Spring 4-2023

Characterization of Physical and Biochemical Traits in Wheat and Corn Plants Using High Throughput Image Analysis

Kantilata Thapa
University of Nebraska-Lincoln, kthapa2@huskers.unl.edu

CHARACTERIZATION OF PHYSICAL AND BIOCHEMICAL TRAITS IN WHEAT AND CORN PLANTS USING HIGH THROUGHPUT IMAGE ANALYSIS

by

Kantilata Thapa

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Mechanized Systems Management

Under the Supervision of Professor Yufeng Ge

Lincoln, Nebraska

May 2023

CHARACTERIZATION OF PHYSICAL AND BIOCHEMICAL TRAITS IN WHEAT AND CORN PLANTS USING HIGH THROUGHPUT IMAGE ANALYSIS

Kantilata Thapa, M.S.

University of Nebraska, 2023

Advisor: Yufeng Ge

Plant phenotyping has been recognized as a rapidly growing field of research due to the labor-intensive, destructive, and time-consuming nature of traditional phenotyping methods. These phenotyping bottlenecks can be addressed by advancements in image-based phenotyping like RGB and hyperspectral imaging for the assessment of plant traits important for breeding purposes. This study aims (1) to characterize the physical and biochemical traits of wheat and corn plants using RGB and hyperspectral imaging in the greenhouse, and (2) to estimate leaf nitrogen (N), phosphorus (P), and potassium (K) content using hyperspectral imaging and an analytical spectral device (ASD spectrometer) and compare the performance from both datasets. Sixty wheat plants with 24 genotypes and 72 corn plants (a single genotype) with four different treatment combinations were manually measured and imaging was performed at different growth stages. RGB and hyperspectral images were processed to extract plant projected area (pixel count) and spectral reflectance, respectively. Partial Least Squares Regression (PLSR), Random Forest (RF), and Support Vector Regression (SVR) models were built to estimate N, P, and K contents from image-generated hyperspectral data, and from the ASD spectrometer. The results showed higher correlation for leaf area with plant pixel count with R^2 of 0.75 for wheat and R^2 of 0.68 for corn plants. For wheat plants, N was predicted more accurately with hyperspectral image datasets with R^2 of 0.69 but P and K

prediction was higher with ASD data using the PLSR model. For hyperspectral image datasets of corn plants, N prediction was higher using PLSR modeling with R^2 0.66 whereas P and K prediction was higher using the RF model with R^2 of 0.74 and 0.87 respectively. For corn plants using data from ASD, N, P, and K were predicted high by using the RF model with R^2 of 0.67,0.41, and 0.69 respectively. RGB and hyperspectral imaging would reduce the need for manual measurement and chemical analysis of leaf tissue, and the technique can be validated in other crops with different architectures for high-throughput macronutrient estimation. The findings from this study can help integrate various disciplines of science, including plant breeding, agronomy, computer vision, mathematics, and engineering, for crop improvement.

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude and deep appreciation to Dr. Yufeng Ge, my advisor, for his unwavering support, motivation, and encouragement throughout my Master's program. His guidance and encouragement pushed me beyond what I thought was unachievable at the beginning, and I am excited to continue working with him on more meaningful problems in coming days. I would also like to extend my sincere appreciation and heartfelt thanks to my committee members, Dr. Katherine Frels and Dr. Yeyin Shi, for their valuable suggestions, cooperative attitude, and constructive feedback during my research project. Additionally, I would like to express my gratitude to my lab members, Sabiha Ferdous, Sadia Mitu, Junxiao Zhang, Nipuna Chamara, Bidhan Ghimire, Husein Harun, and Zhaocheng Xiang, for their help and support during the data collection of my experiment. Special thanks goes to Vincent Stoerger, Troy Pabst, and Nathan Duffy for their help in setting up the experiments in the greenhouse and Dr. Geng (Frank) Bai for helping me to solve technical problems during data analysis. I owe a special debt of gratitude to my beloved parents, Mr. Gorakh Thapa and Mrs. Shanti Maya Thapa, for their continuous love, support, and trust in my decisions. Without their support, I would not have reached where I am today. I would also like to extend equal acknowledgments and appreciation to my sisters, uncle, aunt, grandparents, and all family members for their unbounded love and support during my academic journey.

Lastly, I want to express my gratitude to my friends, Nikee Shrestha, Neetu Regmi, Bhupinder Kaur, Shiva Paudel, Sagar Dahal, Bidhan Ghimire, Shubham Berry, and Gautam Takoo whose encouragement and support were invaluable when I doubted myself. I want to thank everyone who directly or indirectly lent me a hand during my

study. I am committed to paying back the debt by being the best in everything that you all aspire to see in me.

GRANT INFORMATION

The grant for my research was supported by (1) USDA-NIFA grant (High Intensity Phenotyping Sites: Transitioning to a Nationwide Plant Phenotyping Network, Award No. 2020-68013-32371), and (2) the Nebraska Agricultural Experiment Station through the Hatch Multistate funding program (Accession Number 7000908) from USDA-NIFA.

Table of Contents

List of figures

List of Tables

Chