performed on six days depending on the maturity stage of the spikes. After about three weeks, the infected spikes and control spikes were cut and immediately transferred to a dark room for imaging.

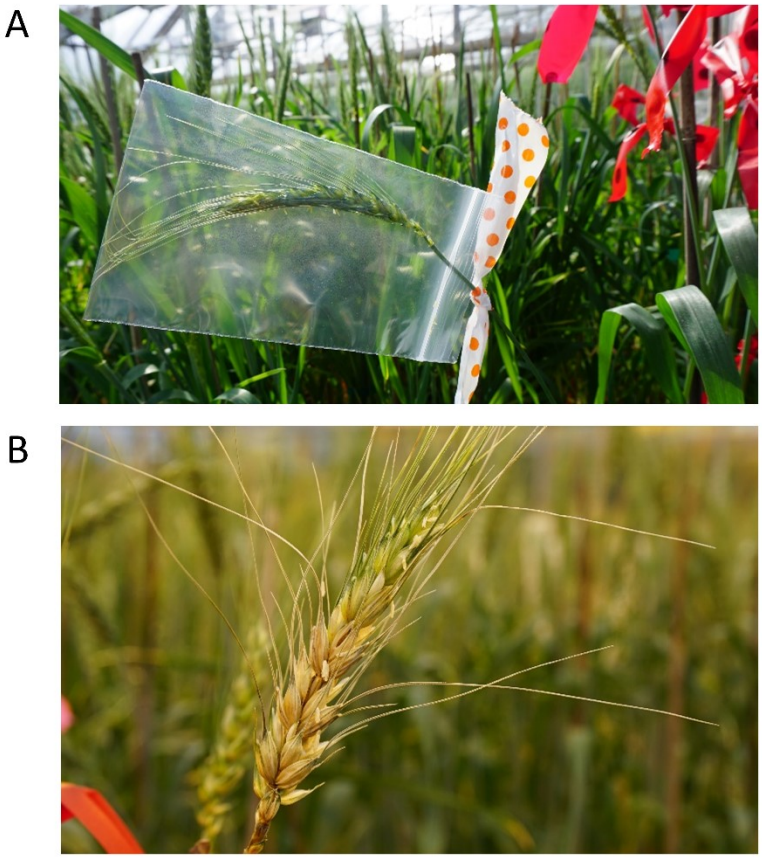

Figure 4.1. An inoculated spike. Colorful ribbons were used to keep track of the inoculation date. Plastic bag was used to provide a humid environment for the fungus. (B) An infected spike became bleached. In an infected spike, awns tend to bend.

4.2.2 Hyperspectral image acquisition

Hyperspectral images were captured using a push-broom (along-track scanner) hyperspectral camera (PIKA II, Resonon, Inc., Bozeman, MT 59715, USA) mounted on a motorized slider moving with a constant speed across a horizontal bar (Figure 4.2). The camera recorded the spectral response of the spikes over 240 spectral channels in visible and near infrared region (400 nm to 900 nm) with about 2.1 nm spectral resolution and 640 spatial channels in the cross-track direction. The scanning frame rate of the camera was set as described by (Moghimi et al., 2017) based on the field of view of the camera lens (33 degrees), the distance between the lens and spikes (about 0.8 meter), and the speed at which the camera was moving (0.025m/s). The lighting source was composed of four halogen lamps (250W), appropriately adjusted to obtain a uniform illumination. The gain and exposure time of the imager was set appropriately to avoid over-exposure while obtaining a maximum dynamic range.

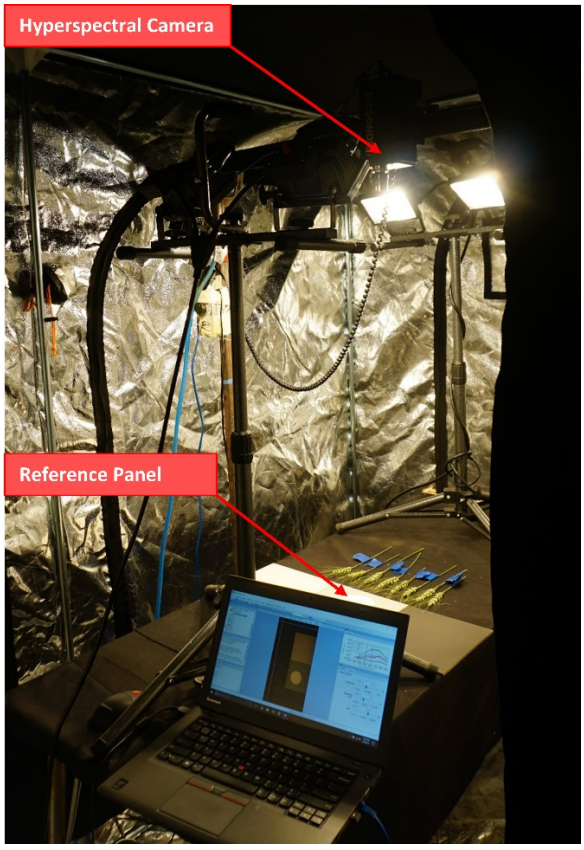

Figure 4.2. Hyperspectral imaging setup.

4.2.3 Preprocessing of hyperspectral images

Two preprocessing steps were required before image analysis steps, radiometric calibration and noisy bands removal.

The radiometric calibration step accounted for the two main inconsistent factors: sensor-dependent factors and time-dependent factors. Sensor-dependent factors include parameters related to the sensor, such as inconsistency among spatial and spectral responses of detectors in the camera (e.g., gain, offset, and quantum efficiency). To account for these sensor-dependent factors, raw images were converted to radiance ($Wm^{-2}sr^{-1}nm^{-1}$) using a vendor-supplied calibration file. Alternatively, time-dependent factors refer to the factors that tend to change over the course of imaging, such as variation in intensity of incident light over time. To account for time-dependent factors, images in radiance were converted to reflectance using a Spectralon panel (Labsphere, Inc., North Sutton, NH, USA) placed as a reference surface in all images. The second preprocessing step was noisy band removal. The first 30 bands (up to ~455 nm) were eliminated from the dataset as they were noisy.

4.2.4 Segmentation of spikes

To identify the most discriminative bands for FHB detection, pixels representing the spikelets with FHB infection and pixels representing healthy spikelets were separately segmented from background, which was a black fabric. At the imaging time, healthy spikelets were still green while the infected spikelets were bleached. To segment the healthy pixels and the infected pixels form background, excessive green index (EGI) and transformed soil-adjusted vegetation index (TSAVI) (Baret, Guyot, & Major, 1989) were used, respectively (Figure 4.3).

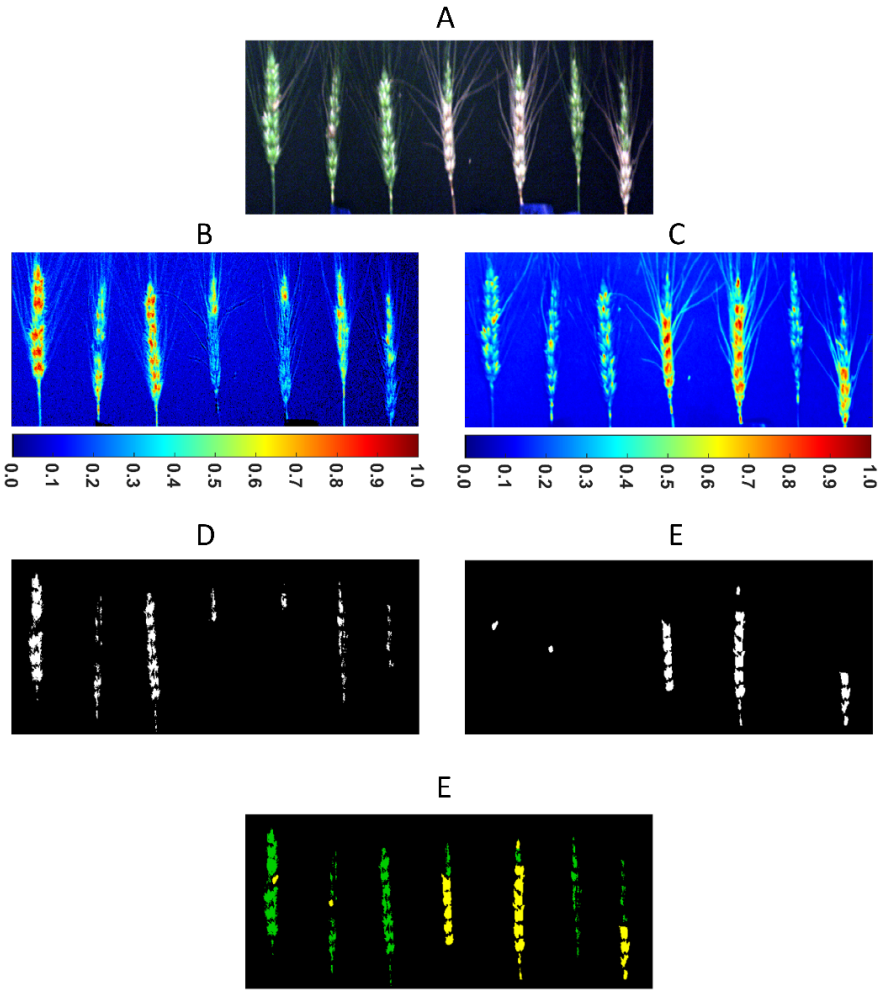

Figure 4.3. (A) RGB representation of a hyperspectral image captured from wheat spikes. (B) Excessive green index (EGI), shown in colormap, was used to segment pixels representing healthy spikelets. (C) Transformed soil-adjusted vegetation index (TSAVI), shown in colormap, was used to segment infected spikelets from background. (D) Thresholding EGI to create a mask for healthy spikelets. (D) Thresholding TSAVI to create a mask for infected spikelets. (E) Healthy and infected spikes segmented from background.

4.2.5 Feature Selection Method

Despite all of the investigations attempting to deploy HSI, still more research is needed to address the challenges in analysis of hyperspectral images. Moghimi, Yang, and Marchetto (2018) leveraged advanced machine learning algorithms for developing a robust ensemble feature selection method to identify sensitive spectral features for salt stress phenotyping in wheat. They identified the six most predominant wavelengths for salt stress assessment in wheat. Their finding suggests the importance of spectral feature selection as a pre-processing step in analysis of hyperspectral images. In the present study, to determine

sensitive wavelengths for FHB detection, their ensemble feature selection pipeline was utilized with a minor modification: instead of the minimum error, the required number of features to obtain an error within one standard error (1SE) of 0.05 was considered as one of the criteria. The reason for changing this criterion was that the cross validation error dropped for the first few feature subset size and reached a steady state for FHB data.

Subsequent to the ensemble process, two methods were used to identify six broad spectral bands around the most informative bands at the top of the ensemble feature list. The first method was the clustering algorithm proposed for the salt stress project (Algorithm 2 in Moghimi et al. (2018)). The second method was a nonparametric method for probability density estimation, called *kernel estimator* also known as *Parzen windows* as follows:

$$\hat{p}(x) = \frac{1}{Nh} \sum_{t=1}^{N} K(\frac{x - x^t}{h}) \tag{4.1}$$

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{u^2}{2}] \tag{4.2}$$

where $\hat{p}(\cdot)$ is the estimator of probability density, $K(\cdot)$ is the Gaussian kernel function, N is the number of spectral features, h is the band interval which was set to 10 nm in the present study.

4.3 Results

4.3.1 Band pair correlation

In hyperspectral imaging, hundreds of spectral bands are scanned. However, a great portion of the scanned bands might be highly correlated. Figure 4.4 depicts a colormap representing the correlation between all pairs of bands. It appeared that there is a high correlation between near infrared bands. This could be anticipated because the spectral response of both control and scab pixels changed very slightly from about 750 nm to 90 nm (Figure 4.5). Alternatively, there is a low correlation between near infrared bands and red bands, shown as dark blue in Figure 4.4. Nevertheless, there is a general pattern in band pair correlation: nearby bands tend to highly correlate with each other.

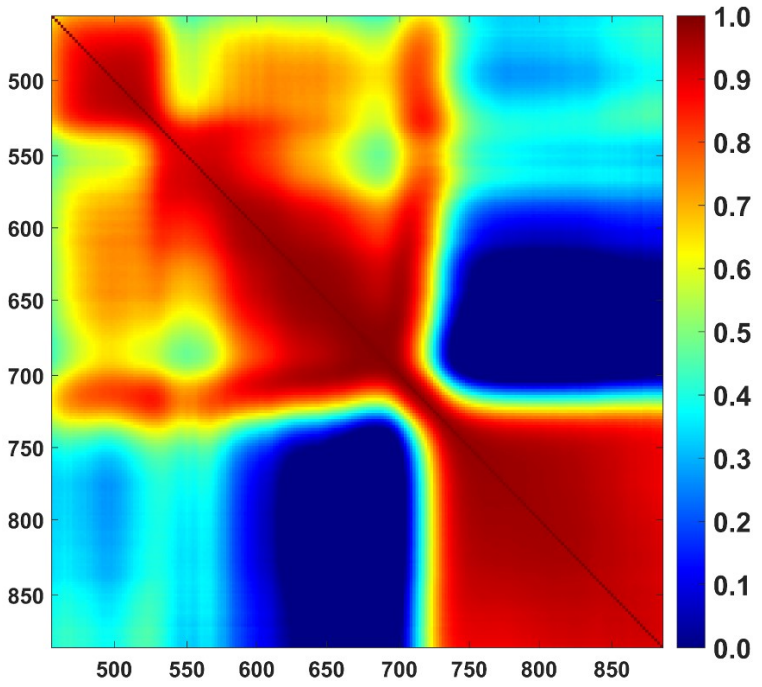

Figure 4.4.Correlation between all pairs of spectral features presented as a colormap image.

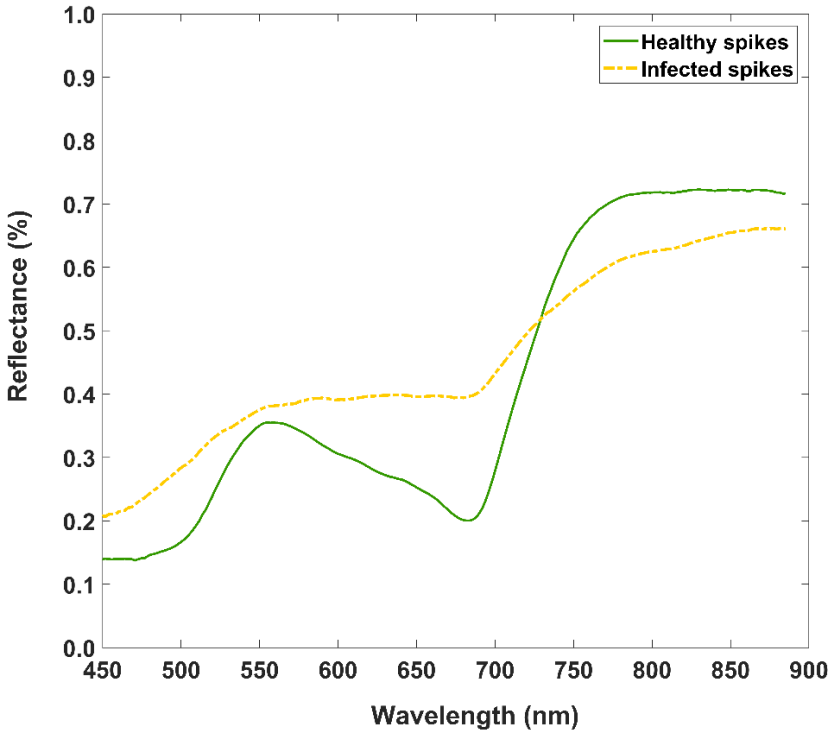

Figure 4.5. Average spectral response from healthy and infected spikes.

4.3.2 Ranking of spectral features by six rankers

Similar to feature selection for salt stress study, feature subset size and F1-mean were computed for each ranker to compare the performance of rankers with each other (Table 4.1). Among the six rankers, CFC and random forest required a significant large number of features to obtain an error within 1SE of the defined error (5% error). On the other hand, SFS-forward, LASSO, and SVM-RFE required only two features to meet the requirement. Although random forest required a large feature subset size, it had the lowest CV error on the training dataset and the maximum F1-mean on the test dataset. However, QDA classifier (i.e., quadratic discriminant analysis) used only the first two features ranked by SFS-forward, LASSO to obtain a similar F1-mean as random forest.

Table 4.1. Classification performance of QDA on the training and test datasets for individual rankers.

	Training dataset with 10-fold CV			Test dataset		
	Feature subset size	Error (%)	StD (%)	$F1_{C0}$ (%)	$F1_{C1}$ (%)	F1-mean (%)
CFS	25	4.65	0.46	95.86	96.13	95.99
ReliefF	4	2.53	0.33	97.42	97.77	97.59
SFS-forward	**2**	0.73	**0.14**	99.09	99.23	99.16
SVM-RFE	**2**	2.72	0.46	96.85	97.26	97.05
LASSO	**2**	0.68	0.19	99.06	99.21	99.13
Random Forest	11	**0.65**	0.16	**99.12**	**99.25**	**99.18**

Table 4.2 presents the first 11 top-ranked features per each ranker. Intriguingly, four rankers belonging to wrapper and embedded categories identified their most informative spectral feature around 760±12 nm. Among these four rankers, random forest required the largest feature size (i.e., 11 spectral features) because the first 10 features were within 20 nm interval (775±10 nm). As shown in Figure 4.4, these 10 spectral features were highly correlated. Therefore, once a spectral feature (686 nm) from outside of this range was added to the feature subset size as the 11^{th} feature, the error dropped to lower than 5%.

Table 4.2. The first 11 top-ranked features per each rankers to classify healthy and infected pixels.

Rankers	First 11 top-ranked spectral features (nm)										
CFS	686	684	688	682	680	690	678	692	676	674	672
ReliefF	489	497	466	770	471	475	499	709	477	458	700
SFS-forward	772	696	752	594	709	764	869	682	754	549	538
SVM-RFE	748	602	659	703	594	620	631	680	690	641	725
LASSO	760	698	709	764	750	766	756	754	752	707	746
Random Forest	772	774	778	776	768	785	783	781	766	770	686

4.3.3 Ensemble of feature selection techniques

The ensemble feature selection method developed for the salt stress project was a recursive elimination process in which one ranker was eliminated at each step provided that its removal resulted in the best performance based on two criteria (F1-mean and feature subset size) among all other potential choices for elimination in that step. However, the recursive elimination method was not used for this scab project because feature subset size was identical for three base rankers (SFS-forward, SVM-RFE, and LASSO), and there was a subtle difference between their F1-mean (Table 4.1). Therefore, the complete linear aggregation method was used to average the ranking assigned to each spectral feature by these three rankers. Afterwards, two clustering methods was used to determine the center of six broad spectral bands. Figure 4.6 demonstrates the first 30 top-ranked spectral features from the ensemble list. Noticeably, three regions are dominant, including spectral bands between 680 to 720 nm, 750 to 800 nm, and between 850 nm to 900 nm. The first interval (680 to 720 nm) refers to the red-edge region and the other two denote the near infrared regions of electromagnetic spectrum. Kernel estimator was able to identify one cluster center per each of these three intervals as well as two other cluster centers with lower probability density (Figure 4.7). The cluster centers identified by Kernel estimator and the clustering algorithm proposed in the salt stress study are presented in Table 4.3. The results revealed that both methods could find similar cluster centers as the center of broad multispectral bands. The difference was the number of cluster centers - the clustering algorithm was required to identify six clusters as the number of clusters was set to six, whereas, Kernel estimator is a nonparametric method and could find five clusters with high probability density.

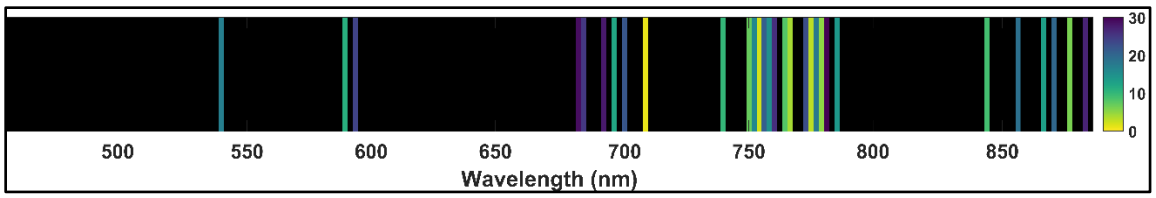

Figure 4.6. The first 30 top-ranked spectral features from the ensemble list mainly located in three regions: 680 to 720 nm, 750 to 800 nm, and between 850 nm to 900 nm.

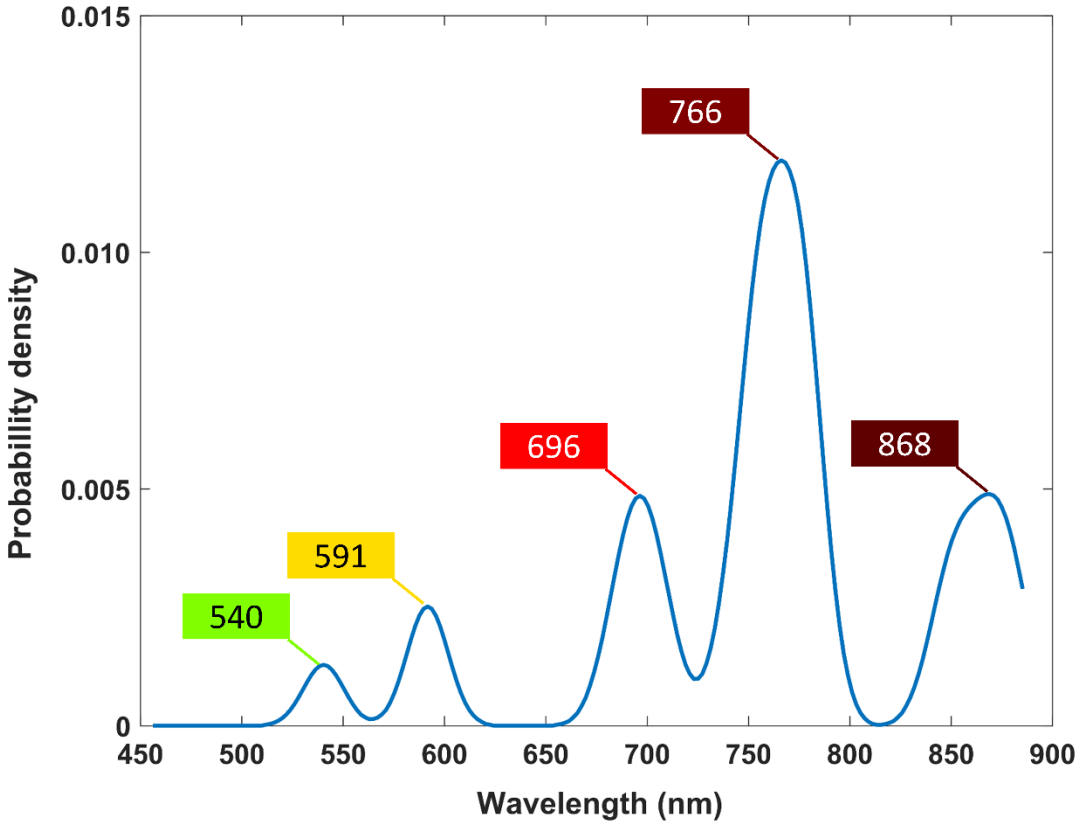

Figure 4.7. Probability density estimation calculated by a Kernel estimator with Gaussian kernel function. Five clusters were generated as the center of broad spectral bands.

Table 4.3. The cluster centers identified by Kernel estimator and the clustering algorithm proposed by (Moghimi, Yang, & Marchetto, 2018)

	Center of clusters (nm)					
Kernel estimator	766	868	696	591	540	-
Clustering algorithm	696	764	874	851	591	540

4.4 Discussion

Each year, wheat growers struggle with major yield losses, costly management practices, and discounts on their delivered crop, all caused by scab disease. One promising solution is to cultivate a resistant wheat variety to scab. Breeders conduct experiments in multiple locations for several years to identify resistant wheat varieties through screening thousands of wheat lines in fields. The process of rating the resistance of each wheat line is performed by visual assessment of spikes, which is prohibitive given the labor, time, and cost requirements. Moreover, visual assessment is subject to human error and bias of raters who score the resistance of wheat lines. FHB is certainly a devastating plant disease that needs an improved phenotyping method for assessment.

In this study, the optimal set of spectral bands appropriate for discriminating control and scab pixels were determined by an ensemble feature selection method. A very high classification accuracy (~99%) was achieved by using two spectral features: 766 nm and 696 nm. This finding evidently demonstrates the redundancy of hyperspectral bands. However, it should be noted that these results were achieved in the laboratory with different conditions than the field environment where FHB screening is performed in a large scale with uncontrolled lighting condition and a more complex background. Therefore, it is critical to find the most informative spectral features to develop a custom designed multispectral camera tailored to recognize FHB infected spikes in field conditions. Based on the results, a regular RGB camera is sufficient for classifying infected spikes from healthy ones since it captures images in red and green regions which are among the informative bands (696 nm and 540 nm). However, incorporating NIR bands might improve the performance of spike recognition from background as well as discriminating healthy and scab infected spikes.

CHAPTER 5

A deep neural network for analysis of aerial hyperspectral imagery to predict the yield of experimental wheat plots

5.1 Introduction

Considering the increasing world population and subsequent demand for food, crop production should double by 2050 (Tilman et al., 2011), indicating the average rate of yield increase of crops should be 2.4% annually (Ray et al., 2013) – the current average rate of increase is only 1.3% (Jose Luis Araus et al., 2014; Ray et al., 2013). These statistics noticeably indicate an urgent need for further efficiency improvement in crop production to alleviate the global concern of food security. Nevertheless, genetic gain in yield of wheat, one of the major crops, was reported to be less than 1%, far behind the necessary yield increase (i.e., 2.4%) (Crain et al., 2018; Ray et al., 2013; Reynolds et al., 2012). Other studies even claimed wheat yields have plateaued in some regions of the world (Acreche et al., 2008; José Luis Araus et al., 2018; Sadras & Lawson, 2011), indicating the importance of phenotyping for developing wheat varieties with high yield potential.

To identify crop varieties with high yield potential, plant scientists and breeders evaluate the performance of hundreds of lines developed through breeding and genotyping. This evaluation is performed through conventional phenotyping methods which rely on demanding, extremely laborious, and time-consuming tasks to measure various crop traits. For instance, in experimental yield nurseries of wheat, the grains of each plot should be individually harvested by a combine harvester followed by packaging, labeling, and sealing to avoid blending grains of plots. These exhausting tasks should be performed at multiple locations in a rather short harvesting time since breeders have yield nurseries at various regions to account for non-uniform climate, soil, and environmental conditions.

Furthermore, conventional yield phenotyping is restricted by the availability of machinery, labor, and weather conditions. Each of these factors could potentially postpone harvesting time for several days during which yield loss can occur because of animals' attack (e.g. birds and rodents) and/or severe weather (e.g. hail and winds). Any of these challenges could deteriorate the quality and validation of the data, therefore, wasting the enormous efforts made thorough growing season.

The other limitation associated with conventional methods is that it ignores the spatial variability of yield within the plots. Various regions in a plot contribute unequally to the measured yield for the plot (i.e., yield is non-uniformly distributed within an experimental plot). Therefore, breeders are unable to study the effect of crop density on yield potential of various varieties. Moreover, ignoring the variability of yield within plots entails an enormous loss of information regarding the marginal effects on yield to identify the lines whose plants located in the middle of plot can compete for nutrition and therefore can contribute to yield as much as the plants located at the margin of the plot. Considering the importance of selecting high-yielding varieties and limitations associated with conventional phenotyping methods, there is a compelling need to predict yield, preferably with high-resolution.

Precise prediction of yield has always been of particular interest worldwide to agricultural communities as it provides insights into making the best site-specific management decisions at the right time and right place. A substantial number of studies have been performed to predict the crop yield using various platforms and sensors.

For large-scale yield prediction (e.g., county, state, and nationwide), satellite imagery has been widely used as a decision support tool for researchers and policy-makers (Aase & Siddoway, 1981; Benedetti & Rossini, 1993; Chaparro et al., 2018; Labus, Nielsen, Lawrence, Engel, & Long, 2002; Lobell, 2013). However, satellite imagery is not an appropriate tool for plant phenotyping which requires high temporal and spatial resolutions given the size of plots and the objective of the problem. Jin et al. (2017) used HJ-1A/B and RADARSAT-2 imaging data as the input variables for AquaCrop model to estimate biomass and canopy cover of winter wheat, and subsequently, predict the yield of 30 field

experiments. They reported that the estimated biomass had a better agreement with the yield (R^2 = 0.42 and RMSE = 0.81 ton/ha) compared to the yield prediction based on the estimated canopy cover (R^2 = 0.31 and RMSE = 0.94 ton/ha).

A decade ago, flexible and maneuverable platforms to carry sensors in experimental field trials comprising hundreds of varieties were considered the bottleneck in plant phenotyping. However, it is not a challenge anymore with the emergence of unmanned aerial vehicles (UAVs) into the agricultural domain and advances in autonomous ground vehicle. On the other hand, image-based remote sensing technologies with higher spectral and spatial resolution has become more lightweight and available at lower cost, and subsequently suitable for mounting on UAVs. In high-throughput field phenotyping, UAVs have been implemented to study several traits of various crops: plant height in maize (Han et al., 2018), drought-adaptive traits in wheat (Condorelli et al., 2018), seasonal growth and development of barley (Burkart, Hecht, Kraska, & Rascher, 2017), plant density of wheat (Jin, Liu, Baret, Hemerlé, & Comar, 2017), yield estimation of soybean (Yu et al., 2016), plant height of sorghum and maize (Shi et al., 2016).

In high-throughput field phenotyping for yield estimation of wheat, researchers have utilized UAVs equipped with various sensors. Madec et al. (2017) attempted to predict the yield of various wheat genotypes based on maximum plant height estimation using RGB images and LiDAR data collected by a UAV. They reported a low correlation between yield and maximum plant height derived from LiDAR data (R^2 = 0.22) and RGB images (R^2 = 0.13). Duan, Chapman, Guo, & Zheng (2017) computed normalized difference vegetation index (NDVI) derived from multispectral images captured by UAV to predict the yield of wheat for high-throughput purposes. Because of the attained low spatial resolution (2-5 cm), the NDVI calculated per each pixel was a combination of vegetation and background, inherently with different spectral characteristics. To address the mixed pixel issue, they proposed a naïve (ineffective) solution in which pixels with NDVI less than a predefined threshold were masked for further analysis. They suggested that there is a high correlation (R^2 = 0.87) between the adjusted NDVI, computed around flowering time, and final yield. However, this finding was achieved from a few number of plots (in

total 12 plots including three cultivars with four treatments). To predict the yield of a particular winter wheat, Du & Noguchi (2017) deployed stepwise regression to analyze five color vegetation indices derived from multi-temporal color images captured by a UAV from heading stage to ripening stage. They performed the analysis on only nine samples of wheat yield. Their results demonstrated a strong correlation ($R^2 = 0.94$ and RMSE = 0.02) between four color vegetation indices and yield for this limited number of samples. In another study, aerial images acquired from UAV were utilized to estimate the yield of twenty wheat varieties under a water limited and heat stressed environment (Kyratzis, Skarlatos, Menexes, Vamvakousis, & Katsiotis, 2017). Two vegetation indices including green normalized difference vegetation index (GNDVI) and NDVI were calculated at various growing stages of plants over two consecutive years. They concluded that GNDVI, compared to NDVI, performed better in explaining variability of grain yield with $R^2 = 0.31$ and $R^2 = 0.21$ for the first and second year of experiment, respectively.

Nowadays, with the commercialization of UAVs and increasing availability of more compact, inexpensive, and sophisticated sensing technologies, the challenge shifted from data collection to data analysis - extracting significant features and recognizing underlying patterns from large datasets captured by autonomous platforms equipped with sophisticated non-contact sensing technologies. The most common approach for analysis of image-based data (RGB, muti- or hyper-spectral images) is to calculate spectral vegetation indices derived by simple arithmetic equation (e.g. ratio) among few spectral bands. Nevertheless, more advanced analysis methods are required to extract valuable information from images for high-throughput field phenotyping rather than simple vegetation indices which entail several limitations. For instance, it has been proved that NVDI, the most widely used index, suffers from saturation issue over vegetation canopy with moderate-to-high level of density (Asrar, Fuchs, Kanemasu, & Hatfield, 1984; Gitelson, 2004; Gitelson, Kaufman, & Merzlyak, 1994).

More advanced techniques are required for high-throughput analysis of large image-based phenotyping datasets. Recently, machine learning and deep learning algorithms have shown considerable promise in developing more efficient and effective pipelines for

analysis of large phenotyping datasets (A. Singh et al., 2016; A. K. Singh et al., 2018). Among a variety of algorithms, techniques, such as convolutional neural network, capable of analyzing large image-based datasets have drawn a substantial attention since imaging has been shown to be a promising component in high-throughput phenotyping.

The primary objective of this study was to develop a sensor-based, automated framework for high-throughput yield phenotyping of wheat in the field. The data from hundreds of wheat varieties were collected by a hyperspectral camera mounted on a UAV flying over three experimental wheat plots during two consecutive growing seasons. To analyze high-dimensional hyperspectral images captured with high spatial and spectral resolution, a deep neural network was trained to predict the yield of wheat plots. In addition to yield prediction at plot scale, the feasibility of yield estimation at a finer spatial resolution (i.e., sub-plot scale) as well as a coarser spatial resolution (i.e., field scale) was investigated in this study.

5.2 Materials and methods

5.2.1 Field site and experimental setup

Field experiments were conducted at three experimental yield trials (C3, C4, and C9) during two consecutive growing seasons 2017 (C3 and C9) and 2018 (C4). Field sites were located at St. Paul Campus Research Facility, University of Minnesota, MN (44°59'28.15"N and 93°10'48.34"W) (Figure 5.1). Yield trials were composed of hundreds of new wheat lines, several check lines, and advanced lines from other breeding programs. Each wheat line was planted as a plot with about 1 meter (3.5 ft) width and 2.4 meter (8 ft) length. The plots were harvested with a combine designed for harvesting small plots. After harvest, the grains were weighted for each wheat plot. Therefore, the unit of yield was gram per plot. Since the area of the plots was the same, yield is presented in terms of gram hereinafter.

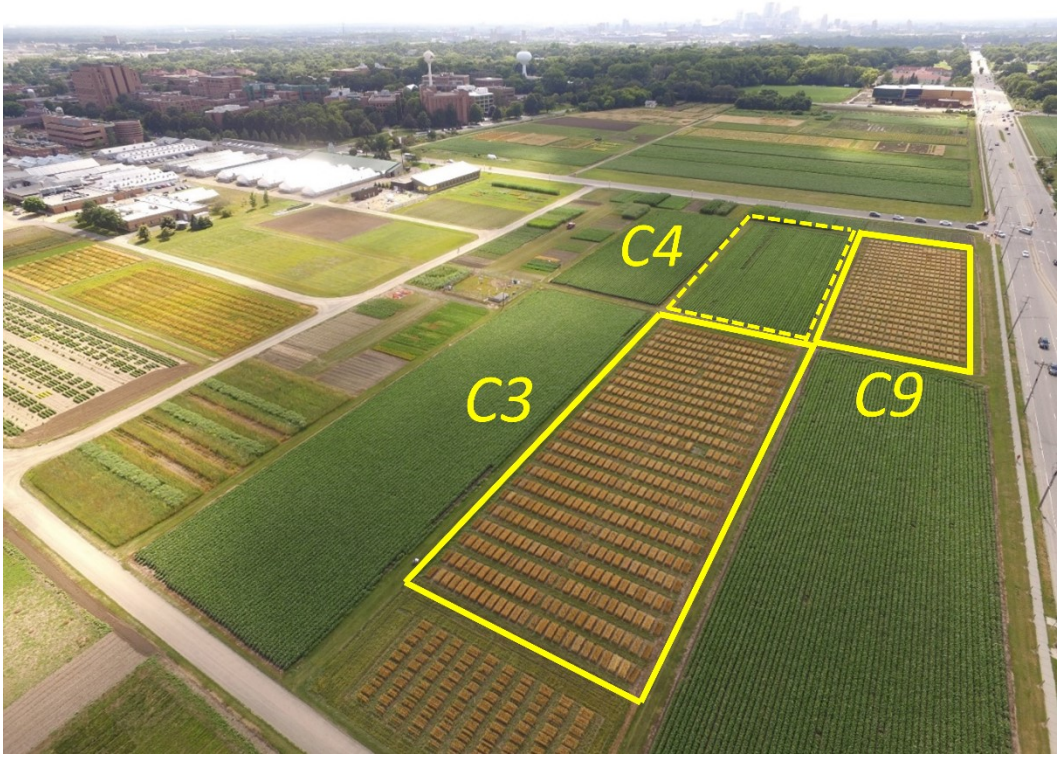

Figure 5.1. Layout of yield plots in 2017 experimental wheat trials (C3 and C9). Crop rotation was done in 2018 and C4 was planted with wheat.

5.2.2 Platform for aerial imagery

The UAV used in this study was DJI Matrice 600 Pro equipped with A3 Pro flight controller. The flight missions were created and executed in a grid mode with DJI Ground Station Pro (Figure 5.2). Table 5.1 presents the detail of the flight mission. For image collection, the entire mission was executed in autonomous mode except the take-off and landing which were performed manually. On the same day of image collection, a manual flight was performed at very low altitude (~ 5 meter) to collect images for endmembers extraction, described in section 5.2.6.1.

A gimbal (DJI Ronin-MX) was used to carry the airborne hyperspectral imaging components and maintain automatically the camera at nadir position regardless of the UAV movements (Figure 5.2B).

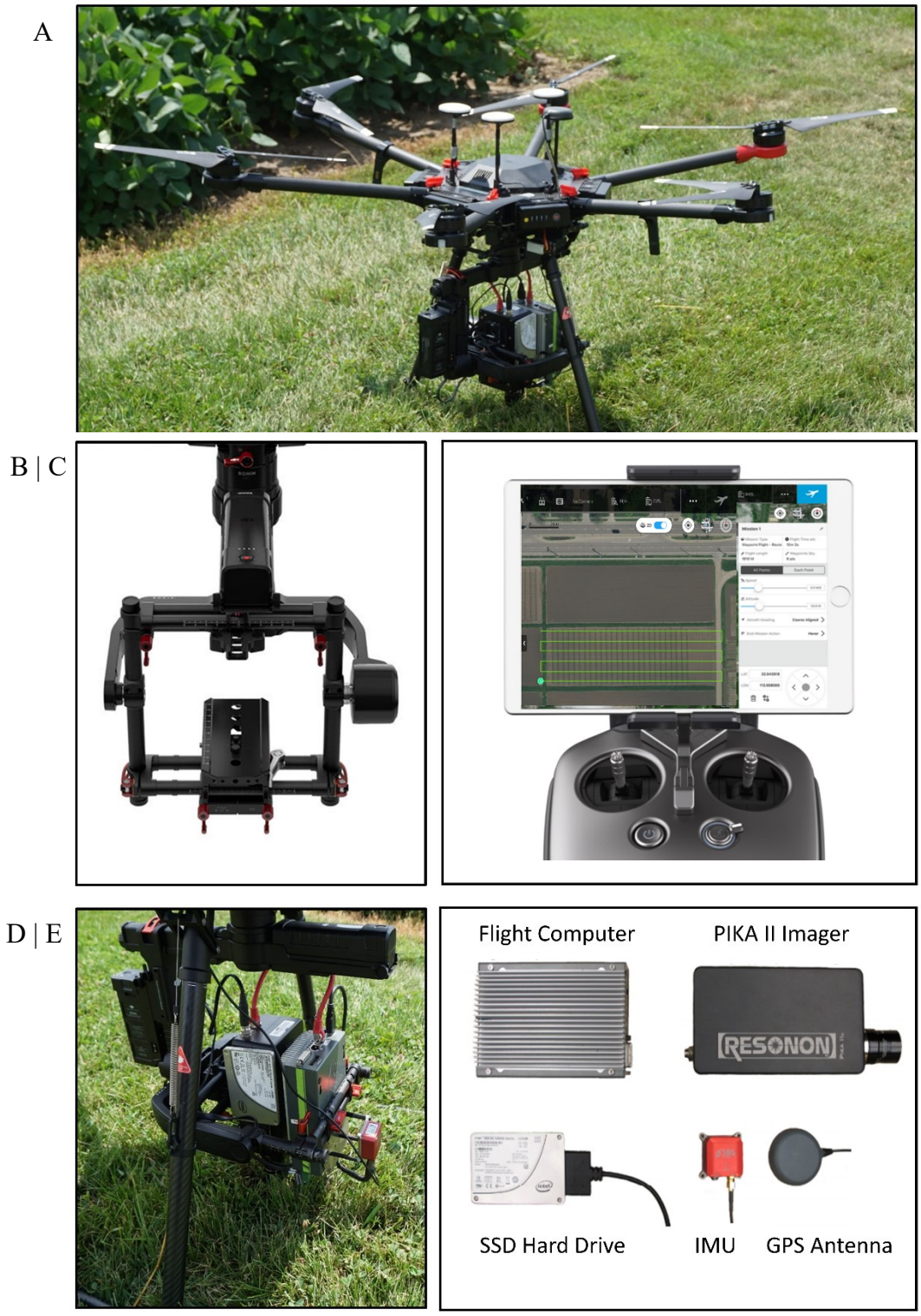

Figure 5.2. (A) Unmanned aerial vehicle: DJI Matrice 600 Pro equipped with A3 Pro flight controller. (B) Gimbal: DJI Ronin-MX. (C) DJI Ground Station Pro. (D) Airborne hyperspectral imaging system mounted on the gimbal. (E) Components of airborne hyperspectral imaging system.

Table 5.1. Flight information for two missions.

Flight mission for:	Flight mode	Altitude (m)	Speed (m/s)	Silelap	Spatial resolution (cm)
Yield prediction	Autonomous	20	2	50%	< 2
Endmember extraction	Manual	~ 5	0.5	-	< 0.5

5.2.3 Airborne hyperspectral imaging setup

The camera used in this study was a push-broom hyperspectral camera (PIKA II, Resonon, Inc., Bozeman, MT 59715, USA) with the specifications presented in Table 5.2. The components of airborne hyperspectral imaging system include the imager, flight computer, GPS antenna, inertial measurement unit (IMU), and a solid-state hard disk (Figure 5.2E).

To have the imager automatically start and stop collecting data, the aircraft should be above a defined minimum altitude (8 meters for yield prediction mission and zero for endmember extraction mission) and the GPS of imaging setup should be inside a predefined target area. To define the area of interest, a polygon was created with appropriate buffer around the experimental field in Google Earth, saved as KML file, and then uploaded into the flight computer. The camera started image collection when the GPS of imaging setup entered the polygon above the predefined threshold, and it stopped the imagery once the aircraft exits from the polygon.

Image acquisition was performed in auto expose mode in which gain and exposure time were automatically adjusted based on the ambient lighting conditions and the brightness of the target. The other parameter that should be adjusted by user was the frame rate of scanning denoting the number of pixel lines scanned at each second perpendicular to the direction of movement.

Frame rate of scanning was a function of two user-defined parameters including aircraft speed and flight altitude as well as two specifications of imager including field of view (33 degree) and spatial channels. In this study, a low flight altitude (20 meter) was defined to attain a high spatial resolution while avoiding the potential turbulence over canopy caused

by the propellers of UAV. The speed of aircraft was set 2 m/s to cover the entire field in one flight while the required scanning frame rate to maintain the spatial integrity (square pixels with aspect ratio of 1:1 in cross and across track) to be less than the maximum frame rate of imager. Once the flight altitude and speed were set, a frame rate of 108 frame per second was calculated as described by (Moghimi et al., 2017).

The hyperspectral pixel lines captured by PIKA II were transferred to the flight computer via an Ethernet cable, synchronized by GPS and IMU data, broken into about 2000 lines, and then saved as a hyperspectral image cube to the hard drive through a USB-3 connection. Therefore, the pixel size of each hyperspectral image cube was 2000×640×240, requiring about 640 megabytes space for saving.

Table 5.2. Specifications of the PIKA II hyperspectral camera

Hyperspectral imager	Spectral range (nm)	Spectral resolution (nm)	Spectral channels	Spatial channels	Maximum frame rate (frame per second)	Bit depth
PIKA II	400 – 900	2.1	240	640	145	12

5.2.4 Pre-processing of hyperspectral images

5.2.4.1 Radiometric Calibration

The hyperspectral images were collected as raw digital numbers (DNs) which is the least useful format with no units or physical meaning. Therefore, raw images were converted to radiance ($Wm^{-2}sr^{-1}nm^{-1}$) using the lab-derived radiometric calibration file provided by the manufacturer of imager. This conversion is a key step required in radiometric calibration of hyperspectral images to compensate the non-uniform spectral and spatial responses of the instrument (Moghimi, Yang, Miller, et al., 2018).

To account for potential variation in solar illumination, hyperspectral images in radiance were then converted to reflectance using reference panels (60×60 cm) placed in the field before image collection. The panels were pained with gray paint mixed with Barium Sulfate to diffuse the incoming solar irradiance to various directions (no specular

reflection). In a laboratory setup, the actual reflectance of gray panels were measured by a ASD FieldSpec 4 spectroradiometer (Analytical Spectral Devices, Inc., Longmont, CO, USA) with respect to the reflection of a Spectralon panel (Labsphere, Inc., North Sutton, NH, USA) used as a standard reference panel with highly Lambertian surface. Radiance and reflectance conversion were performed using SpectrononPro software (Resonon, Inc., Bozeman, MT 59715, USA). The gray panels were placed north side of black panels in alleys. The location of panels was calculated based on image swath to maximize the probability of capturing at least one set of reference panel in each image. The unique ID of the plots located at both side of gray panels were recorded in an inventory for further processing to recognize the ID of all plots across the image.

5.2.4.2 Noisy Band Removal

Prior to any further analysis, the first and last few bands were disregarded because of high noise (any bands before 430 nm and after 870 nm). In addition, spectral bands near the absorption region of O_2 and H_2O were removed from the hyperspectral data cube (Moghimi, Yang, Miller, et al., 2018). In total, 190 spectral bands out of 240 bands were kept for further analyses.

5.2.5 Plot segmentation and identification

5.2.5.1 Segmentation of plots from background

When aerial images were collected, wheat plots were at the senescence stage. While chlorophyll a and chlorophyll b in a green, healthy leaf of a wheat plant require absorbing a high extent of light at blue and red region of electromagnetic spectrum for photosynthesis, a senescent leaf tends to absorb less light at these two regions, this is because of a significant decline in chlorophyll content (C. Lu, Lu, Zhang, & Kuang, 2001). However, the extent of enhancement in reflection from senescence leaves of wheat at red region is higher than the reflection at blue region. The reason for this change in reflectance pattern is that carotenoid, with a high absorption at blue region (Lichtenthaler, 1987), is much less affected compared to chlorophyll a and b during the leaf senescence, meaning the illuminated light is still highly absorbed at blue region during senescence (Biswal, 1995;

Grover, Sabat, & Mohanty, 1986). Therefore, to segment wheat plots from background, a vegetation index referred to as *normalized difference plant senescence index* (NDPSI) was proposed in this study. NDPSI is essentially a vegetation index derived from two broad bands: red (670±5 nm) and blue (450±5), as follows:

$$NDPSI = \frac{Red - Blue}{Red + Blue} = \frac{\frac{1}{n}\sum_{i=665}^{675}\rho_i - \frac{1}{m}\sum_{j=445}^{455}\rho_j}{\frac{1}{n}\sum_{i=665}^{675}\rho_i + \frac{1}{m}\sum_{j=445}^{455}\rho_j} \quad (5.1)$$

where ρ denotes reflectance at particular wavelength, n and m refer to the number of bands used to generate broad red and blue spectral bands ($n = m = 5$), respectively. While single band at 670 and 450 nm can be also used to calculate NDPSI, consolidating five bands as broad red and blue bands rendered a NDSPI gray-scale image with effectively reduced salt and pepper noise.

Pixels representing wheat plots displayed a tendency to exhibit large values of NDPSI compared to the background pixels, which were mainly reference panels, green winter wheat planted in alleys, soil, and shadow caused by plants. A threshold was defined for pixel values of NDPSI to segment wheat plots from the background (Figure 5.3). Several morphological operations were applied on this binary image. First, a flood-fill operation was conducted to fill the holes generated inside the objects. The second mathematical morphology was opening operation, an erosion followed by a dilation with a rectangle structuring element (10×5), to remove small objects and break the potential connection between adjacent plots due to the lodging of plants. To assure small objects are disregarded, a threshold was defined for the area of the objects in terms of pixels. The obtained binary mask was then used to segment the plots and fit bounding boxes enclosing the plots (Figure 5.3).

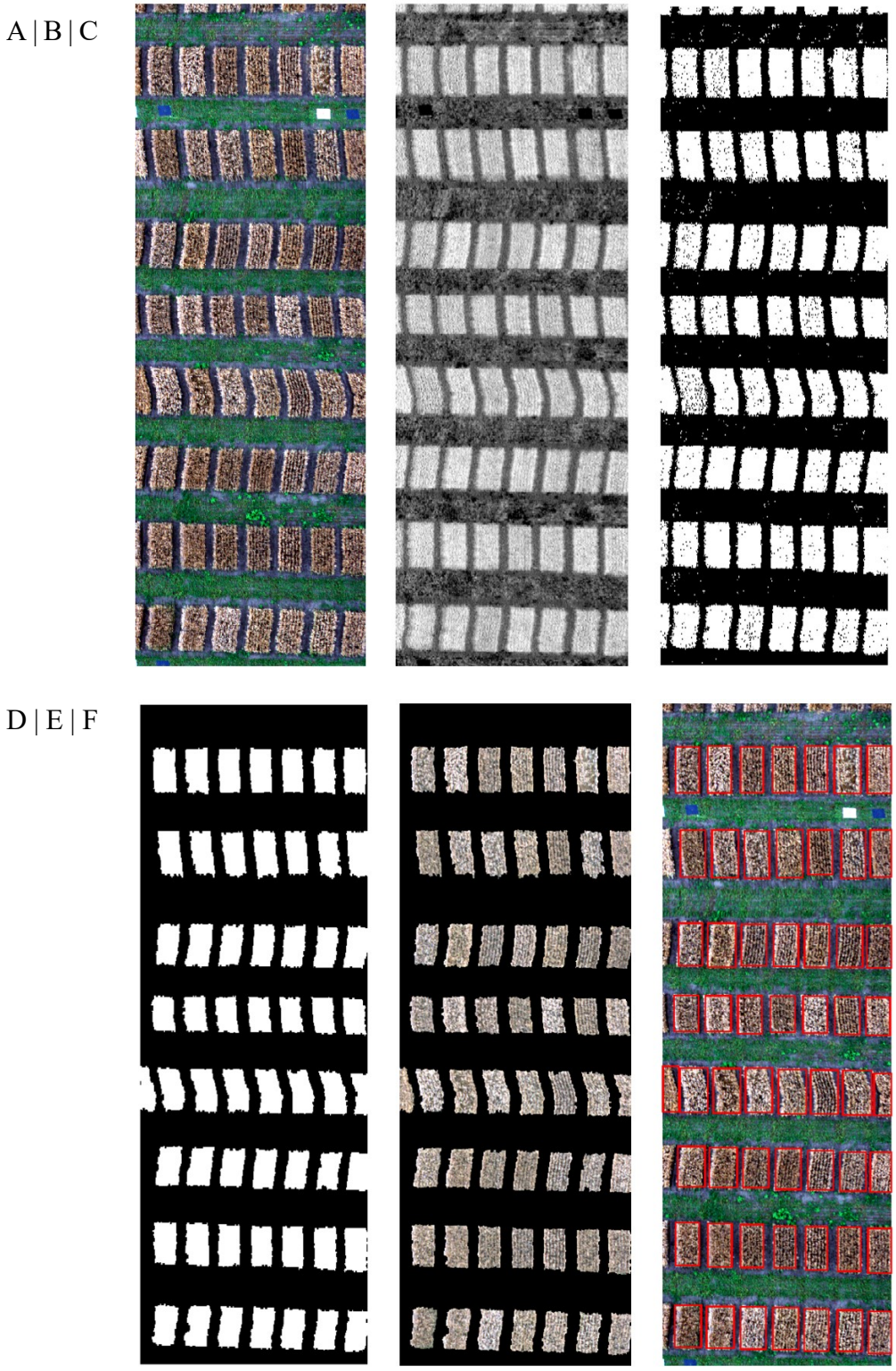

Figure 5.3. (A) RGB representation of a hyperspectral image. (B) Gray scale image of normalized difference plant senescence index (NDPSI). (C) Binary image obtained by thresholding NDPSI. (D) Binary mask obtained by morphological operations including flood-fill, opening, and area thresholding. (E) RGB representation of hyperspectral image of plots segmented from background using the binary mask. (F) Fitting bounding boxes enclosing the segmented wheat plots.

5.2.5.2 Recognizing plots ID

The geo-rectification process of hyperspectral images failed largely because the IMU data was not accurate enough due to the magnetic interference. Therefore, a semi-automatic pipeline was developed to identify the plot ID of segmented plots in each image.

Manual processing

GPS/IMU data was utilized to generate an approximate swath outline (i.e., image boundary) per each image. These swath outlines were saved as KML files and imported into QGIS to have an estimate for the geographical position of the field area scanned in each hyperspectral image. To create a detailed basemap with reference features, such as location of reference panels, an orthomosaic was generated using high-resolution RGB images orthorectified and stitched together using Pix4Dmapper. These images were captured at low altitude (7 meter) by DJI Inspire UAV equipped with a double 4K sensor (Sentera, Inc., Minneapolis, MN 55423, USA). Using the orthomosaic as the basemap for the swath outlines assisted in identification of at least two plots per image since the plot ID of the plots located at both side of gray panels was previously recorded.

Automatic processing

The automatic processing part was mostly based on image processing techniques. Pixels representing the top-left corner of bounding boxes in a single hyperspectral image were clustered based on row- and column-wise pixel distance from each other. The number of clusters obtained by row- and column-wise clustering denoted the number rows and columns of plots exist in a given image, respectively. Afterwards, a grid was created from horizontal and vertical lines obtained by taking the average over the row as well as column arrays of the pixels (i.e., top-left corner of bounding boxes) grouped in one cluster.

Given the fact that the ID of two plots per each image are previously identified through the manual processing, the procedure of assigning a plot ID to each grid cell can be automatically propagated across the image using plot ID matrix. Theoretically, each cell of the grid should entail a wheat plot with unique ID, and potentially a bounding box

encompassing the plot. However, a few plots were not detected by the segmentation algorithm since they could not pass the thresholds defined in morphological operations. If a bounding box exists in a given cell gird, the algorithm allocates the ID assigned to the cell to the bounding box. If there is no bounding box in a given cell, then the algorithm skips to the next grid cell.

Subsequent to plot segmentation and identification, plots were cropped from the hyperspectral images using fitted bounding boxes and saved as 3-D matrices ($x \times y \times \lambda$) to preserve the spatial ($x \times y$) and spectral (λ) integrity of plots for further workflow. These 3-D matrices will be referred to as plots hyperspectral cube (P-HSC) hereinafter.

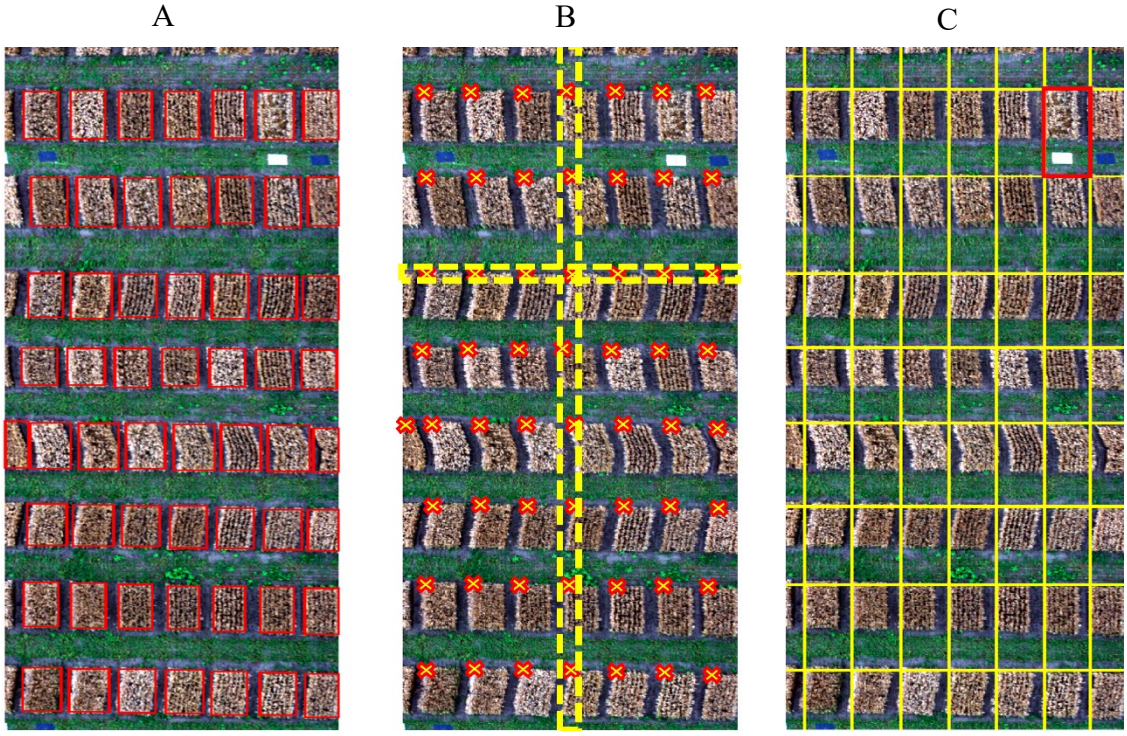

Figure 5.4. (A) Bounding boxes enclosing the wheat plots. (B) Top-left corner of bounding boxes are clustered based on row- and column-wise pixel distance. It is shown how top-left corners are clustered together in a row and a column. In this example, there are eight rows and eights columns of clusters (C) horizontal and vertical lines of the grids were generated by taking average over the row array of pixels group in a row cluster and column array of the pixels grouped in a column cluster, respectively. Based on the plot ID of the plot next to the gray panel (red box), a plot ID was assigned to the grid cells, and subsequently to the bounding boxes inside them.

5.2.6 Hyperspectral image analysis

5.2.6.1 Endmember selection

Despite the high-spatial resolution (~2 cm) attained by flying at low altitude (20 m), each pixel of a plot might exhibit spectral characteristics of a mixed pixel, largely due to properties of the objects of interest (spikes and leaves) such as size, angle, and curvature. Within a plot image, there were mainly four classes, including spikes and leaves, two classes of our interest, as well as soil and shadow, two classes considered as background. Therefore, the spectral response of a plot pixel can be composed of these four classes, each contributing with various extent and with a distinct spectral signature. To distinguish the abundance of these four classes in each pixel, the pure spectral signature, called endmember, should be identified per each of the classes. Nevertheless, it should be noted that the notion of endmember existence in the form of perfectly pure pixel is for conceptual convenience because of uncertainty caused by sensor noise and spectral signature variability within a class (Schowengerdt, 2012). In practice, each pixel is essentially a mixed pixel to a certain extent in remote sensing. Therefore, the most pure pixels in the scene with the most distinct spectral response are often considered as endmembers. One of the widely-used techniques to identify the endmembers is N-FINDR algorithm, in which n endmembers are selected as the n vertices of a $(n-1)$-simplex with a maximum volume encompassing the majority of pixels in the feature space spanned by all pixels (Winter, 1999, 2004). However, N-FINDR algorithm suffers from issues such as long processing time, and inconsistency in selecting the final set of endmembers due to the random initial endmember selection. Various automated techniques, all inspired by N-FINDR, were proposed to ameliorate the process of endmember extraction (T. H. Chan et al., 2011; Chang, Wu, & Tsai, 2011; Zortea & Plaza, 2009).

In the present study, successive volume maximization (SVMAX), proposed by Chan et al. (2011), was utilized to identify the endmembers through a successive optimization problem.

In hyperspectral images collected at 20 meter AGL, it was impossible to find a pixel that contains only a spike because of the spike geometry from the sensor perspective. To obtain a sufficient resolution for capturing spikes, a low altitude flight (5 meter AGL) was performed in which the resolution of hyperspectral images was approximately 0.5 cm. This flight mission was executed in manual mode because DJI Ground Station Pro prohibited autonomous flight at such a low altitude.

There were six endmembers in both hyperspectral image datasets collected at 5- and 20-meter AGL, each representing one of the six distinct classes, including spikes, wheat leaves, soil, shadow, winter wheat, and gray panel.

5.2.6.2 Spectral Mixture Analysis

Once the endmembers were identified from the images captured at low altitude, each pixel of a P-HSC can be represented as a convex combination of the endmembers. Since P-HSCs contained mainly the four classes (spikes, wheat leaves, soil, and shadow), their endmembers were used for the un-mixing process. In this study, to determine the fractional abundance of the endmembers in the pixels of P-HSCs, a matrix factorization problem with two constraints was defined as per Thurau, Kersting, Bauckhage, Iais, & Augustin (2010) in which a Frobenius norm is minimized as follows:

$$min \|X - WH\|_F$$
$$s.t. \begin{cases} 1^T.\ h_j = 1 \\ 0 \leq h_{ij} \leq 1 \end{cases} \quad (5.2)$$

Where $X(d \times N)$ is the matrix of data obtained by reshaping a P-HSC (i.e., 3D matrix) to a 2-dimenionsal matrix such that pixels (N: number of pixels) were extracted in column-wise order and were placed as the columns of matrix X, and bands (d: number of bands) were placed as the rows. $W(d \times e)$ and $H(e \times N)$ are the endmembers matrix (e: number of endmembers), and the abundance matrix, respectively. Each columns (h_j) of matrix H was calculated by resolving a quadratic optimization problem (Moghimi, Yang, Miller, et al., 2018) iteratively N times with similar constraints as follows:

$$\min_{h_j} \quad \frac{1}{2}\, h_j^T Q h_j + \, c^T h_j \,, \qquad j = 1, \dots, N$$
$$s.t. \begin{cases} 1^T .\, h_j = 1 \\ 0 \le h_{ij} \le 1 \end{cases} \tag{5.3}$$

where

$$Q = 2W^T W$$
$$c = \, -2W^T x_j \tag{5.4}$$

5.2.6.3 Sub-plot image analysis

The distribution of the measured yield for a plot was not homogeneous over the plot because of the factors such as spatial variability of soil, available nutrient, and marginal effects. While studying the yield variation within a plot can provide valuable insights into the breeding for selecting advanced wheat lines, harvesting the wheat grains at sub-plot resolution in a large yield trial is a tedious, unrealistic, and impractical task. One of the objective of this study was to leverage the high spectral and spatial resolution of aerial hyperspectral images in order to examine the yield variation within a plot.

Each plot was divided into square sub-plots (15×15 pixel). To assure P-HSC can be divided into 15×15 grids, zero-padding was applied at the margins of P-HSC, meaning each pixel can be fitted in a 15×15 grid. Once a plot was divided into sub-plots, a yield should be assigned to each sub-plot. For this purpose, we hypothesized that the yield of a sub-plot is proportion to the number of spikes and leaves (SL) pixels which are representing the above-ground biomass in the subplot (i.e., subplots with higher density of spikes and leaves contribute more in plot yield). To count the number of SL pixels within each sub-plot, sub-plot pixels were classified into two classes: SL class or soil-shadow (SS) class. A given pixel was classified to SL class if the summation of abundance for spikes and leaves endmembers in that pixel was more than 0.5; otherwise, it was assigned to SS class (background). Afterwards, if the measured yield of a plot was y, a normalized yield value y_i was assigned to the ith sub-plot as follows:

$$y_i = \frac{n_i}{N} \times y \qquad j = 1, \dots, m \tag{5.5}$$

$$N = \sum_{i=1}^{m} n_i \tag{5.6}$$

where m is the number of sub-plots in the plot, N is the total number of SL pixels in the plot, and n_i is the number of SL pixels in ith sub-plot. Since the yield assigned to a sub-plot was normalized based on the total number of SL pixels in the plot, the summation of yield for all sub-plots in the plot was equal to the measured yield of the plot ($\sum_{i=1}^{m} y_i = y$).

5.2.6.4 Extracting input features from sub-plots

Each sub-plot was composed of several SL pixels segmented from SS pixels. These SL pixels were considered as one object per each sub-plot window. Object-based image analysis (OBIA) approach was then used to leverage extracting features (such as size, area, texture, mean and standard deviation per band) associated with a set of pixels as oppose to per-pixel analysis (Blaschke, 2010).

In the present study, mean and standard deviation (std) per band (in total 190 bands) were extracted as input features because they provide adequate information to estimate the distribution of pixels' reflectance per band per each subplot. The other input features extracted from sub-plots was the area of the SL object in terms of pixels (i.e., the number of SL pixels). This referes the number of samples used to calculate the mean and std of the distribution. In total, the number of input features per sub-plot was 381 (190+190+1).

5.2.6.5 Dataset

There were three sets of data each from an adjacent field C3 and C9 collected in 2017, and C4 collected in 2018. After removing the damaged plots, a set of 50 plots was selected for test dataset using stratified sampling to assure test dataset has an akin yield distribution to the training and validation datasets. The number of plots selected for test dataset from each field was proportion to the total number of training samples (i.e., number of sub-plots)

belonging to that field. The sub-plots of these 50 plots were held out as the test dataset for an unbiased evaluation of the final trained model (Figure 5.5). Subsequent to selecting the test dataset, other plots of the three fields were divided into sub-plots and merged together to form a dataset for training and validation of the model. Using stratified sampling, these sub-plots were split into training (90%) and validation (10%) datasets to train and validate the model during the training process (Table 5.3).

In another experiment, the dataset of each field was separately divided into training, validation, and test datasets with a similar approach described above to develop an individual model per each field.

After splitting the data, the training dataset was normalized to make each feature have zero-mean and unit-variance. Subsequently, validation and test datasets were standardized using the mean and variance obtained from training dataset.

Table 5.3. Number of plots and sub-plots in each field and size of training, validation, and test datasets

Year	Field	Number of plots	Number of sub-plots	Training	Validation	Test
2017	C3	422	19287	14491	2557	2239
	C9	345	19650	14343	2531	2776
2018	C4	254	12773	8726	1540	2507
All fields		1021	51710	44,261	4,919	2,530

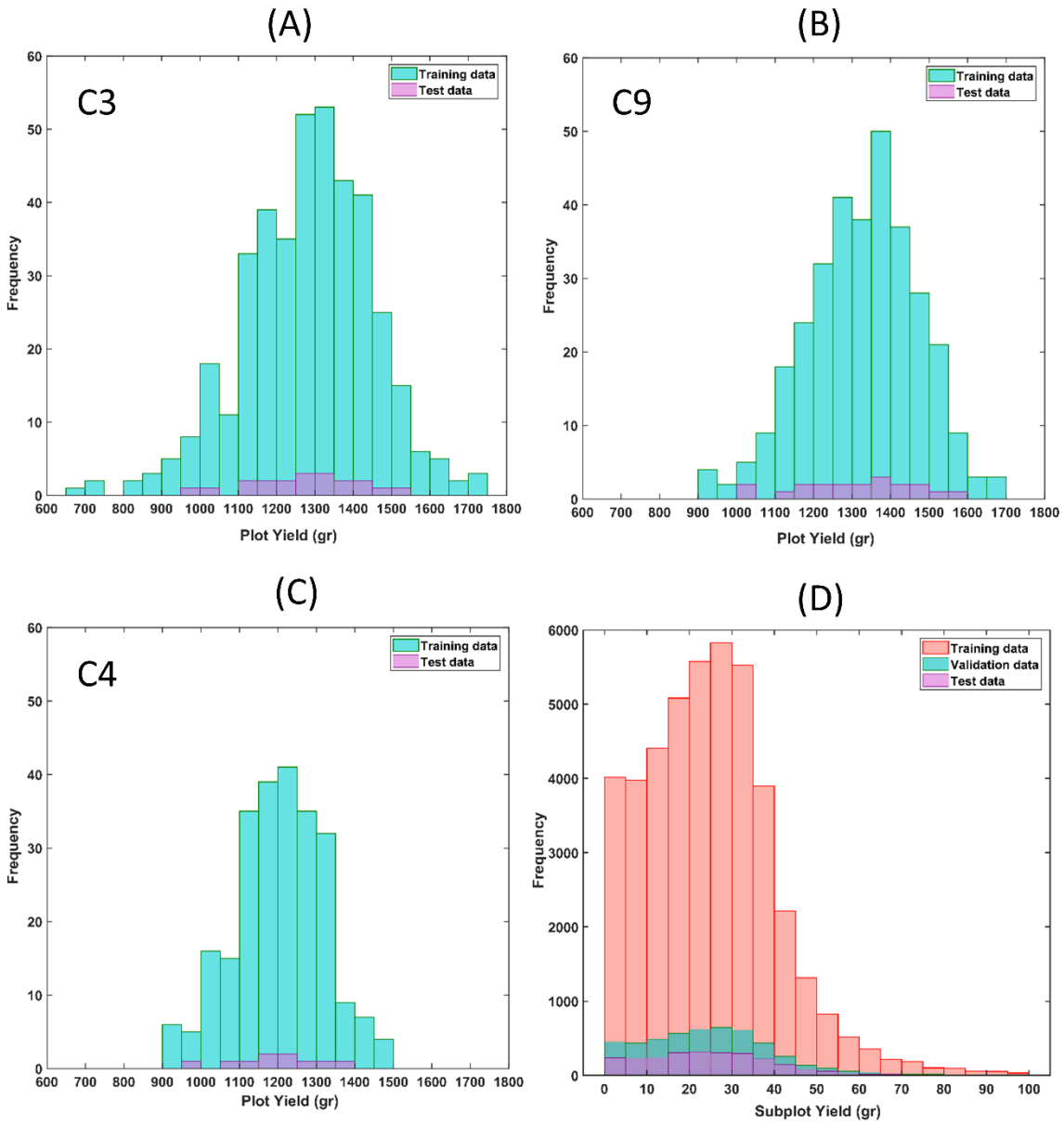

Figure 5.5. Yield histogram of the plots used for training and test per each field: 20 plots from C3 (A), 20 plots from C9 (B), and 10 plots from C4 (C) were selected for test dataset. (D) Yield histogram of the sub-plots used for training, validation, and test datasets.

5.2.6.6 Deep Neural Network

Deep learning (DL), inspired by the biological neural structure in the human brain, refers to computational non-linear models composed of various processing layers in which an abstract representation from the output of the previous layer is learned up to the output layer where a complex functions is learned in terms of these abstract representations (Lecun, Bengio, & Hinton, 2015).

Among various type of architectures, convolutional neural network (CNN) (Krizhevsky, Sutskever, & Hinton, 2012; Y LeCun et al., 1989; Yann LeCun et al., 1990) is well suited for data with spatial structure. However, the spatial information within sub-plots was lost since the yield assigned to the sub-plots was based on the number of SL pixels, regardless of the spatial location of SL pixels with respect to each other in the sub-plot window. Consequently, a vector of features for each sub-plot was considered as the input layer for a deep neural network (DNN) with fully connected layers in preference to CNN. In this study, the network was a feedforward neural network, also known as multilayer perceptron (MLP) (Goodfellow, Bengio, & Courville, 2016), composed of an input and output layer as well as four hidden layers.

The input layer represented the input features, thus it had 381 units (190+190+1). The output layer was a single unit representing the predicted yield (Figure 5.6). The number of hidden layers and their units were two important hyper-parameters of the network defined through an empirical process in which the performance of various network architectures, selected based on the domain knowledge, were evaluated. Since a large portion of wavelengths scanned by the hyperspectral camera are redundant or irrelevant to the desired phenotyping trait (Moghimi, Yang, & Marchetto, 2018), the number of units in the hidden layer was selected among a set of small numbers relative to the input layer. Alternatively, the number of hidden layers was limited by the size of the training dataset since an additional hidden layer increased the required number of data to train the model parameters (weights and biases).

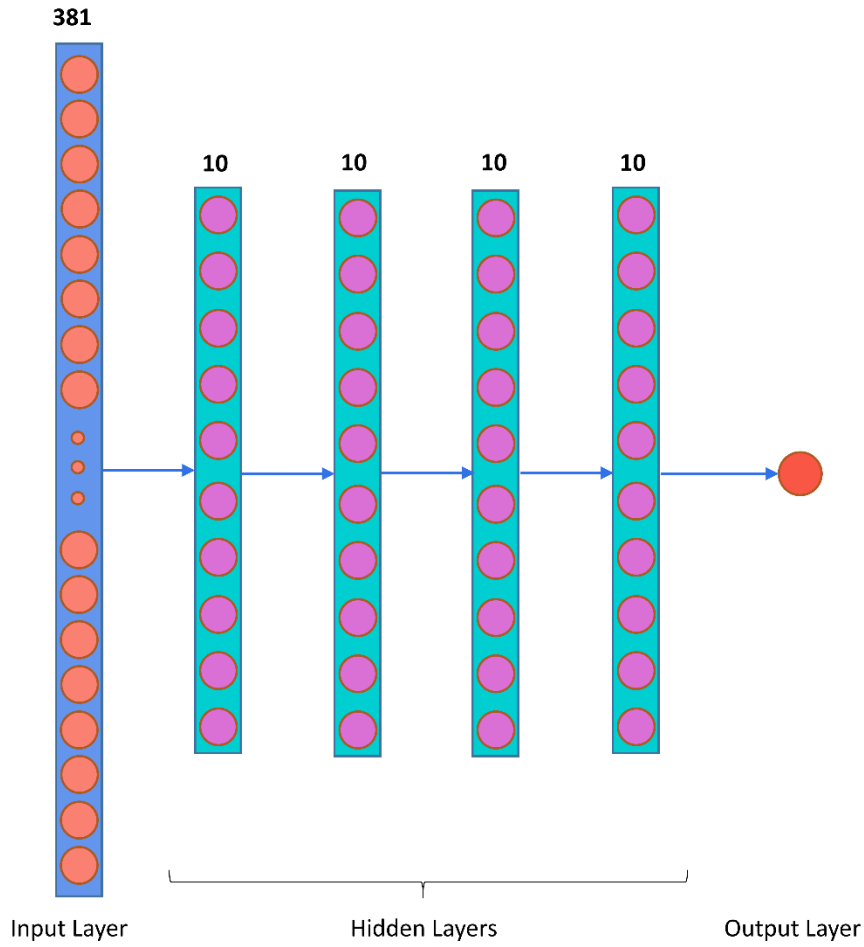

Figure 5.6. The architecture of deep neural network with fully connected layers.

Layers were fully connected, meaning each unit in layer l was connected via weighted linkage to the all units in the layer $l+1$. Therefore, the input of unit j in layer $l+1$ (z_j^{l+1}) was the weighted sum of the output of all units in layer l (a_i^l) plus a bias term (b_j), and the output of the unit j in layer $l+1$, denoted by a_j^{l+1}, were calculated as follows:

$$z_j^{l+1} = \sum_{i=1}^{d} w_{ij}^l \, a_i^l + b_j^l \tag{5.7}$$

$$a_j^{l+1} = f(z_j^{l+1}) \tag{5.8}$$

where d is the number of units in layer l, w_{ij}^{l} refers to the weights of links connecting all the units in layer l into unit j in layer $l+1$, and $f(.)$ was the activation function to introduce non-linearity into the network. The transfer function was rectified linear unit (ReLU) $f(x) = \max(0, x)$ (Glorot, Bordes, & Bengio, 2011; Jarrett, Kavukcuoglu, Ranzato, & LeCun, 2009; Nair & Hinton, 2010), a non-linear function that allows the network learn faster and avoids saturation for large positive inputs.

The weights and biases were adjusted through an iterative process using training and evaluation datasets. The training began with initial random values as per Glorot & Bengio, (2010) for weights and biases. In the forward phase of an iteration, weights and biases were used to calculate the network output using equations (5.7) and (5.8). In the backward phase, weights and biases were updated based on the network error. The cost function used to measure the network error was mean squared error (MSE) which is the squared difference between the output of the network (i.e., the predicted yield for a sub-plot) and the desired value (i.e., the yield value assigned to the sub-plot), averaged across all training samples.

During the training process, an optimization algorithm was used to identify the network parameters for which the cost function was minimized (returned a low value for the error). The optimization algorithm was Adam (Kingma & Ba, 2014), a computationally efficient optimization algorithm with adaptive learning rate. Adam embraces the benefits of adaptive gradient algorithm (Duchi, Hazan, & Singer, 2011) and root mean square propagation (Tieleman & Hinton, 2012), which are suitable for sparse gradients and non-stationary setting, respectively (Kingma & Ba, 2014). The number of epochs[1] was set to 100.

5.2.7 Computational environment

The DNN model was developed and tested in Keras 2.2.2 (Chollet & others, 2015) with TensorFlow 1.9.0 (Abadi et al., 2015) backend running on an NVIDIA (GeForce GTX 750 Ti) GPU. All other computations and image analysis were performed by MATLAB R2017b (MathWorks, Inc., Natick, MA, USA).

1. The number of times that the entire dataset is passed through the network to find the optimized values for parameters of the network.

5.3 Results

5.3.1 Endmember extraction

The number of endmembers in SVMAX was set to six, as there were six distinct classes in the image collected at 5-meter altitude. Pixels identified by SVMAX as the endmembers were the vertices of a simplex with the maximum volume compared to any other possible simplex formed by pixels in the feature space spanned by all pixels. To account for uncertainty caused by factors such as sensor noise, the reflectance of pixels within a specified Euclidean distance of the identified endmembers were averaged as the new set of endmembers. For visualization of the endmembers location with respect to the other pixels, all pixels were projected onto a 2- and 3-dimensional feature space, respectively spanned by the first two and three principal components (PC) obtained by principal components analysis (Figure 5.7). It should be noted that the location of endmembers might not be the vertices in the new feature space because of projection onto a lower dimension. For instance, in a 2-dimensional feature space, the endmembers of gray panel, winter wheat, and shadow were the vertices of a triangle while the endmembers of spike, leaves, and soil were placed in the established triangle. Alternatively, in a 3-dimensional feature space spanned by the first three PCs, a different set of endmembers might be the vertices depending on the viewing angle (Figure 5.7).

Figure 5.8 illustrates the spectral signature of endmembers. Based on the spectral signature of endmembers and configuration of endmembers' location in the 3-dimensional feature space, it can be inferred that spectral response of spikes and senescence leaves as well as soil and shadow tend to be similar, whereas, the spectral response of gray panel and shadow had the most distinct spectral signature.

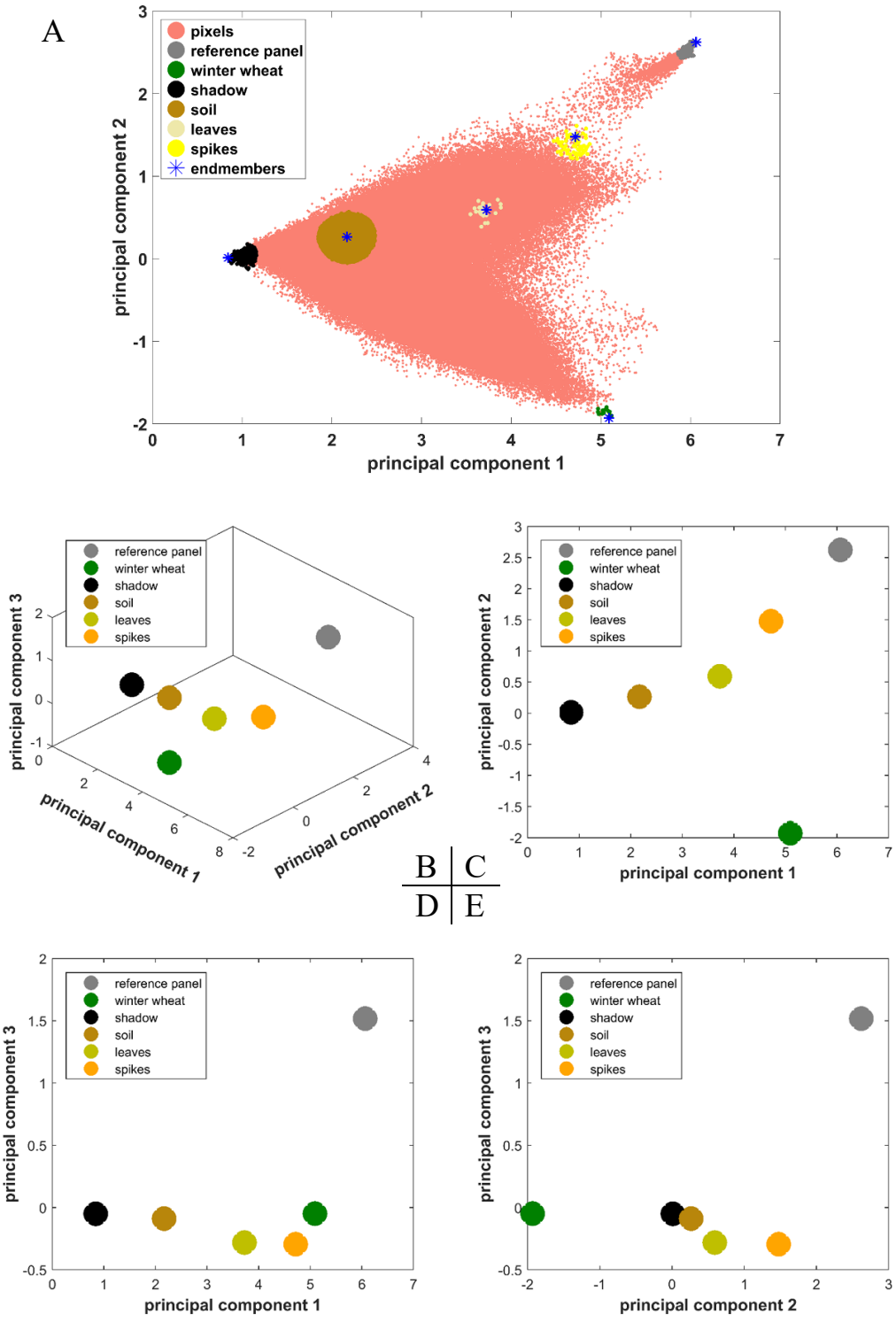

Figure 5.7. Location of the endmembers in the feature space spanned by the first two principal components (PC) (A), and the first three PCs (B). (C) Projecting the endmembers on the PC1 and PC2 plane. (D) Projecting the endmembers on the PC1 and PC3 plane. (E) Projecting the endmembers on the PC2 and PC3 plane. Depending on the projection, different set of endmembers become the vertices.

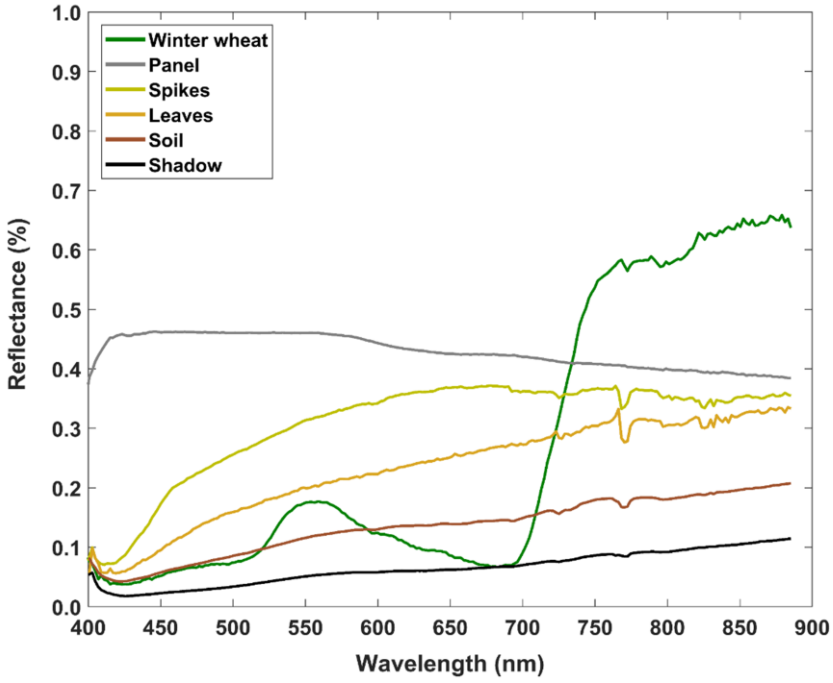

Figure 5.8. Spectral response of the six endmembers.

5.3.2 Spectral un-mixing

The spectral response of four endmembers, including spikes, leaves, soil, and shadow, were used for un-mixing analysis of P-HSC because the pixels representing winter wheat and gray panels were masked out during the segmentation process. The quadratic optimization problem, defined to minimize the Frobenius norm, returned four gray scale images, each of which representing the abundance of a particular endmember (Figure 5.9B). Therefore, for a given pixel in a P-HSC, there are four values denoting the abundance of endmembers such that the summation of these four values is equal to one due to the applied constraints in solving the optimization problem. To segment pixels representing biomass (i.e., SL class), the abundance of spikes and leaves were added pixel-wise. A binary mask was created to segment SL pixels (Figure 5.9C). A pixel was assigned to SL class if the summation of spikes and leaves abundances was more than the summation of soil and shadow abundances.

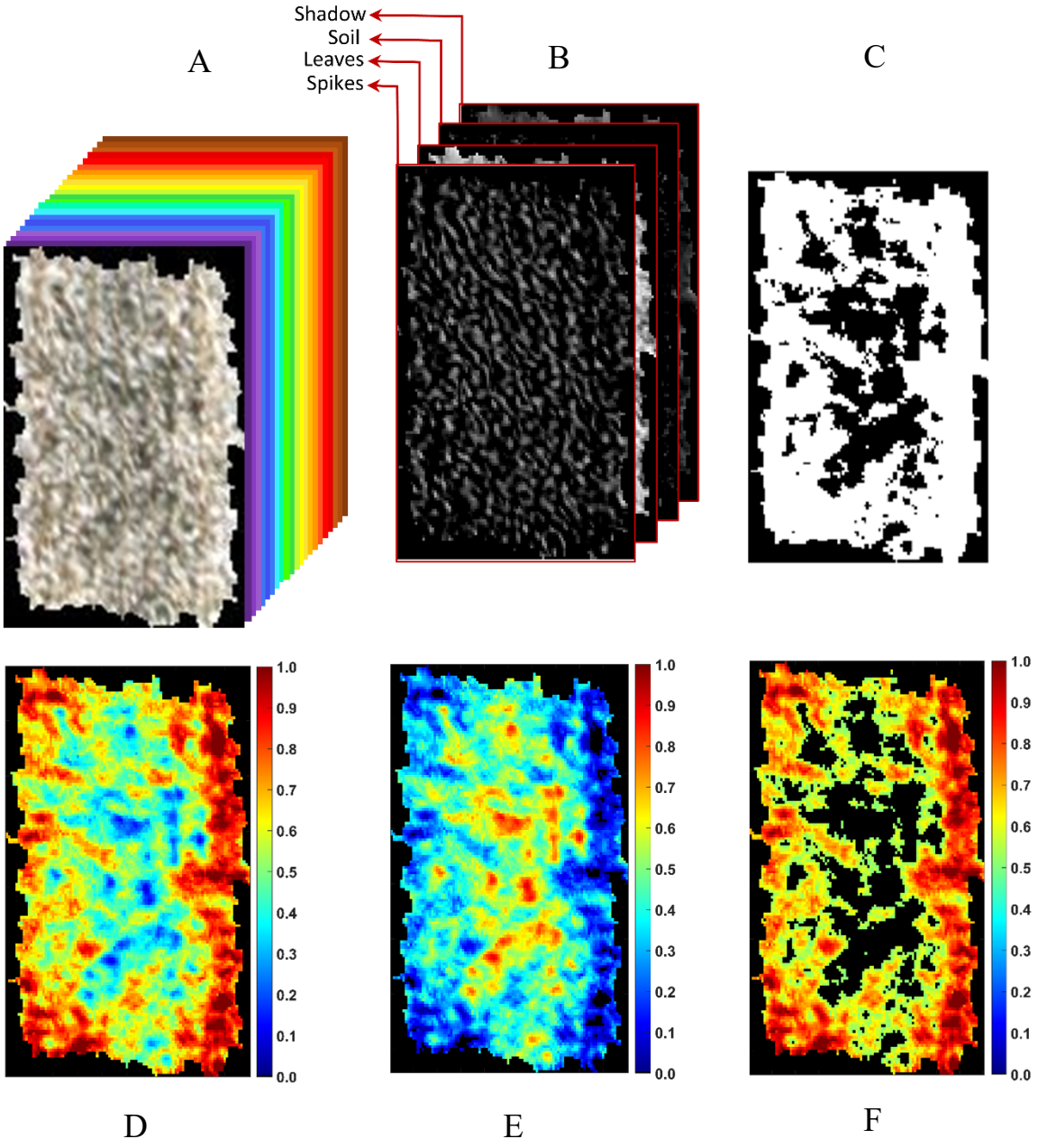

Figure 5.9. (A) Hyperspectral cube of a plot (P-HSC). (B) Abundance of endmembers in each pixel shown as gray scale images. (C) Binary mask of spikes and leaves class. A threshold of 0.5 was applied on the summation of spikes and leaves abundances. (D) Summation of spikes and leaves abundances shown as a colormap. (E) Summation of Soil and shadow abundances shown as a colormap. (F) Spikes and leaves pixels masked from background.

5.3.3 Yield allocation to sub-plots

The measured yield for a plot was distributed among the sub-plots based on the ratio between the number of SL pixels in sub-plots to the total number of SL pixels in the plot. Each sub-plot represents an area about 30×30 cm on the ground because the size of sub-plot was 15×15 pixels and the size of pixels were about 2 cm.

Several sizes for sub-plot window were evaluated to find an appropriate window size. While a small window size allows investigating the yield variation at a higher spatial resolution, the allocated yield to the sub-plots becomes very small as the number of SL pixels per windows decreases. In addition, the probability of having sub-plots with the same number of SL pixels increases, meaning an identical yield is assigned to a significant portion of sub-plots within a given plot. This can deteriorate the process of training the model since a significant portion of sub-plots has identical target variables. For instance, by dividing a plot shown in Figure 5.10 with a window size of 10×10, more than 40 sub-plots had yield values varying between 12.5 and 15 grams, and more than 54 percent of sub-plots had identical yield (Figure 5.11). Alternatively, for a larger window size, the assigned yield to sub-plots varied substantially at the cost of sacrificing the spatial resolution for investigating the yield variation in a plot. By dividing the same plot shown in Figure 5.10 using a window size of 20×20, the yield of sub-plots varied from zero to about 60 grams with only about 8% identical sub-plots yield. To maintain the possibility of investigating the yield variation at a higher spatial resolution and avoid numerous sub-plots with identical yield, the size of window was set to 15×15, compromising the benefits of 10×10 and 20×20 window size.

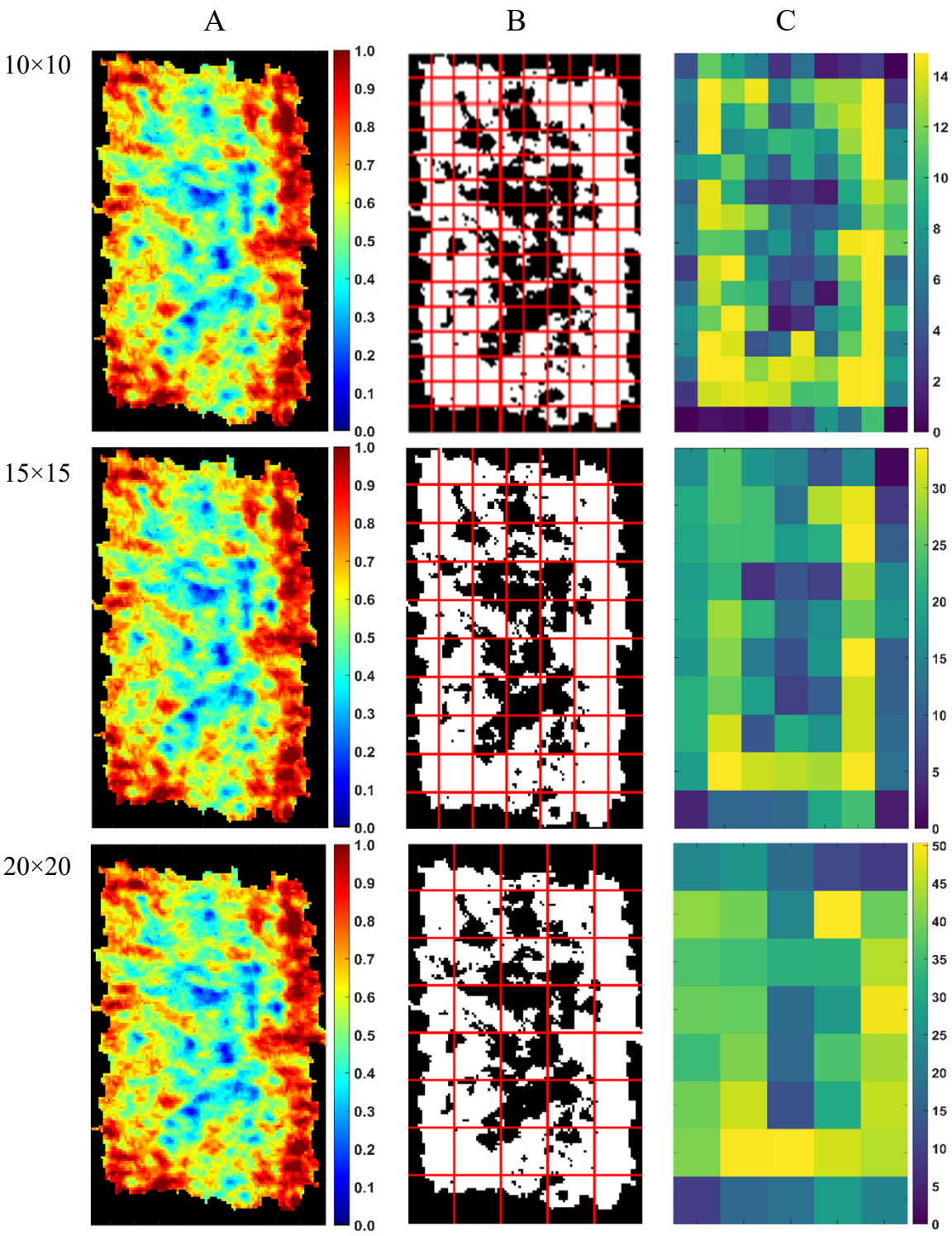

Figure 5.10. Dividing a plot using various window size of 10×10, 15×15, and 20×20. (A) Summation of spikes and leaves abundances in a plot. (B) Dividing the binary mask of spikes and leaves into sub-plots using various window size. (C) Yield allocated to the sub-plots.

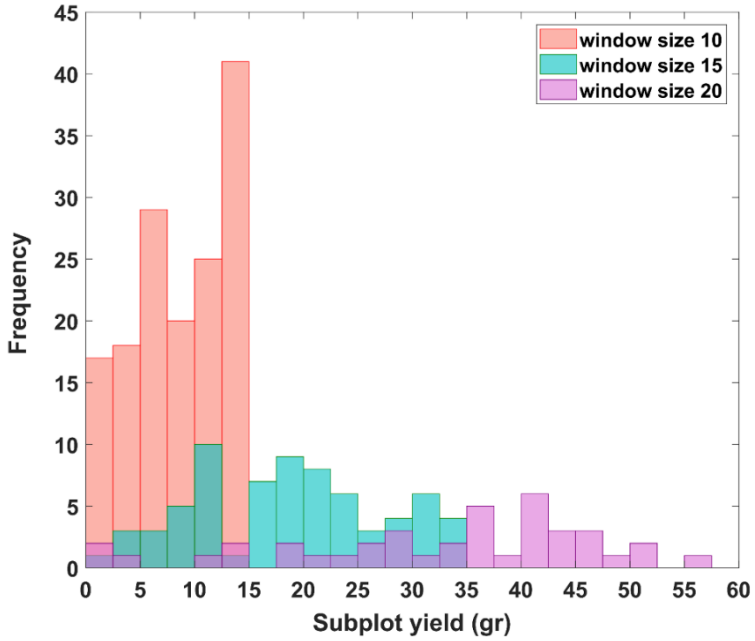

Figure 5.11. Yield histogram of sub-plots generated by various window sizes. For the window size of 10×10, the variation range of the allocated yield values to sub-plots was small, whereas the variation range of the allocated yield values to sub-plots generated by a window size of 20×20 was wide.

5.3.4 Deep Neural Network

5.3.4.1 Yield prediction at sub-plot scale

The training dataset of sub-plots was used to train a DNN model. In training the model, the main goal was to identify a set of model parameters (weights and biases) that minimize the cost function's value (i.e., RMSE). As training continued, model parameters were updated. To achieve an interpretable unit (gram) as the target value (yield), root mean square error (RMSE) was calculated for presenting the variation of cost function over epochs. Figure 5.11 illustrates how RMSE changes over training epochs for three individual models developed for each field as well as the model trained on the large training dataset obtained by merging all three fields. For all four models, RMSE decreased rapidly over the first training epochs for both training and validation datasets, subsequently, reached a plateau where RMSE remained rather unchanged. However, for the merged dataset, there was a sharp decrease in RMSE within the first few epochs, meaning that the convergence

occurred faster than other models. Among the 100 epochs, the weights and biases returning the lowest RMSE for validation dataset was saved as the model parameters to predict the yield of test dataset.

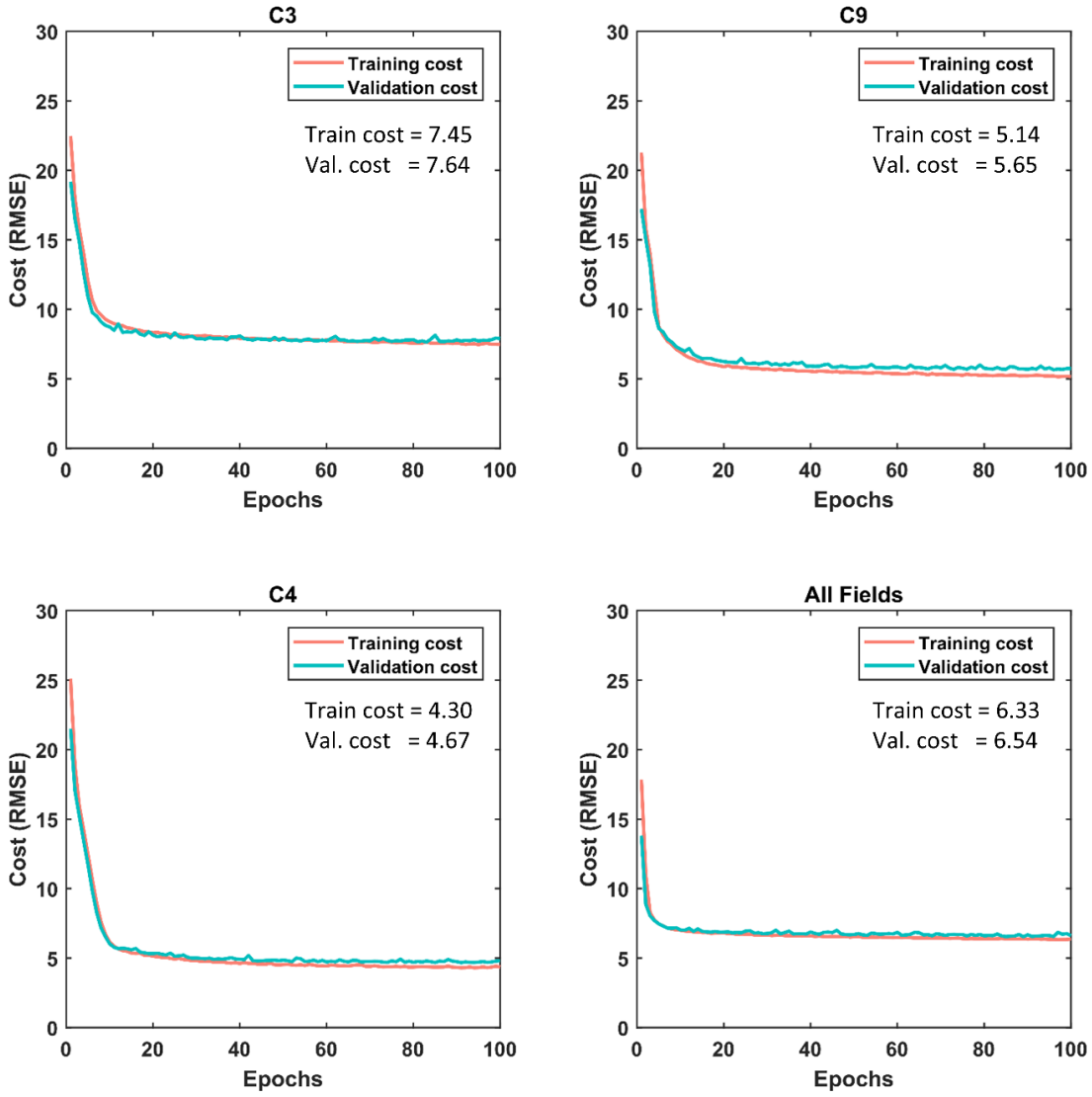

Figure 5.12.Variation of root mean squared error over epochs for C3, C9, C4, and merged dataset.

Figure 5.13 demonstrated the performance of the trained models in predicting the yield of the sub-plots in test datasets. The model trained on C9 dataset had the largest coefficient of determination (R^2) and lowest RMSE in predicting the yield. Alternatively, C3 model had the lowest R^2 and largest RMSE, indicating the generalization of the trained model on

an unseen dataset was not as satisfactory as the C9 model. This could be anticipated because the train and validation cost for C3 model during the training process was largest among the models (Figure 5.12). One reason that might explain this is the difference between the dates that imaging were captured from these two fields in 2017. The time interval between imagery and harvesting of C3 was one week more than C9.

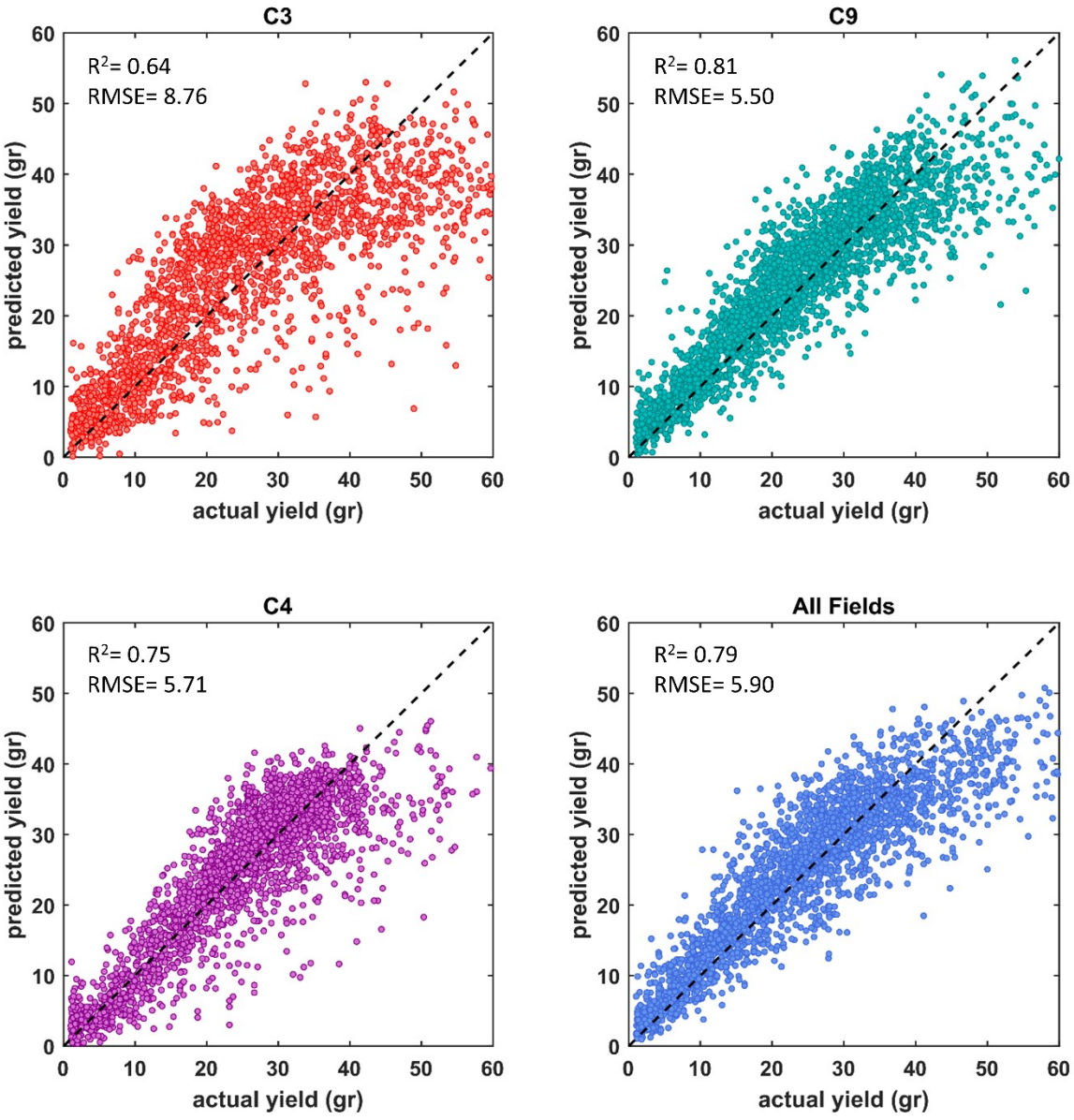

Figure 5.13. Performance of deep neural network models on yield prediction of sub-plot test datasets. The model trained on C9 training dataset had the best performance (R^2=0.81 and RMSE=5.5 gram), while the model trained on C3 had the lowest R^2 and largest RMSE.

The model trained on the merged dataset had a promising R^2 ($R^2 = 0.79$) and low RMSE of 5.90 gram, indicating the DNN model could explain 79 percent of yield variation among the 2444 subplots in the test dataset. The DNN model was able to predict the yield of a significant portion of the test sub-plots with a low error as they located close to the 1:1 line (the black dashed line). However, the model demonstrated a tendency to under estimate the yield of sub-plots with the large yield value (more than 40 gram). This might be because the number of sub-plots with the yield more than 40 gram in the training dataset was substantially lower than the number of sub-plots with the yield less than 40 gram (Figure 5.5D). Therefore, the network was moderately successful to learn the yield estimation based on the input data with a large sub-plot yield. A similar pattern was observed in yield prediction of the individual fields (Figure 5.13).

5.3.4.2 Yield prediction at plot scale

As described before, these test sub-plots were obtained from the 50 test plots selected from the three fields using stratified sampling. To observe the performance of the trained model in predicting the yield at plot scale, the summation of the predicted yield for the sub-plots belonging to a test plot were compared to the measured yield for that plot (Figure 5.14). For yield prediction at plot scale, R^2 dropped to 0.41 compared to R^2 at sub-plots scale, which was 0.79. To compare the prediction error obtained for plot and sub-plot analysis and account for the difference in the scale of yield variation, the normalized RMSE was calculated by dividing the RMSE of plot and sub-plots to their mean of yield. The normalized RMSE for predicting the yield at plot level was 0.14, while it was 0.24 for yield prediction of sub-plots, indicating the error in yield prediction of plots improved although R^2 deteriorated compared to the sub-plots.

Moreover, Figure 5.14 demonstrates that the DNN model, trained on the merged sub-plots of three fields, tends to overestimate the yield of the high-yielding plots and underestimate the yield of low-yielding plots. This suggests that DNN could not adequately learn the complicated pattern from the spectral response of the low- and high- yielding plots, largely because there were limited number of low- and high-yielding plots in the training dataset. Figure 5.5 showed the low number of plots at the two tails of the yield histogram for each field.

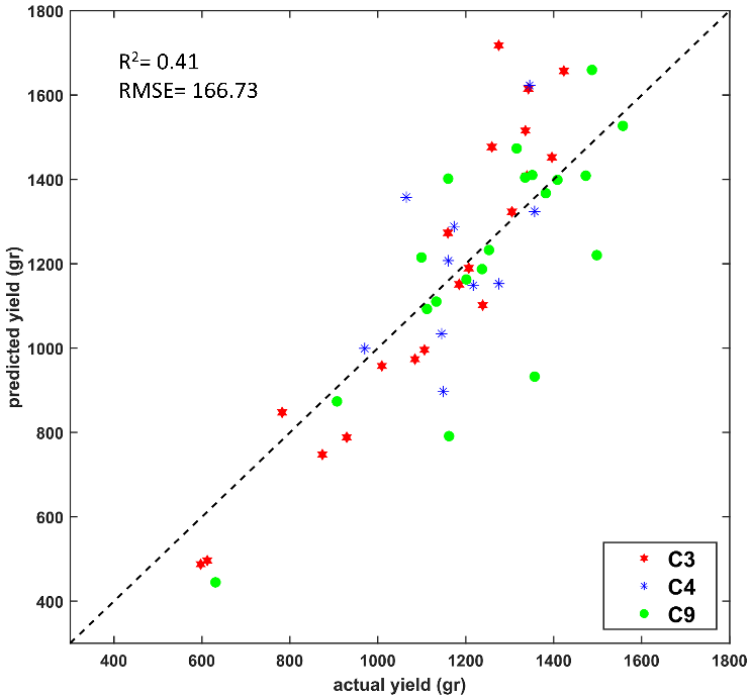

Figure 5.14. Performance of the model on yield prediction at plot scale.

Among the test plots, there were plots that the network could accurately predict the yield of their sub-plots. Figure 5.14 illustrates an example of such a plot that the network could explain about 96 percent of yield variation among its 62 sub-plots with RMSE of 1.90 gram. In addition, Figure 5.15 shows a test plot that the network overestimated the yield of the substantial number of its sub-plots, and an example of a test plot that the network underestimated the yield of the majority of its sub-plots.

5.3.4.3 Yield prediction at a larger scale

To evaluate the feasibility of yield prediction at a large scale, the measured yield for all 50 test plots were added as a yield of a large field composed of 50 wheat plots. Alternatively, the predicted yield for these 50 plots were also added together as the predicted yield for such a large field. The total actual yield of the test plots was 59,356 gram, and the total predicted yield of these plots was 59,494 gram. Such an impressive result (i.e., about 0.2% error in yield prediction) indicates the capability of the proposed pipeline for yield prediction at a large field scale.

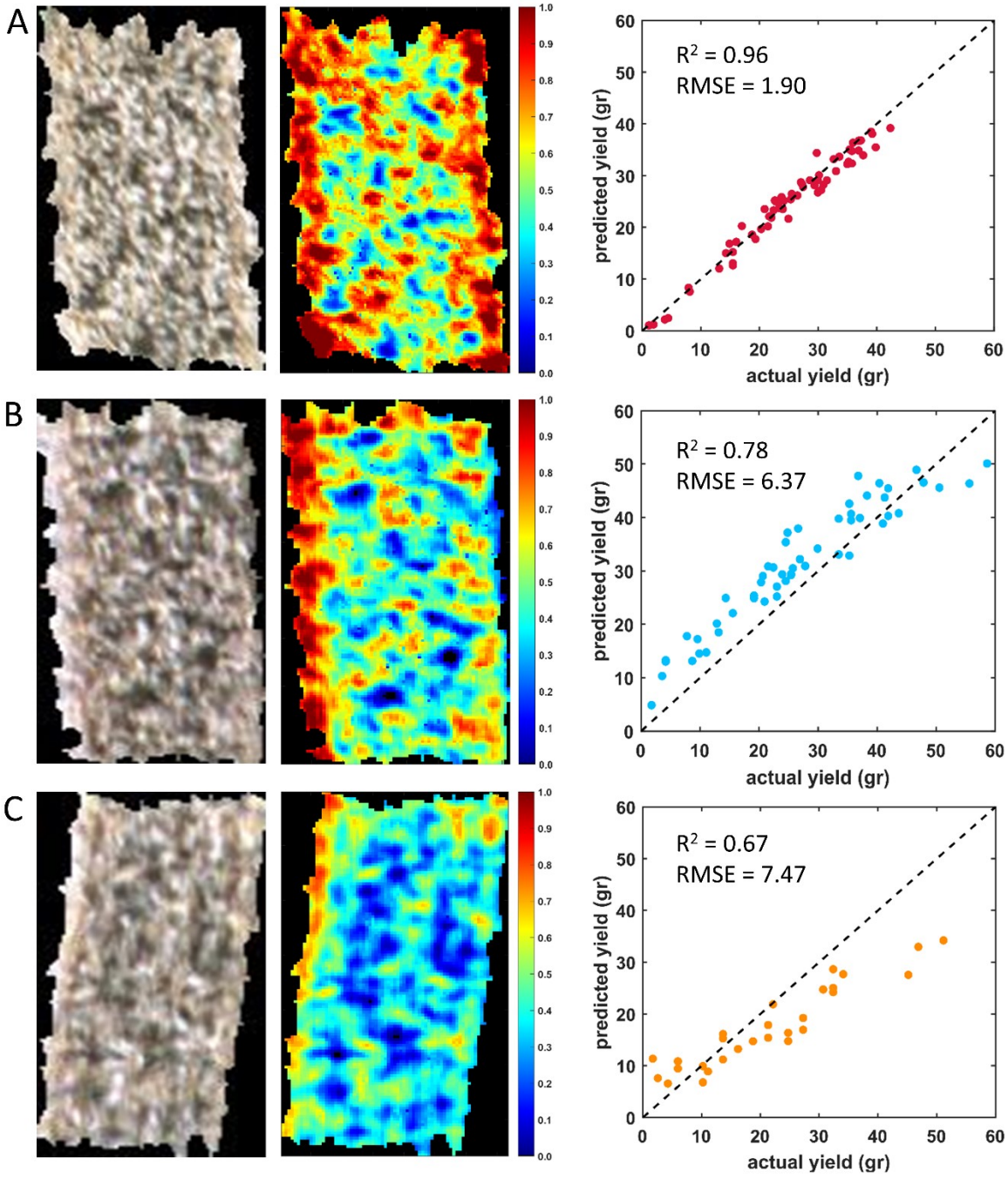

Figure 5.15. Examples of test plots that the network accurately estimated (A), overestimated (B), and underestimated (C) the yield of their sub-plots. Wheat plots are wavy because the gimbal could not restrict the amplitude of vibrations caused by UAV.

5.4 Discussion

Yield is the most fundamental trait in plant breeding since almost every other characteristic of crops, treatments, and management decisions are evaluated through the lens of whether they promote or hinder the yield potential. Two years of yield experiments were performed in three experimental fields, two fields, C3 and C9, in 2017 and one field, C4, in 2018. The grain of individual wheat plots was harvested by a combine, weighted, and recorded as the ground truth data. Several plots were disregarded for analysis because of damages caused by animals (e.g., rodent and birds) and severe weather (e.g., hail and strong wind). Aerial hyperspectral images were captured at about 20 meter above the ground level, resulting in spatial resolution about 2 cm. Image pre-processing, including radiometric calibration and noisy band removal, was performed to compensate for the non-uniform spectral and spatial responses of the instrument and suppress the impact of noisy bands, respectively. Subsequent to the pre-processing steps, various methods could be used to analyze hyperspectral images for yield prediction.

One of the widely used approaches is to utilize spectral indices, mostly NDVI. While yield prediction would be more accurate toward the end of growing season and prior to harvesting when the density of crop canopy is moderate to high, a method based on NDVI suffers from saturation issues at this stage of crop growth (Asrar et al., 1984; Gitelson, 2004; Gitelson et al., 1994). Therefore, this method was not suitable for yield prediction.

The other possible method was to train a model to predict the yield at plot scale. For such a model, the spectral response of pixels representing spikes and leaves in a single plot should be averaged to have one feature vector since there was a single target value (i.e., measured yield) per each plot. One main drawback of this naïve approach is that substantial spectral information is suppressed by taking the spectral average over hundreds of SL pixels in a plot. Moreover, the spatial information attained with high resolution is diminished through the averaging process. The other disadvantage of taking average across the plot pixels is that the number of samples is limited to the number of plots which was about 1000 in this study. This low number of samples might be insufficient to recognize the pattern from a high dimensional dataset with 190 features (i.e., number of bands) to

develop a robust model for yield prediction.

This study proposed an innovative method for analysis of high-dimensional hyperspectral images captured at high spatial and spectral resolution to estimate the yield of hundreds of wheat lines. Aerial hyperspectral images were captured in less than 10 minutes from each field using an autonomous platform. Several image processing techniques and an optimization algorithm were integrated with the domain knowledge to segment the plots from background, divide them into sub-plots, unmix the plot pixels, and assign a yield value to each sub-plot. Subsequent to these analyses, OBIA approach was deployed to extract features from each sub-plots. Finally, deep neural networks were used to estimate the yield at sub-plot and plot scale. The results achieved by the proposed analysis framework are discussed in this section.

5.4.1 Spectral mixture analysis

With the spatial resolution of 2 cm, each pixel could potentially be a mixed-pixel, a spectral mixture of more than one particular endmember. Once the spectral signature of the endmembers were discovered from the hyperspectral image with 0.5 cm spatial resolution, the spectral mixture analysis was performed to identify the abundance of the endmembers in a given pixel. The benefits of un-mixing the pixels can be summarized into twofold. First, it allowed segmenting the plot pixels with high abundance of wheat leaves and spikes and disregarding the pixels representing background for further processing. Second, this approach provided the opportunity to assign a yield value to a given sub-plot based on the number of SL pixels in that sub-plot.

5.4.2 Yield analysis at sub-plot scale

Besides the yield potential of a variety, evaluation of producing a uniform yield across the plot is valuable information that could assist breeders in selecting advanced lines. However, harvesting the grains at sub-plot scale to study the yield variation of various wheat varieties within plots is not practical particularly in a large nursery. The novel approach proposed in this study offered this chance to investigate the feasibility of yield

estimation at sub-plot scale with very high spatial resolution, and further evaluate the performance of various wheat lines in terms of producing yield uniformly distributed across the plot.

The results of the yield analysis at sub-plot scale revealed the significance of marginal effects on the distribution of spikes and leaves for various wheat varieties. While a particular variety might be capable of producing a uniform yield across the plot because the plants were able to compete with their neighbors, another variety might be sensitive to the plant density, causing non-uniform yield production. A uniform yield production is a fundamental trait because plants should maintain their potential yield in a competitive environment at field scale. Figure 5.16 shows two wheat lines A and C suffering from marginal effects while producing more yield compared to the wheat lines presented in B and D producing less yield but more uniformly, less affected from the marginal effect. According to the colormaps in Figure 5.16 showing the distribution of the spikes and leaves, lines A and C produced less yield inside of the plot and more yield at the margins of the plot. Therefore, breeders prefer line B and D because of their potential for uniform yield production.

It should be noted that although the existence of SL pixels (presented in red color) is a reasonable indicator for more yield within a plot (i.e., one variety), it is not necessarily a valid metric to compare the yield across the plots (i.e., varieties). It is because the plants within a plot share similar characteristics as they all belong to a single variety, whereas, plants of different plots may demonstrate various characteristics since they belong to diverse wheat varieties. For instance, one variety with more spikes might produce less yield compared to a variety that has fewer spikes but with larger size and more weight. Therefore, spectral information is required to account for variability across the varieties since the spectral response of a plant carries valuable information about the internal activity.

Alternatively, the existence of more SS pixels (presented in blue color) implies less yield regardless of the variety because it indicates the pixels representing soil and shadow, which does not contribute to the yield. A large proportion of a plot covered with background pixels (blue spots in colormap) refers to a low-yield plot.

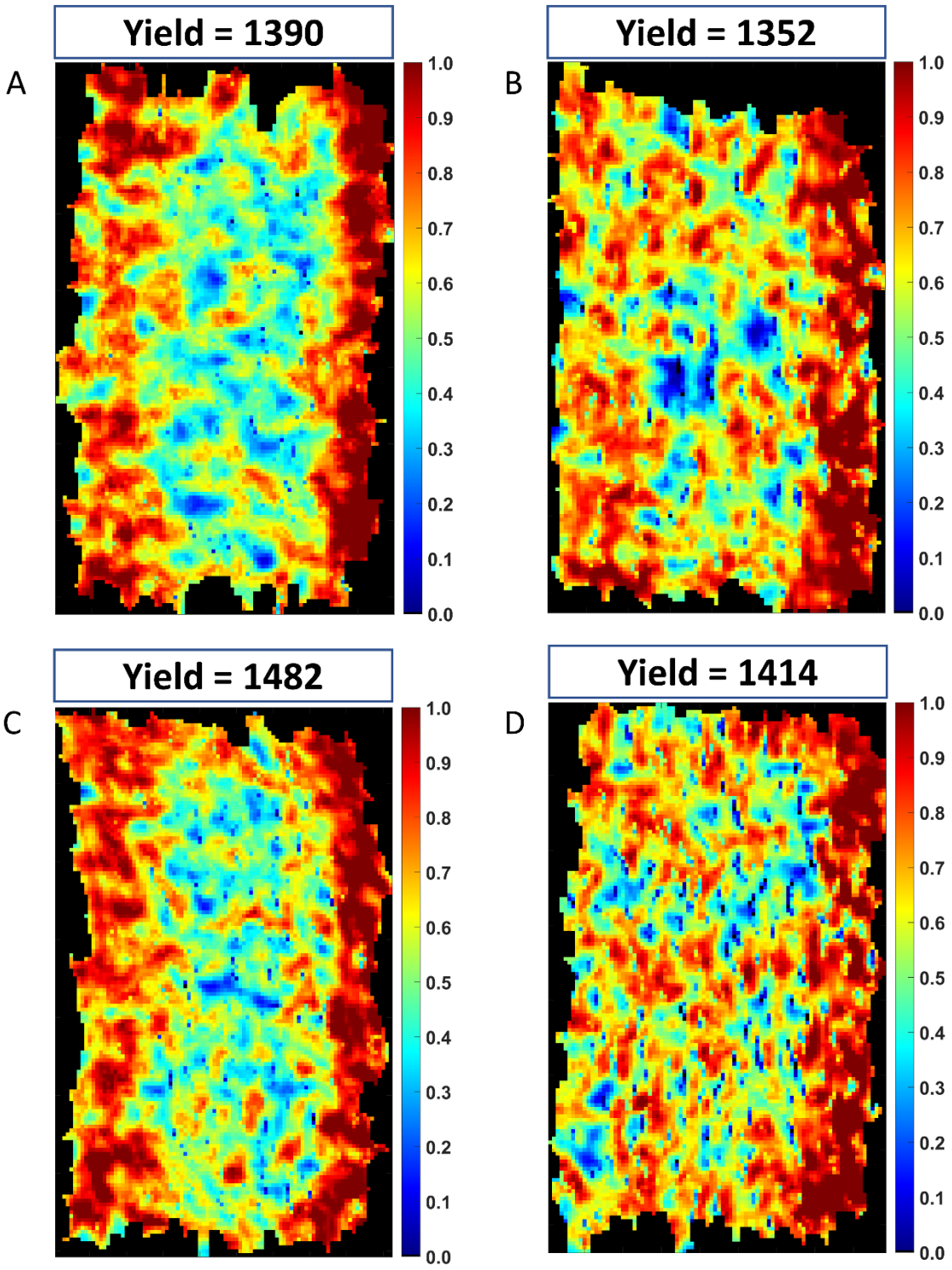

Figure 5.16. Distribution of spikes and leaves pixels in four wheat lines. Two wheat lines suffer from marginal effects (A) and (C); while the other two wheat lines produce less yield with a uniform distribution (B) and (D). Despite less yield production, breeders may prefer (B) and (D) because of uniform yield production. Wheat plots are wavy because the gimbal could not restrict the amplitude of vibrations caused by UAV.

5.4.3 Yield production in the middle one-third of the plot

An additional analysis was performed to determine the potential of each wheat variety in producing yield in the middle one-third of the plot. The results suggest that about 90 percent of the plots produced more yield at one side of the plot (Figure 5.17). Lower productivity in the middle one-third can be a result of the high competition between the plants in the middle of the plot and/or receiving less light compared to the plants at the border of the plots.

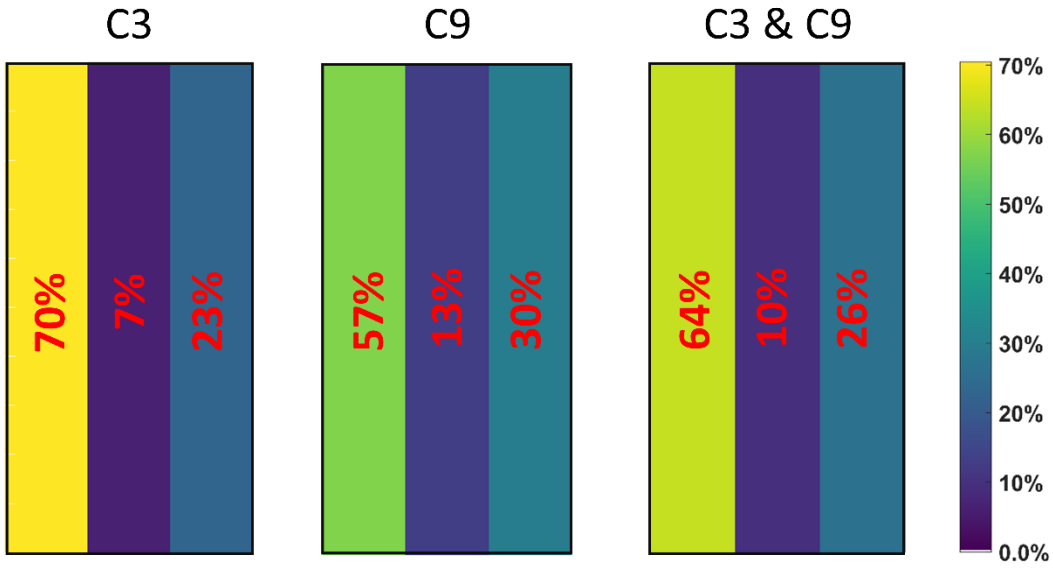

Figure 5.17. Analysis of yield production in the middle one-third of the C3 and C9 plots.

5.4.4 Aerial inspection of the field

Breeders visually inspect the nursery multiple times during the growing season to record any incidents that might affect their screening, such as damages caused by animal or severe weather condition. Obviously, this is an extremely demanding, time-consuming, and subjective task. Aerial imaging followed by the proposed automated analysis pipeline can facilitate the visual inspection to be performed with high temporal resolution and across all nursery in multiple locations. Figure 5.18 shows the SL colormap obtained by analysis of aerial hyperspectral images conformed to the notes taken by an expert in the field. This observation again demonstrates that breeder can identify quickly and effectively the plots with low yield with large proportion of blue spots representing background.

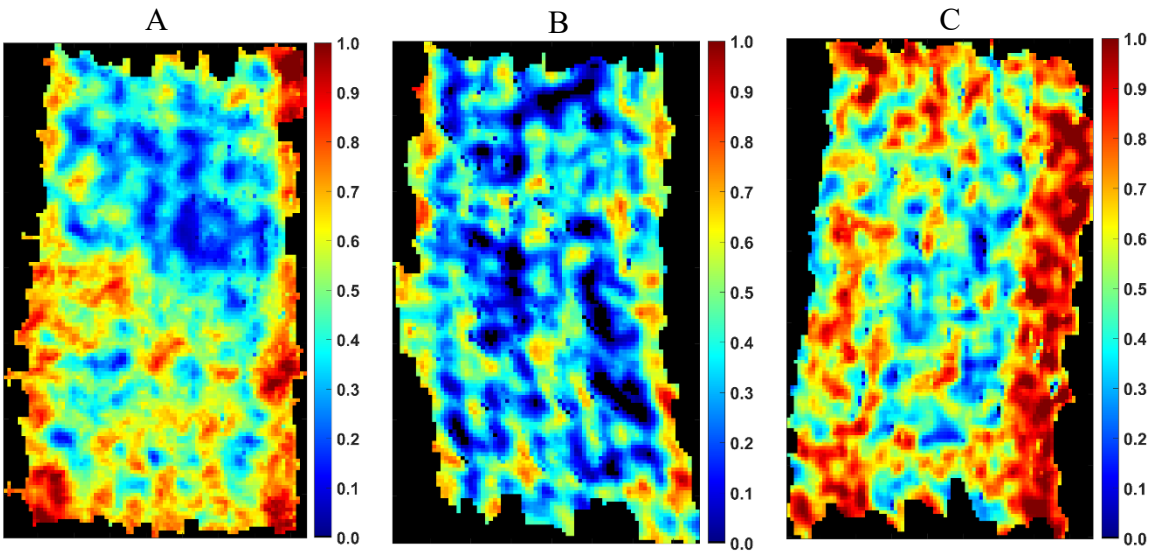

Figure 5.18. Analysis of aerial imagery to replace the visual inspection of the plots. The notes made by an expert for these plots in the field were: (A) ½ of the plot is shorter, (B) weak plot, (C) strong plot. Wheat plots are wavy because the gimbal could not restrict the amplitude of vibrations caused by UAV.

5.5 Conclusion

Crop production needs to increase in a sustainable manner to meet the growing global demand for food. To identify crop varieties with high yield potential, plant scientists and breeders evaluate the performance of hundreds of lines in multiple locations over several years. To facilitate the process of selecting advanced varieties, an automated pipeline was developed by mounting a hyperspectral camera on a UAV to collect aerial imagery in a fast, cost-effective manner, and integrating domain knowledge with computer vision and deep learning to analyze the images. The proposed method offers novel insights for making decisions in selecting high-yielding wheat varieties in an efficient and effective manner by predicting the yield at sub-plot level. The coefficient of determination for predicting the yield at sub-plot scale was 0.79 with RMSE of 5.90 gram. In addition to yield prediction at sub-plot scale, the findings of this study revealed that the proposed method was able to accurately predict the yield at larger scale: plot size and large field. In addition to yield estimation, this study can benefit plant scientist and breeders in various ways, including: (i) optimizing the plot size to mitigate the alley effects on yield, (ii) investigating the effect of plant density on yield at high spatial resolution, (iii) studying the impact of side trimming

on yield across various varieties, and (iv) facilitating the process of visual assessment of the plots by collecting images at high spectral, spatial, and temporal resolution. In future work, the dimension of hyperspectral images can be reduced to avoid the issues associated with high dimensional data. Furthermore, a deeper network can be trained with the same sample size because the number of model weights will be significantly reduced due to fewer dimension of input features at the first layer of the network.

CHAPTER 6

Deep Autoencoder for unsupervised feature learning from hyperspectral images to predict the yield of experimental wheat plots

6.1 Introduction

The importance of yield estimation to identify high-yielding varieties were discussed in Chapter 5. Alternatively, in Chapter 3 and 4, we discussed the significance of dimensionality reduction in hyperspectral image analysis. In summary, we argued the high-dimensionality of hyperspectral images hinders the analysis of images by increasing the required sample size, risk of overfitting, complexity of interpretation, and running time. Therefore, selecting an innovative data representation is a crucial step in hyperspectral image analysis using machine learning algorithms. However, f*eature engineering* to extract meaningful features and discriminative information that are appropriate for the subsequent machine learning algorithms is a challenging task. Goodfellow, Bengio, & Courville (2016) described why this is a difficult task through a car detection example. They mentioned we might consider the presence of wheels as a feature since cars have wheels. Then, they argued that:

"Unfortunately, it is difficult to describe exactly what a wheel looks like in terms of pixel values. A wheel has a simple geometric shape, but its image may be complicated by shadows falling on the wheel, the sun glaring off the metal parts of the wheel, the fender of the car or an object in the foreground obscuring part of the wheel, and so on."

Similar problems exist in analysis of images captured in agricultural fields with even more complex environment. Goodfellow et al. (2016) proposed using *representation learning* as an approach to address this problem:

"One solution to this problem is to use machine learning to discover not only the mapping from representation to output but also the representation itself. This approach is known as ***representation learning****. Learned representations often result in much better performance than can be obtained with hand-designed representations. They also enable AI systems to*

rapidly adapt to new tasks, with minimal human intervention."

The present study focused on learning a representation of hyperspectral data that reduces the dimensionality of the images while retaining the yield prediction accuracy. The method used for representation learning was autoencoder, which is a deep leaning method. The objectives of this study were to: (i) develop a deep autoencoder network to project the hyperspectral data onto lower dimensional feature spaces while retaining invaluable information for yield prediction, and (ii) train a deep neural network based on the new data representation to predict the yield potential of experimental wheat plots.

6.2 Deep fully-connected autoencoder

Autoencoders are categorized as self-supervised learning techniques composed of two symmetrical networks: encoder and decoder. While an autoencoder network attempts to reconstruct the input layer at the output layer, the hidden layer is forced to learn the best representation of the input data (Goodfellow et al., 2016). Figure 6.1 shows the encoder and decoder part of a shallow autoencoder network with one hidden layer. In encoder part of the network, the input data x is encoded as a compressed representation (projecting to a 3-dimenional feature space), and subsequently, the encoded data is decoded back to the same dimension of the input data x, meaning the input data is reconstructed in decoder part.

In this study, we developed a deep autoencoder consisted of an encoder with three hidden layers and a symmetric decoder (Figure 6.2). The encoded part of the network entailed three fully connected hidden layers each of which had a ReLU activation function. The decoded part of the network had a similar architecture except the transfer function of the last layer, which was a sigmoid function. The reason for using the sigmoid function for the last layer was to keep the range of the output layer values between zero and one, similar to the input layer values, which were the reflectance of a pixel. To measure the network error, the cost function was defined to mean squared error (MSE) which is the squared difference between the decoded data (i.e., the reconstructed spectral response) and the input data (i.e., the spectral response of a pixel), averaged across all pixels used for training. Number of epochs was set to 100, and the model returning the minimum MSE for validation dataset

was saved for further processing. Since the size of the training dataset was large, it was divided to mini-beaches. The batch size[1] was set to 2^{12}, meaning 4096 samples were propagated through the network before updating the network's parameters. Consequently, in each epoch[2], there were 479 iterations[3] (number of samples divided by the batch size). At each iteration, one mini-batch is used to update the network's parameters. It should be noted that the size of the samples in the final mini-batch was fewer than the other batches. Adam algorithm (Kingma & Ba, 2014) was used as the optimization algorithm to identify the network parameters aimed at minimizing the cost function.

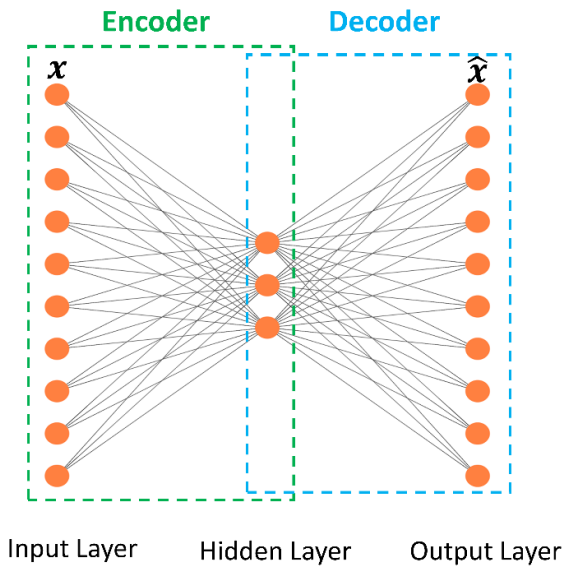

Figure 6.1. A fully-connected autoencoder with one hidden layer, x is the input to the network and $\hat{x}$ is the reconstructed input at the output layer.

The number of hidden layers and nodes in each layer was defined through an empirical process. The first and second hidden layers had 90 and 30 units, respectively. For the third hidden layer, the number of units was set to three and five to evaluate the projection of hyperspectral data onto a 3- and 5-dimensoinal feature space. In addition, another autoencoder was trained to project the hyperspectral data onto a 10-dimensional feature space. The first two hidden layers in this autoencoder had 100 and 50 units, respectively, and the number of units in the third layer was set to 10.

1. Batch size indicates the number of samples passed through the network before updating the parameters.
2. Epoch refers to the number of times that the entire training dataset is passed through the network.
3. Iteration denotes the number of batches in each epoch.

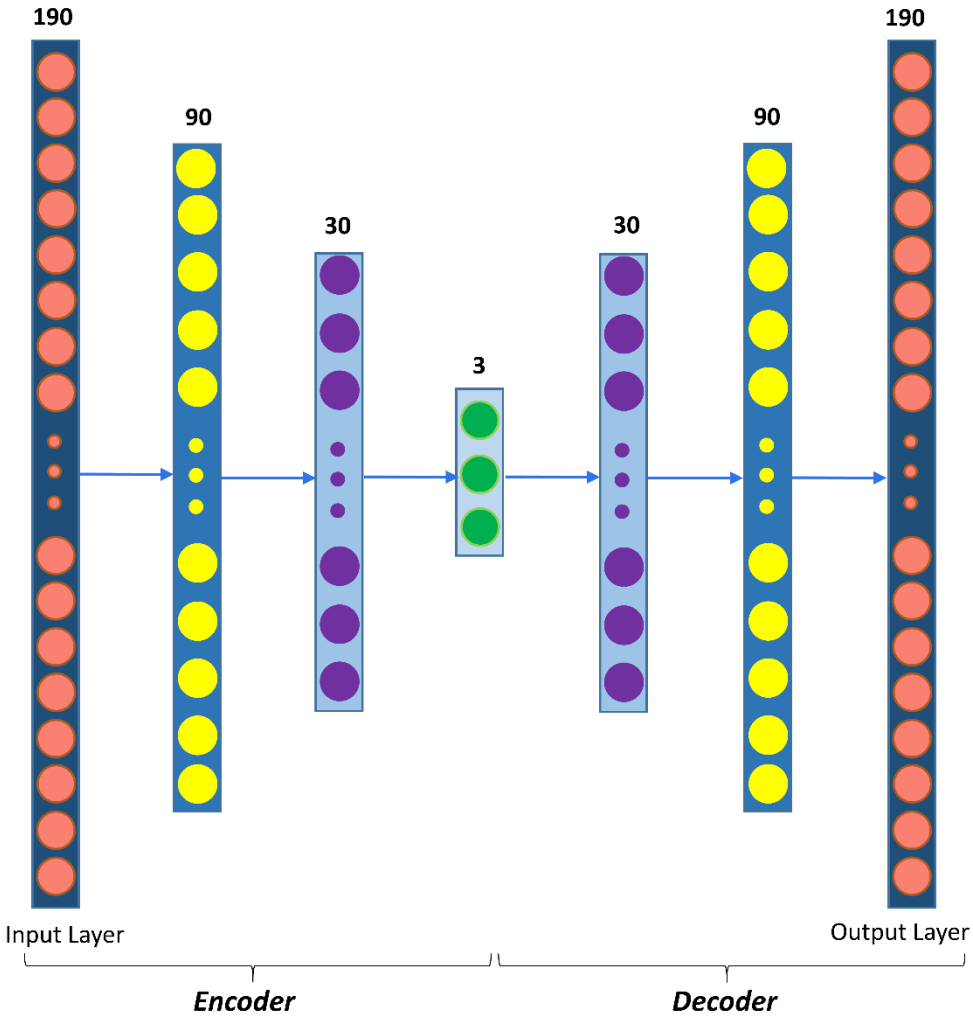

Figure 6.2. Architecture of a fully-connected autoencoder to project the input data onto a low dimensional feature space. Since the last layer of encoder has three units, this network projects the input data onto a 3-dimensional feature space. This layer with three units is the input of the decoder. To reconstruct the input data, the decoder projects it back the same dimension of the input data at the output layer.

6.3 Hyperspectral dataset

Since the objective was yield prediction, the pixels representing the wheat plots were used for the training the network. In total, there were more than eight million pixels segmented as the wheat plots. The results obtained during the tuning of hyper-parameters demonstrated that about two million pixels were adequate to achieve a reasonable MSE. Therefore, a subset of two million were randomly selected, 1,960,000 pixels to train the hyper-parameters of the network, and a subset of 40,000 pixels for validation of the network.

6.4 Results

6.4.1 Feature learning

The Adam optimization algorithm updated the network's parameters (weights and biases) at each iteration to minimize the MSE. Figure 6.3 demonstrates how MSE for training and validation dataset changed over various epochs. It should be noted that the MSE presented for a particular epoch was obtained by averaging across all the iterations in that epoch. After the first few epochs, MSE remained rather unchanged for the three networks. This might be because of the large training dataset, and the fact that the network's parameters were actually updated 479 times at each epoch.

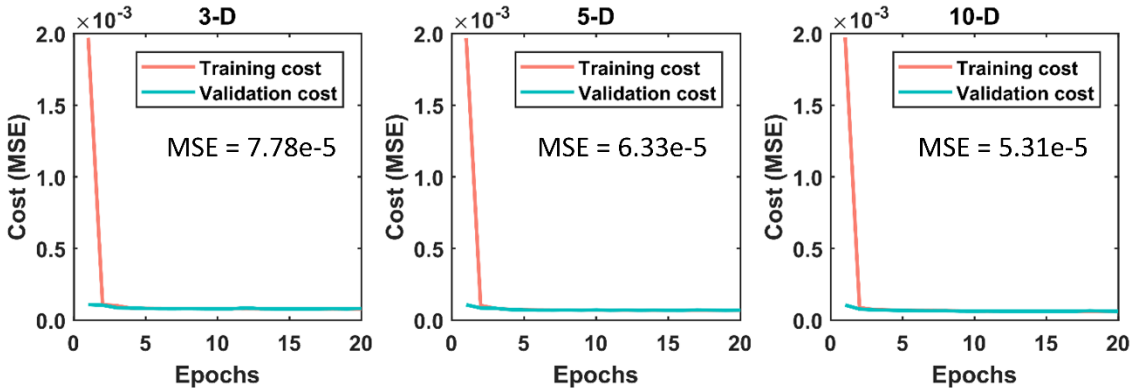

Figure 6.3. The variation of MSE over epochs for the three autoencoders projecting the input data onto 3-, 5-, and 10-dimensional feature space. Only the first 20 epochs were shown because MSE remained unchanged afterwards.

The error of reconstructing the input data from three, five, and 10 dimensions was about 7e-5, 6e-5, and 5e-5, respectively. The result demonstrated that the trained autoencoder could substantially reduce the dimension of hyperspectral data, and subsequently, project these low dimensional data back to the same dimension of the input with minor error. Through this reconstruction process, a substantial noise was eliminated from the raw data. Figure 6.4 depicts the spectral response of two randomly selected pixels verses their reconstructed reflectance obtained by decoding the 3-, 5-, and 10-dimensional data. The reconstructed data has considerably lower noise compared to the raw data. This indicates that autoencoders can be used for data denoising besides feature learning.

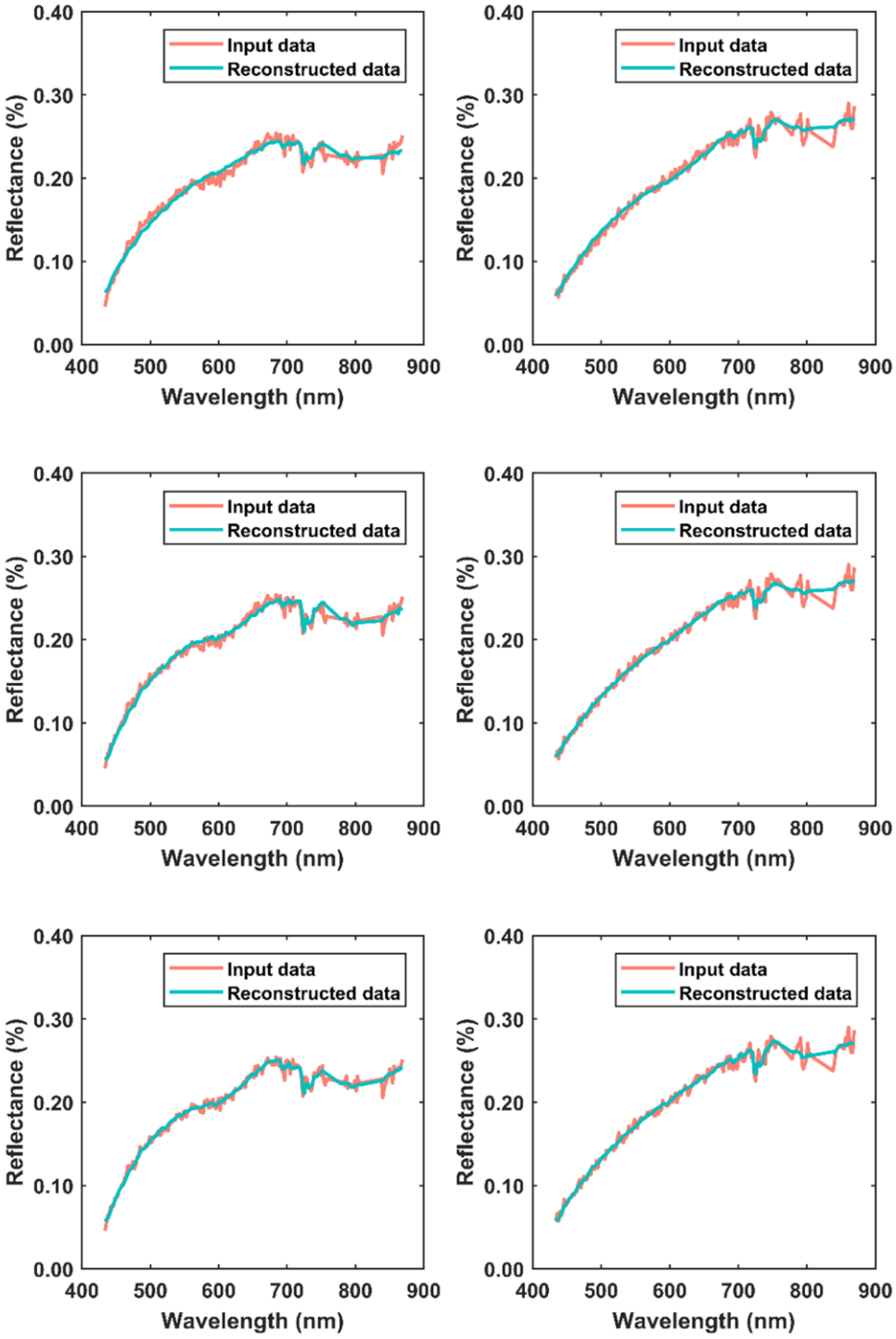

Figure 6.4. The reflectance of two randomly selected pixels and the reconstructed spectral reflectance obtained by decoding the 3-, 5-, and 10-dimensional encoded data. The reconstructed reflectance is less noisy compared to the input raw reflectance.

6.4.2 Low dimensional feature space

Each hidden layer of the encoder is a new representation of the input data in a different feature space. Through the encoding process, the dimension of the input data was reduced until the last layer of the encoder. For instance, the first layer projected the input data with full-dimension (190 bands) onto a 90-dimensional feature space. This process continued until the last layer of the encoder where the input data was projected on 3-, 5-, or 10-dimensional feature space.

The results achieved during the hyper-parameter tuning revealed that one or more units of the last layer of encoder (bottleneck layer) become zero if the number of units for this layer was set to a number more than five. For instance, when the autoencoder was trained to project the hyperspectral images onto a 10-dimensioal feature space, the decoder required eight units of the encoder output (bottleneck layer) to reconstruct the input data, meaning two units of the last layer of the encoder were zero. This might be because the decoder could learn how to reconstruct the input layer using only eight units while minimizing the cost function. To assure this is not because of the ReLU transfer function, the transfer function of the last layer in encoder (bottleneck layer) was set to a sigmoid function, which returns values between zero and one. In the new autoencoder, the values of two units in the last layer of the encoder were saturated to one, regardless of the input data. The result of the new autoencoder verified that the decoder required eights units when the number of units in the last layer of encoder was set to 10.

6.4.3 Deep neural network for yield prediction

Once the autoencoder network was trained, the encoder part of the network was used to project the hyperspectral images onto a lower dimensional feature space, a new representation of the input data in three, five, and 10 dimensions.

The pipeline developed in Chapter 5 was used for splitting the data into training, validation, and test datasets, and dividing the plots into subplots. Furthermore, the same deep neural network (DNN) architecture was utilized for yield prediction based on the new datasets (i.e., 3-, 5-, 10-dimenionsal datasets). Table 6.1 compares the performance of DNN on

yield prediction of the test dataset.

Table 6.1. Performance of deep neural network on yield prediction of the test dataset. The network were trained on three datasets: full-, 10-, 5-, and 3-dimensional datasets.

		Hyperspectral dataset (full-dimensional)		10-dimensional dataset		5-dimensional dataset		3-dimensional dataset	
Year	Fields	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
2017	C3	64.13	8.76	65.60	8.54	61.86	9.00	58.30	9.41
	C9	81.49	5.50	79.59	5.78	79.33	5.81	76.31	6.22
2018	C4	75.35	5.71	75.18	5.73	73.79	5.89	74.48	5.81
All fields		79.89	5.90	78.52	6.10	76.19	6.42	68.13	7.43

6.5 Discussion

Selecting an optimal set of informative bands is the most practical and appropriate solution for dimensionality reduction of hyperspectral images. However, feature selection with supervised learning techniques requires a large labeled dataset to identify the best informative bands. Yield study was a regression problem with limited number of samples, about 1000 plots each with one target value (i.e., measured yield). Therefore, a deep autoencoder network was used in this study to learn a new representation of hyperspectral images in a low dimensional feature space through a self-supervised learning approach by leveraging the large amounts of unlabeled pixels. Identifying a suitable representation oftentimes makes the subsequent learning processes more interpretable and less complex.

The results achieved in the present study revealed that the trained autoencoder could effectively project the high-dimensional hyperspectral images onto a 3-, 5-, and 10-dimensional feature space with mean squared error of 7e-5, 6e-5, and 5e-5, respectively. In addition, the reconstructed data appeared to be less noisy compared to the input data. While the encoder reduced the dimensionality, the decoder eliminated the noise from the input data.

The results of training deep neural networks on the low dimensional datasets demonstrated that the dimension of hyperspectral images could be significantly reduced without deteriorating the performance of the networks on yield prediction. Furthermore, a low dimensional dataset reduces the required sample size, risk of overfitting, complexity of interpretation, and running time, because the number of model parameters (weights and biases) that should be tuned becomes significantly fewer. For instance, if the hyperspectral dataset in Chapter 5 with 381 features (190+190+1) is used as the input layer of a fully-connected DNN model with 10 units in the first layer, there will be 3820 ((381×10) + 10) parameters that should be trained. However, if the 5-dimensional dataset (5+5+1) is used as the input layer of the same DNN model, the number of parameters will be 120 ((11×10)+10), significantly lower compared to using the hyperspectral image dataset.

CHAPTER 7

Conclusion

Developing crop varieties with desired traits through plant breeding and genetics is crucial in promoting crop production to meet the projected demand for food. To improve genetic gain for developing advanced lines, it is required to investigate the breeding and genotyping data in association with the corresponding information acquired by plant phenotyping.

This dissertation focused on applying innovative technologies, automation, and artificial intelligence to develop sensor-based, intelligent frameworks for high-throughput phenotyping of three major wheat traits: yield potential, salt stress tolerance, and *Fusarium* head blight resistance.

In this dissertation, several disciplines were integrated to develop automated phenotyping frameworks for the desired traits in wheat, tailored around improving wheat production. These disciplines include remote sensing (radiometric calibration of hyperspectral images, designing and execution of missions for aerial imagery), computer vision (segmentation of the object of interest from background, eliminating mixed pixels, morphological operations for processing binary images), machine learning/deep learning (feature selection, feature learning, quantitative ranking, making prediction), and plant science (defining the problems, collecting ground truth data, incorporating domain knowledge).

7.1 Summary of projects

Project 1 was conducted with collaboration of the USDA cereal disease laboratory to identify wheat varieties tolerant to salt stress as early as one day after applying salt treatment. In this project, a sensor-based phenotyping framework was developed by leveraging the unique advantages of hyperspectral imaging in providing valuable information about the internal activities of plants. By integrating image processing, optimization and machine learning algorithms, we were able to quantitatively rank wheat varieties based on their tolerance to salt stress in a non-invasive manner only one day after salt application when there were no visual symptoms. The quantitative ranking of salt stress tolerance helps breeders integrate salt tolerance results with other desired traits (e.g., grain yield) to mitigate yield losses due to salinity, and to ultimately maintain or improve production on saline soils.

Project 2 focused on selecting the most informative spectral bands from high-dimensional hyperspectral images to mitigate the challenges associated with analysis of hyperspectral images. In this project, an ensemble feature selection pipeline was developed to aggregate the benefits of multiple feature selection methods. The performance of algorithm was evaluated on the salt stress dataset. The proposed method could drastically reduce the dimension of hyperspectral images from 215 to 15 while improving the accuracy of classifying healthy and stressed vegetation pixels by 8.5%. Furthermore, we were able to determine the center of six broad bands that returned similar ranking of wheat lines obtained by analysis of hyperspectral dataset. Such a discovery can substantially contribute in the development of custom-designed multispectral sensors for detecting crop stresses. A multispectral camera designed for a specific application can benefit scientists, farmers, agricultural practitioners, and field consultants across a wide range of use cases.

Project 3 was performed with collaboration of Dr. James Anderson's lab and Dr. Brian Steffenson's lab to develop a framework for *Fusarium* head blight (FHB) phenotyping. FHB is among the most widespread and devastating diseases of small grain crops throughout the world. Besides the negative impact on yield and grain quality, the FHB pathogens produce mycotoxins, such as deoxynivalenol (DON), that contaminate the grain,

making it unsuitable for human and livestock consumption. In this study, the most informative spectral bands for classifying healthy spikes and FHB-infected spikes were identified using the ensemble feature selection method developed in Project 2. The results of this study revealed that a set of two broad spectral bands (766 nm and 696 nm) returns the maximum classification accuracy of healthy spikes from FHB infected ones.

Project 4 aimed to develop an autonomous robotic framework for high-throughput yield phenotyping of wheat in the field through collaboration with Dr. James Anderson's research group. The data were captured by a hyperspectral camera mounted on an unmanned aerial vehicle (UAV) flying over three experimental wheat plots during two consecutive growing seasons. In this study, a deep neural network was trained to predict the yield of wheat plots. In addition to yield prediction at plot scale, the findings of this study revealed that the proposed method was able to accurately predict the yield at a finer spatial resolution (i.e., sub-plot scale) as well as a coarser spatial resolution (i.e., field scale). In addition to yield estimation, this study can benefit plant scientists and breeders in various ways by providing further insights into the process of screening. These benefits include: (i) optimizing the plot size to mitigate the alley effects on yield, (ii) investigating the effect of plant density on yield in such a high spatial resolution, (iii) studying the impact of side trimming on yield across various varieties, and (iv) facilitating the process of visual assessment of the plots by collecting images at high spectral, spatial, and temporal resolution.

Project 5 focused on developing a deep autoencoder network to learn an optimal feature representation of hyperspectral images in a low dimensional feature space for yield prediction. In this study, a deep fully-connected autoencoder network was trained on a large unlabeled dataset (2 million pixels) to project the hyperspectral images onto a 3-, 5-, and 10-dimenionsal feature space. The result demonstrated that the trained autoencoder could substantially reduce the dimension of hyperspectral images with minimum error, while retaining the relevant information for yield prediction.

7.2 Broader potential impacts

Although the underlying aim of this dissertation was to apply automation and artificial intelligence to phenotyping, it potentially has long-term benefits for scientists, growers, and other stakeholders. At a higher level, this work contributes to improving economic, ecological, and social impacts by improving crop production, reducing pesticides use, and properly leveraging salt-affected farmlands. From an environmental perspective, a cultivar with high yield potential and a cultivar resistant to FHB disease both promote sustainability in crop production and environment by reducing the required fertilizer and pesticide to meet the anticipated farmers' profit. From food security perspective, identifying crop varieties tolerant to salt stress, resistant to diseases, and with high yield potential can promote crop production, and subsequently, mitigate food security concerns.

The analytical tools established through this study can be deployed for other applications in addition to plant phenotyping. For instance, the pipeline proposed in yield study could be applied for yield prediction in large-scale agricultural fields. Yield prediction at large scale can provide insights into crop management, business decisions (e.g., crop insurance), and policymaking. Furthermore, the feature selection method developed for selecting the most informative bands can be deployed for feature selection, and subsequently, dimensionality reduction of other types of high-dimensional datasets (e.g., microarray datasets). In addition, this feature selection framework can be utilized to extend the digital library, initiated in this dissertation, by identifying spectral features associated with various crop diseases and stresses. This discovery of the most discriminative spectral bands for detecting crop diseases can be a foundation for developing custom-designed sensors that can be mounted on advanced autonomous vehicles to generate precise disease maps in large-scale fields, aimed at leveraging variable rate technologies for pesticides spraying at hot spots.

7.3 Future work

Based on the findings of this dissertation, we suggest several directions for potential future work per each project:

Salt stress project

The proposed framework can be used for salt stress phenotyping of a large number of wheat lines. For such a project, we suggest to evaluate the performance of the two proposed techniques (Bayesian approach and MDPA) on several wheat lines based on the conventional phenotyping and/or historical knowledge. Once the robustness and performance of each technique was determined, a large number of wheat lines can be assessed in a short time period in greenhouse because these techniques are able to discriminant the difference between the salt stress tolerance of wheat lines as early as one day after applying salt treatment. Furthermore, the results of such a project can be incorporated with the findings of phenotyping for other desired traits, such as yield, to identify a variety with multiple desired traits; for instance, a cultivar with high tolerance to salt stress as well as high yield potential.

Feature selection project

The proposed ensemble feature selection pipeline can be used to identify the most informative bands for other crop stresses/diseases and extend the digital spectral library initiated in this dissertation. For instance, the proposed feature selection method can be utilized to determine an optimal set of spectral bands for nitrogen deficiency detection in corn to alleviate the nitrogen loss that contaminates ground/surface water.

FHB (scab) project

The results of the FHB project suggest that a RGB or a multispectral camera with three bands are adequate for FHB detection. We recommend training a Convolutional Neural Network model on a large annotated RGB dataset. To facilitate the process of gathering a large annotated data, transfer learning can be used such that only the last few layers of the

network is tuned since the weights and biases of the Convolutional layers are borrowed from a network already trained on an available large annotated dataset. As of now, there are two annotated datasets that can be sued for transfer learning to detect the wheat spikes. One of the dataset is for detection of wheat spikes (Hasan, Chopin, Laga, & Miklavcic, 2018), and the other is for locating spikes and spikelets (Pound, Atkinson, Wells, Pridmore, & French, 2017).

Yield prediction project

For yield project, we recommend incorporating a temporal component and evaluate the feasibility of yield prediction at various phases of plant development as well as various locations. An accurate yield prediction at earlier growth stage could provide invaluable information for making appropriate agronomic decisions such as nutrient management.

Feature learning by deep autoencoder

Supervised learning techniques can be used to find a set of informative bands for yield prediction. Since these methods require large labeled datasets, we suggest increasing the number of samples in yield dataset by using the methodology proposed for dividing the plots into sub-plots and assigning yield to each of sub-plots. Subsequent to identifying the optimal set of bands, the new dataset can be used for training a deep neural network to evaluate the performance of the selected bands in yield prediction.

References

Aase, J. K., & Siddoway, F. H. (1981). Spring Wheat Yield Estimates from Spectral Reflectance Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-19*(2), 78–84. https://doi.org/10.1109/TGRS.1981.350356

Abadi, M., Ashish Agarwal, Paul Barham, E. B., Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, … Yuan Yu. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Retrieved from https://www.tensorflow.org/

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., & Saeys, Y. (2009). Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics*, *26*(3), 392–398. https://doi.org/10.1093/bioinformatics/btp630

Acreche, M. M., Briceño-Félix, G., Sánchez, J. A. M., & Slafer, G. A. (2008). Physiological bases of genetic gains in Mediterranean bread wheat yield in Spain. *European Journal of Agronomy*, *28*(3), 162–170. https://doi.org/10.1016/J.EJA.2007.07.001

Agarwal, P. K., Shukla, P. S., Gupta, K., & Jha, B. (2013). Bioengineering for Salinity Tolerance in Plants: State of the Art. *Molecular Biotechnology*, *54*(1), 102–123. https://doi.org/10.1007/s12033-012-9538-3

Aktar, M. W., Sengupta, D., & Chowdhury, A. (2009). Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary Toxicology*, *2*(1), 1–12. https://doi.org/10.2478/v10102-009-0001-7

Alisaac, E., Behmann, J., Kuska, M. T., Dehne, H. W., & Mahlein, A. K. (2018). Hyperspectral quantification of wheat resistance to Fusarium head blight: comparison of two Fusarium species. *European Journal of Plant Pathology*, 1–16. https://doi.org/10.1007/s10658-018-1505-9

Anderson, J. A., Stack, R. W., Liu, S., Waldron, B. L., Fjeld, A. D., Coyne, C., … Frohberg, R. C. (2001). DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, *102*(8), 1164–1168. https://doi.org/10.1007/s001220000509

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. *Trends in Plant Science*, *23*(5), 451–466. https://doi.org/10.1016/J.TPLANTS.2018.02.001

Araus, J. L., Li, J., Parry, M. A. J., & Wang, J. (2014). Phenotyping and other breeding approaches for a New Green Revolution. *Journal of Integrative Plant Biology*, *56*(5), 422–424. https://doi.org/10.1111/jipb.12202

Archibald, R., & Fann, G. (2007). Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and Remote Sensing Letters*, *4*(4), 674–677. https://doi.org/10.1109/LGRS.2007.905116

Aslett, Z., Taranik, J. V., & Riley, D. N. (2018). Mapping rock forming minerals at Boundary Canyon, Death Valley National Park, California, using aerial SEBASS thermal infrared hyperspectral image data. *International Journal of Applied Earth Observation and Geoinformation*, *64*(October 2017), 326–339. https://doi.org/10.1016/j.jag.2017.08.001

Asrar, G., Fuchs, M., Kanemasu, E. T., & Hatfield, J. L. (1984). Estimating Absorbed Photosynthetic Radiation and Leaf Area Index from Spectral Reflectance in Wheat1. *Agronomy Journal*, *76*(2), 300. https://doi.org/10.2134/agronj1984.00021962007600020029x

Barbedo, J. G. A., Tibola, C. S., & Fernandes, J. M. C. (2015). Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. *Biosystems Engineering*, *131*, 65–76. https://doi.org/10.1016/j.biosystemseng.2015.01.003

Baret, F., Guyot, G., & Major, D. J. (1989). TSAVI: A Vegetation Index Which Minimizes Soil Brightness Effects On LAI And APAR Estimation. In *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium,* (Vol. 3, pp. 1355–1358). IEEE. https://doi.org/10.1109/IGARSS.1989.576128

Bauckhage, C., & Kersting, K. (2013). Data Mining and Pattern Recognition in Agriculture. *KI - Künstliche Intelligenz*, *27*(4), 313–324. https://doi.org/10.1007/s13218-013-0273-0

Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U., & Herppich, W. B. (2011). Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture*, *75*(2), 304–312. https://doi.org/10.1016/j.compag.2010.12.006

Bauriegel, E., & Herppich, W. (2014). Hyperspectral and Chlorophyll Fluorescence Imaging for Early Detection of Plant Diseases, with Special Reference to Fusarium spec. Infections on Wheat. *Agriculture*, *4*(1), 32–57. https://doi.org/10.3390/agriculture4010032

Beaulieu, R. J., Goldstein, S. D., Singh, J., Safar, B., Banerjee, A., & Ahuja, N. (2018). Automated diagnosis of colon cancer using hyperspectral sensing. *The International Journal of Medical Robotics and Computer Assisted Surgery*, (March 2017), e1897. https://doi.org/10.1002/rcs.1897

Bedini, E., & Rasmussen, T. M. (2018). Use of airborne hyperspectral and gamma-ray spectroscopy data for mineral exploration at the Sarfartoq carbonatite complex, southern West Greenland. *Geosciences Journal*, *22*(4), 1–11. https://doi.org/10.1007/s12303-017-0078-5

Behmann, J., Mahlein, A. K., Paulus, S., Kuhlmann, H., Oerke, E. C., & Plümer, L. (2015). Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2015.05.010

Ben Brahim, A., & Limam, M. (2017). Ensemble feature selection for high dimensional data: a new method and a comparative study. *Advances in Data Analysis and Classification*, 1–16. https://doi.org/10.1007/s11634-017-0285-y

Benedetti, R., & Rossini, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment*, *45*(3), 311–326. https://doi.org/10.1016/0034-4257(93)90113-C

Biswal, B. (1995). Carotenoid catabolism during leaf senescence and its control by light. *Journal of Photochemistry & Photobiology. B - Biology*, *30*(1), 3–13.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. https://doi.org/10.1016/J.ISPRSJPRS.2009.06.004

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burkart, A., Hecht, V. L., Kraska, T., & Rascher, U. (2017). Phenological analysis of unmanned aerial vehicle based time series of barley imagery with high temporal resolution. *Precision Agriculture*, *19*(1), 1–13. https://doi.org/10.1007/s11119-017-9504-y

Carillo, P., Annunziata, M. G., Pontecorvo, G., Fuggi, A., & Woodrow, P. (2011). Salinity stress and salt tolerance. In A. K. Shanker & B. Venkateswarlu (Eds.), *Abiotic Stress in Plants – Mechanisms and Adaptations* (Vol. 1, pp. 21–38). Rijeka: InTech. https://doi.org/10.5772/22331

Centers for Disease Control and Prevention. (2014). Pesticide and Health. Retrieved October 15, 2016, from http://ephtracking.cdc.gov/showPesticidesHealth.action

Cha, S. H., & Srihari, S. N. (2002). On measuring the distance between histograms. *Pattern Recognition*, *35*(6), 1355–1370. https://doi.org/10.1016/S0031-3203(01)00118-2

Chan, J. C. W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*(6), 2999–3011. https://doi.org/10.1016/j.rse.2008.02.011

Chan, T. H., Ma, W. K., Ambikapathi, A., & Chi, C. Y. (2011). A simplex volume maximization framework for hyperspectral endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11 PART 1), 4177–4193. https://doi.org/10.1109/TGRS.2011.2141672

Chang, C. I., Wu, C. C., & Tsai, C. T. (2011). Random N-finder (N-FINDR) endmember extraction algorithms for hyperspectral imagery. *IEEE Transactions on Image Processing*, *20*(3), 641–656. https://doi.org/10.1109/TIP.2010.2071310

Chaparro, D., Piles, M., Vall-llossera, M., Camps, A., Konings, A. G., & Entekhabi, D. (2018). L-band vegetation optical depth seasonal metrics for crop yield assessment. *Remote Sensing of Environment*, *212*(April), 249–259. https://doi.org/10.1016/j.rse.2018.04.049

Chollet, F., & others. (2015). Keras. Retrieved from https://keras.io

Chutia, D., Bhattacharyya, D. K., Sarma, J., & Raju, P. N. L. (2017). An effective ensemble classification framework using random forests and a correlation based feature selection technique. *Transactions in GIS*, *21*(6), 1165–1178. https://doi.org/10.1111/tgis.12268

Coleman-Jensen, A., Rabbitt, M. P., Gregory, C. A., Singh, G., & Singh, A. (2018). Household Food Security in the United States in 2017. *Household Food Security in the United States in 2017, ERR-256, U.S. Department of Agriculture, Economic Research Service*. https://doi.org/10.2139/ssrn.2504067

Condorelli, G. E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J. W., French, A. N., … Tuberosa, R. (2018). Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. *Frontiers in Plant Science*, *9*(June), 1–17. https://doi.org/10.3389/fpls.2018.00893

Coppens, F., Wuyts, N., Inzé, D., & Dhondt, S. (2017). Unlocking the potential of plant phenotyping data through integration and data-driven approaches. *Current Opinion in Systems Biology*, *4*, 58–63. https://doi.org/10.1016/j.coisb.2017.07.002

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., & Poland, J. (2018). Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. *The Plant Genome*, *11*(1), 0. https://doi.org/10.3835/plantgenome2017.05.0043

Damodaran, B. B., Courty, N., & Lefevre, S. (2017). Sparse Hilbert Schmidt Independence Criterion and Surrogate-Kernel-Based Feature Selection for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(4), 2385–2398. https://doi.org/10.1109/TGRS.2016.2642479

De La Fuente, G. N., Frei, U. K., & Lübberstedt, T. (2013). Accelerating plant breeding. *Trends in Plant Science*, *18*(12), 667–672. https://doi.org/10.1016/J.TPLANTS.2013.09.001

Delwiche, S. R., & Kim, M. S. (2000). <title>Hyperspectral imaging for detection of scab in wheat</title>, *4203*, 13–20. https://doi.org/10.1117/12.411752

Delwiche, S. R., Kim, M. S., & Dong, Y. (2011). Fusarium damage assessment in wheat kernels by Vis/NIR hyperspectral imaging. *Sensing and Instrumentation for Food Quality and Safety*, *5*(2), 63–71. https://doi.org/10.1007/s11694-011-9112-x

Du, M., & Noguchi, N. (2017). Monitoring of wheat growth status and mapping of wheat yield's within-field spatial variations using color images acquired from UAV-camera System. *Remote Sensing*, *9*(3). https://doi.org/10.3390/rs9030289

Duan, T., Chapman, S. C., Guo, Y., & Zheng, B. (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research*, *210*(May), 71–80. https://doi.org/10.1016/j.fcr.2017.05.025

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, *12*, 2121–2159. https://doi.org/10.1109/CDC.2012.6426698

Duvick, D. N. (2005). The Contribution of Breeding to Yield Advances in maize (Zea mays L.). *Advances in Agronomy*, *86*, 83–145. https://doi.org/10.1016/S0065-2113(05)86002-X

FAO and ITPS. (2015). *Status of the World's Soil Resources (SWSR) – Main Report.* Roma, Italy: Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils.

FAO, IFAD, UNICEF, WFP, & WHO. (2018). *The State of Food Security and Nutrition in the World 2018*. *Global Food Insecurity Report*. Rome: Building climate resilience for food security and nutrition. https://doi.org/10.1093/cjres/rst006

Feng, X., Yu, C., Liu, X., Chen, Y., Zhen, H., Sheng, K., & He, Y. (2018). Nondestructive and rapid determination of lignocellulose components of biofuel pellet using online hyperspectral imaging system. *Biotechnology for Biofuels*, *11*(1), 1–12. https://doi.org/10.1186/s13068-018-1090-3

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, *161*(2), 165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1994). 1996.Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS.pdf, *4257*(96).

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Vol. 9, pp. 249–256). Chia Laguna Resort, Sardinia, Italy: PMLR. Retrieved from http://proceedings.mlr.press/v9/glorot10a.html

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In G. Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (Vol. 15, pp. 315–323). Fort Lauderdale, FL, USA: PMLR. Retrieved from http://proceedings.mlr.press/v15/glorot11a.html

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press. Retrieved from http://www.deeplearningbook.org

Grover, A., Sabat, S. C., & Mohanty, P. (1986). Relative sensitivity of various spectral forms of photosynthetic pigments to leaf senescence in wheat (Triticum aestivum L.). *Photosynthesis Research*, *10*(3), 223–231. https://doi.org/10.1007/BF00118287

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46*(1), 389–422. https://doi.org/10.1023/A:1012487302797

Hall, M. A. (2000). Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 359–366). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=645529.657793

Han, L., Yang, G., Yang, H., Xu, B., Li, Z., & Yang, X. (2018). Clustering Field-Based Maize Phenotyping of Plant-Height Growth and Canopy Spectral Dynamics Using a UAV Remote-Sensing Approach. *Frontiers in Plant Science*, *9*(November), 1–18. https://doi.org/10.3389/fpls.2018.01638

Hasan, M. M., Chopin, J. P., Laga, H., & Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods*, *14*(1), 100. https://doi.org/10.1186/s13007-018-0366-8

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*. *Springer Series in Statistics*. New York: Springer-Verlag New York. https://doi.org/10.1007/b94608

Hira, Z. M., & Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*, (1). https://doi.org/10.1155/2015/198363

Hwang, C.-L., & Yoon, K. (1981). *Methods for Multiple Attribute Decision Making*. Berlin: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-48318-9_3

Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? *Proceedings of the IEEE International Conference on Computer Vision*, 2146–2153. https://doi.org/10.1109/ICCV.2009.5459469

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment*, *198*, 105–114. https://doi.org/10.1016/j.rse.2017.06.007

Jovic, A., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1200–1205). Opatija, Croatia. https://doi.org/10.1109/MIPRO.2015.7160458

Kalaji, H. M., Rastogi, A., Živčák, M., Brestic, M., Daszkowska-Golec, A., Sitko, K., … Cetner, M. D. (2018). Prompt chlorophyll fluorescence as a tool for crop phenotyping: an example of barley landraces exposed to various abiotic stress factors. *Photosynthetica*, *56*(3), 953–961. https://doi.org/10.1007/s11099-018-0766-z

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/J.BUSHOR.2018.08.004

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *CoRR*, *abs/1412.6*. Retrieved from http://dblp.uni-trier.de/db/journals/corr/corr1412.html#KingmaB14

Kira, K., & Rendell, L. A. (1992). A Practical Approach to Feature Selection. In *Proceedings of the Ninth International Workshop on Machine Learning* (pp. 249–256). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=141975.142034

Koebner, R. M. D., Martin, P. K., Orford, S. M., Miller, T. E., & Lupton, F. G. H. (1996). Responses to salt stress controlled by the homoeologous group 5 chromosomes of hexaploid wheat. *Plant Breeding*, *115*, 81–84.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Kononenko, I., Šimec, E., & Robnik-Šikonja, M. (1997). Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, *7*(1), 39–55. https://doi.org/10.1023/A:1008280620621

Koonsanit, K., Jaruskulchai, C., & Eiumnoh, A. (2012). Band Selection for Dimension Reduction in Hyper Spectral Image Using Integrated Information Gain and Principal Components Analysis Technique. *International Journal of Machine Learning and Computing*, *2*(3), 248–251.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kuncheva, L. I. (2014). Ensemble Feature Selection. In *Combining Pattern Classifiers: Methods and Algorithms* (Second Edi, pp. 290–325). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118914564.ch9

Kuska, M., Wahabzada, M., Leucker, M., Dehne, H.-W., Kersting, K., Oerke, E.-C., … Mahlein, A.-K. (2015). Hyperspectral phenotyping on the microscopic scale: towards automated characterization of plant-pathogen interactions. *Plant Methods*, *11*(28), 28. https://doi.org/10.1186/s13007-015-0073-7

Kyratzis, A. C., Skarlatos, D. P., Menexes, G. C., Vamvakousis, V. F., & Katsiotis, A. (2017). Assessment of Vegetation Indices Derived by UAV Imagery for Durum Wheat Phenotyping under a Water Limited and Heat Stressed Mediterranean Environment. *Frontiers in Plant Science*, *8*(June), 1–14. https://doi.org/10.3389/fpls.2017.01114

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R., & Long, D. S. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, *23*(20), 4169–4180. https://doi.org/10.1080/01431160110107653

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). Handwritten Digit Recognition with a Back-Propagation Network. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 396–404). Morgan-Kaufmann. Retrieved from http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf

Li, G., Wan, S., Zhou, J., Yang, Z., & Qin, P. (2010). Leaf chlorophyll fluorescence , hyperspectral reflectance , pigments content , malondialdehyde and proline accumulation responses of castor bean (Ricinus communis L .) seedlings to salt stress levels, *31*, 13–19. https://doi.org/10.1016/j.indcrop.2009.07.015

Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors*, *14*(11), 20078–20111. https://doi.org/10.3390/s141120078

Liberatore, K. L., Dukowic-schulze, S., Miller, M. E., Chen, C., & Kianian, S. F. (2016). Free Radical Biology and Medicine The role of mitochondria in plant development and stress tolerance. *Free Radical Biology and Medicine*, *100*, 238–256. https://doi.org/10.1016/j.freeradbiomed.2016.03.033

Lichtenthaler, H. K. (1987). [34] Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods in Enzymology*, *148*, 350–382. https://doi.org/10.1016/0076-6879(87)48036-1

Liu, Z., Wang, H., & Li, Q. (2012). Tongue tumor detection in medical hyperspectral images. *Sensors*, *12*(1), 162–174. https://doi.org/10.3390/s120100162

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008

Lopes, M. S., Reynolds, M. P., Manes, Y., Singh, R. P., Crossa, J., & Braun, H. J. (2012). Genetic Yield Gains and Changes in Associated Traits of CIMMYT Spring Bread Wheat in a "Historic" Set Representing 30 Years of Breeding. *Crop Science*, *52*(3), 1123. https://doi.org/10.2135/cropsci2011.09.0467

Lu, C., Lu, Q., Zhang, J., & Kuang, T. (2001). Characterization of photosynthetic pigment composition, photosystem II photochemistry and thermal energy dissipation during leaf senescence of wheat plants grown in the field. *Journal of Experimental Botany*, *52*, 1805–1810. https://doi.org/10.1016/0893-6080(88)90069-X

Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, *19*(1), 010901. https://doi.org/10.1117/1.JBO.19.1.010901

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., … Comar, A. (2017). High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Frontiers in Plant Science*, *8*(November), 1–14. https://doi.org/10.3389/fpls.2017.02002

Mahlein, A.-K., Steiner, U., Hillnhütter, C., Dehne, H.-W., & Oerke, E.-C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods*, *8*(1), 3. https://doi.org/10.1186/1746-4811-8-3

Mahlein, A. K., Oerke, E. C., Steiner, U., & Dehne, H. W. (2012). Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, *133*(1), 197–209. https://doi.org/10.1007/s10658-011-9878-z

Makdessi, N. Al, Jean, P. A., Ecarnot, M., Gorretta, N., Rabatel, G., & Roumet, P. (2017). How plant structure impacts the biochemical leaf traits assessment from in-field hyperspectral images: A simulation study based on light propagation modeling in 3D virtual wheat scenes. *Field Crops Research*, *205*, 95–105. https://doi.org/10.1016/j.fcr.2017.02.001

Masia, F., Pope, I., Watson, P., Langbein, W., & Borri, P. (2018). Bessel-Beam Hyperspectral CARS Microscopy with Sparse Sampling: Enabling High-Content High-Throughput Label-Free Quantitative Chemical Imaging. *Analytical Chemistry*, *90*(6), 3775–3785. https://doi.org/10.1021/acs.analchem.7b04039

Mbarki, S., Sytar, O., Cerda, A., Zivcak, M., Rastogi, A., He, X., … Abdelly, C. (2018). Strategies to Mitigate the Salt Stress Effects on Photosynthetic Apparatus and Productivity of Crop Plants (Vol. 1). Springer, Cham. https://doi.org/10.1007/978-3-319-75671-4_4

Moghimi, A., Hossein Aghkhani, M., Reza Golzarian, M., Rohani, A., & Yang, C. (2015). A Robo-vision Algorithm for Automatic Harvesting of Green Bell Pepper. In *ASABE Annual International Meeting*. New Orleans, United States: ASABE. https://doi.org/https://doi.org/10.13031/aim.20152189355

Moghimi, A., & Yang, C. (2018). Hyperspectral image dataset for salt stress phenotyping of wheat. https://doi.org/10.13020/D69Q3K

Moghimi, A., Yang, C., & Marchetto, P. M. (2018). Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging. *IEEE Access*, *6*, 56870–56884. https://doi.org/10.1109/ACCESS.2018.2872801

Moghimi, A., Yang, C., Miller, M. E., Kianian, S. F., & Marchetto, P. M. (2018). A Novel Approach to Assess Salt Stress Tolerance in Wheat Using Hyperspectral Imaging. *Frontiers in Plant Science*, *9*, 1182. https://doi.org/10.3389/fpls.2018.01182

Moghimi, A., Yang, C., Miller, M. E., Kianian, S., & Marchetto, P. (2017). Hyperspectral imaging to identify salt-tolerant wheat lines. In *SPIE. Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II 2017*. Anaheim, California, United States. https://doi.org/10.1117/12.2262388

Mohd Asaari, M. S., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., & Scheunders, P. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, *138*, 121–138. https://doi.org/10.1016/j.isprsjprs.2018.02.003

Munns, R., & James, R. A. (2003). Screening methods for salinity tolerance: A case study with tetraploid wheat. *Plant and Soil*, *253*(1), 201–218. https://doi.org/10.1023/A:1024553303144

Munns, R., James, R. A., & Läuchli, A. (2006). Approaches to increasing the salt tolerance of wheat and other cereals. *Journal of Experimental Botany*, *57*(5), 1025–1043. https://doi.org/10.1093/jxb/erj100

Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., & Ganapathysubramanian, B. (2017). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean, 1–20. Retrieved from http://arxiv.org/abs/1710.04681

Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 807–814). USA: Omnipress. Retrieved from http://dl.acm.org/citation.cfm?id=3104322.3104425

Nxele, X., Klein, A., & Ndimba, B. K. (2017). Drought and salinity stress alters ROS accumulation, water retention, and osmolyte content in sorghum plants. *South African Journal of Botany*, *108*, 261–266. https://doi.org/10.1016/J.SAJB.2016.11.003

O'Donnell, K., Kistler, H. C., Cigelnik, E., Ploetz, R. C., & Taylor, J. W. (1998). Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proceedings of the National Academy of Sciences*, *95*(5), 2044–2049. https://doi.org/10.1073/pnas.95.5.2044

Ondrasek, G., Rengal, Z., & Veres, S. (2011). Soil salinisation and salt stress in crop production. In A. K. Shanker & B. Venkateswarlu (Eds.), *Abiotic stress in plants - Mechanisms and adaptations* (pp. 171–190). Rijeka: InTech. https://doi.org/10.5772/22248

Pavurala, N., Xu, X., & Krishnaiah, Y. S. R. (2017). Hyperspectral imaging using near infrared spectroscopy to monitor coat thickness uniformity in the manufacture of a transdermal drug delivery system. *International Journal of Pharmaceutics*, *523*(1), 281–290. https://doi.org/10.1016/j.ijpharm.2017.03.022

Peleman, J. D., & van der Voort, J. R. (2003). Breeding by Design. *Trends in Plant Science*, *8*(7), 330–334. https://doi.org/10.1016/S1360-1385(03)00134-1

Polder, G., Van Der Heijden, G. W. A. M., Waalwijk, C., & Young, I. T. (2005). Detection of Fusarium in single wheat kernels using spectral imaging. *Seed Science and Technology*, *33*(3), 655–668. https://doi.org/10.15258/sst.2005.33.3.13

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., & French, A. P. (2017). Deep Learning for Multi-task Plant Phenotyping. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (pp. 2055–2063). IEEE. https://doi.org/10.1109/ICCVW.2017.241

Qian, Y., Ye, M., & Zhou, J. (2013). Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 2276–2291. https://doi.org/10.1109/TGRS.2012.2209657

Rady, A., Ekramirad, N., Adedeji, A. A., Li, M., & Alimardani, R. (2017). Hyperspectral imaging for detection of codling moth infestation in GoldRush apples. *Postharvest Biology and Technology*, *129*, 37–44. https://doi.org/10.1016/j.postharvbio.2017.03.007

Rawat, N., Pumphrey, M. O., Liu, S., Zhang, X., Tiwari, V. K., Ando, K., … Gill, B. S. (2016). Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. *Nature Genetics*, *48*(12), 1576–1580. https://doi.org/10.1038/ng.3706

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066428

Reynolds, M., FOULKES, J., FURBANK, R., GRIFFITHS, S., KING, J., MURCHIE, E., … SLAFER, G. (2012). Achieving yield gains in wheat. *Plant, Cell & Environment*, *35*(10), 1799–1823. https://doi.org/10.1111/j.1365-3040.2012.02588.x

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third ERTS Symposium* (Vol. 1, pp. 309–317). Washington DC OR - NASA. Retrieved from citeulike-article-id:9507328

Sadras, V. O., & Lawson, C. (2011). Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. *Crop and Pasture Science*, *62*(7), 533. https://doi.org/10.1071/CP11060

Saeys, Y., Abeel, T., & de Peer, Y. (2008). Robust Feature Selection Using Ensemble Feature Selection Techniques. In W. Daelemans, B. Goethals, & K. Morik (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 313–325). Berlin, Heidelberg: Springer Berlin Heidelberg.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

Schachtman, D. P., Lagudah, E. S., & Munns, R. (1992). The expression of salt tolerance from Triticum tauschii in hexaploid wheat. *Theoretical and Applied Genetics*, *84*(5–6), 714–719. https://doi.org/10.1007/BF00224174

Schmid, T., Rodriguez-Rastrero, M., Escribano, P., Palacios-Orueta, A., Ben-Dor, E., Plaza, A., … Chabrillat, S. (2016). Characterization of Soil Erosion Indicators Using Hyperspectral Data from a Mediterranean Rainfed Cultivated Region. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(2), 845–860. https://doi.org/10.1109/JSTARS.2015.2462125

Schowengerdt, R. A. (2012). *Remote sensing: Models and methods for image processing: Second edition*. *Remote Sensing: Models and Methods for Image Processing: Second Edition*. https://doi.org/10.1016/C2009-0-21902-7

Shahin, M. A., & Symons, S. J. (2011). Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Computers and Electronics in Agriculture*, *75*(1), 107–112. https://doi.org/10.1016/j.compag.2010.10.004

Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food and Energy Security*, *4*(3), 178–202. https://doi.org/10.1002/fes3.64

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., … Yang, C. (2016). Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. *PloS One*, *11*(7), e0159781. https://doi.org/10.1371/journal.pone.0159781

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, *5*(3), 291–317. https://doi.org/10.1007/s12571-013-0263-y

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science*, *23*(10), 883–898. https://doi.org/10.1016/J.TPLANTS.2018.07.004

Sobota, D. J., Compton, J. E., McCrackin, M. L., & Singh, S. (2015). Cost of reactive nitrogen release from human activities to the environment in the United States. *Environmental Research Letters*, *10*(2), 025006. https://doi.org/10.1088/1748-9326/10/2/025006

Sun, Z. W., Ren, L. K., Fan, J. W., Li, Q., Wang, K. J., Guo, M. M., … Li, X. (2016). Salt response of photosynthetic electron transport system in wheat cultivars with contrasting tolerance. *Plant, Soil and Environment*, *62*, 515–521. https://doi.org/10.17221/529/2016-PSE

Sytar, O., Brestic, M., Zivcak, M., Olsovska, K., Kovar, M., Shao, H., & He, X. (2017). Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance. *Science of the Total Environment*, *578*(August 2016), 90–99. https://doi.org/10.1016/j.scitotenv.2016.08.014

Tang, J., Alelyani, S., & Liu, H. (2014). *Feature selection for classification: A review*. *Data Classification: Algorithms and Applications.* CRC Press. Retrieved from https://www.crcpress.com/Data-Classification-Algorithms-and-Applications/Aggarwal/9781466586741%5Cnhttp://eprints.kku.edu.sa/170/%5Cnhttp://eprints.kku.edu.sa/170/1/feature_selection_for_classification.pdf

Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., & Mahlein, A.-K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, *14*(1), 45. https://doi.org/10.1186/s13007-018-0313-8

Thorp, K. R., Wang, G., Bronson, K. F., Badaruddin, M., & Mon, J. (2017). Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. *Computers and Electronics in Agriculture*, *136*, 1–12. https://doi.org/10.1016/j.compag.2017.02.024

Thurau, C., Kersting, K., Bauckhage, C., Iais, F., & Augustin, S. (2010). Yes We Can – Simplex Volume Maximization for Descriptive Web-Scale Matrix Factorization Categories and Subject Descriptors. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM,* (pp. 1785–1788). Toronto, Ontario, Canada.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*. https://doi.org/10.2307/2346178

Tieleman, T., & Hinton, G. (2012). Lecture 6.5---RmsProp: Divide the gradient by a running average of its recent magnitude. Retrieved from https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–4. https://doi.org/10.1073/pnas.1116437108

Tsunewaki, K., Wang, G.-Z., & Matsuoka, Y. (2002). Plasmon analysis of Triticum (wheat) and Aegilops. 2. Characterization and classification of 47 plasmons based on their effects on common wheat phenotype. *Genes & Genetic Systems*, *77*(6), 409–27. https://doi.org/10.1266/ggs.77.409

Tsunewaki, K., Wang, G. Z., & Matsuoka, Y. (1996). Plasmon analysis of Triticum (wheat) and Aegilops. 1. Production of alloplasmic common wheats and their fertilities. *Genes & Genetic Systems*. https://doi.org/10.1266/ggs.71.293

United Nations, Department of Economic and Social Affairs, P. D. (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. New York. https://doi.org/Working Paper No. ESA/P/WP/248

United States Department of Agriculture. (2018a). *Minnesota Ag News – 2018 Spring Wheat County Estimates*. Retrieved from www.nass.usda.gov

United States Department of Agriculture. (2018b). *Minnesota Ag News – Crop Production Minnesota*.

United States Department of Agriculture. (2018c). *World Agricultural Production*. Retrieved from https://apps.fas.usda.gov/PSDOnline/app/index.html#/app/home,

Valin, H., Sands, R. D., van der Mensbrugghe, D., Nelson, G. C., Ahammad, H., Blanc, E., … Willenbockel, D. (2014). The future of food demand: Understanding differences in global economic models. *Agricultural Economics (United Kingdom)*, *45*(1), 51–67. https://doi.org/10.1111/agec.12089

Van Grinsven, H. J. M., Holland, M., Jacobsen, B. H., Klimont, Z., Sutton, M. A., & Willems, W. J. (2013). Costs and Benefits of Nitrogen for Europe and Implications for Mitigation. *Environ. Sci. Technol*, *47*, 45. https://doi.org/10.1021/es303804g

Veraverbeke, S., Dennison, P., Gitas, I., Hulley, G., Kalashnikova, O., Katagis, T., … Stavros, N. (2018). Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. *Remote Sensing of Environment*, *216*(May), 105–121. https://doi.org/10.1016/J.RSE.2018.06.020

Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., & Kersting, K. (2016). Plant Phenotyping using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants. *Scientific Reports*, *6*(1), 22482. https://doi.org/10.1038/srep22482

Wei, D., Zhang, W., Wang, C., Meng, Q., Li, G., Chen, T. H. H., & Yang, X. (2017). Genetic engineering of the biosynthesis of glycinebetaine leads to alleviate salt-induced potassium efflux and enhances salt tolerance in tomato plants. *Plant Science*, *257*, 74–83. https://doi.org/10.1016/J.PLANTSCI.2017.01.012

Wendel, A., & Underwood, J. (2017). Illumination compensation in ground based hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, *129*, 162–178. https://doi.org/10.1016/j.isprsjprs.2017.04.010

Whetton, R. L., Hassall, K. L., Waine, T. W., & Mouazen, A. M. (2018). Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. *Biosystems Engineering*, *166*, 101–115. https://doi.org/10.1016/j.biosystemseng.2017.11.008

Winter, M. E. (1999). N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Proc. SPIE 3753, Imaging Spectrometry V*. Denver, CO, United States. https://doi.org/10.1117/12.366289

Winter, M. E. (2004). A proof of the N-FINDR algorithm for the automated detection of endmembers in a hyperspectral image. In *Proc. SPIE 5425, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X* (p. 31). Orlando, Florida, United States. https://doi.org/10.1117/12.542854

Yu, N., Li, L., Schmitz, N., Tian, L. F., Greenberg, J. A., & Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sensing of Environment*, *187*, 91–101. https://doi.org/10.1016/j.rse.2016.10.005

Zhang, Y., Liu, Z., Khan, A. A., Lin, Q., Han, Y., Mu, P., … Xin, M. (2016). Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (Triticum aestivum L.). *Scientific Reports*, *6*(1), 21476. https://doi.org/10.1038/srep21476

Zortea, M., & Plaza, A. (2009). A quantitative and comparative analysis of different implementations of N-FINDR: A fast endmember extraction algorithm. *IEEE Geoscience and Remote Sensing Letters*, *6*(4), 787–791. https://doi.org/10.1109/LGRS.2009.2025520

Glossary

Artificial intelligence:	"a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." (Kaplan & Haenlein, 2019).
Cultivar:	Cultivar is a commercialized variety.
Deep learning:	"Deep learning is a class of machine learning algorithms that: • use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. • learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners. • learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts." (Wikipedia on "Deep Learning" around February 2019).
Genetic gain:	Genetic gain: "amount of increase in performance achieved annually per unit time through artificial selection." (José Luis Araus et al., 2018).
Hyperspectral imaging:	Hyperspectral imaging (HSI) integrates imaging and high-resolution spectroscopy whereby a continuous spectrum is scanned for each pixel, typically across the visible and near infrared range of electromagnetic spectrum.
Line:	A line is a wild species, landrace, historical cultivar or a possible new genotype that is still being tested for its potential value in a given environmental setting.
Machine learning:	"Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task." (Wikipedia on "Machine learning" around February 2019).
Phenotyping:	Phenotyping is a quantitative assessment of plant structure and function related to a desirable trait.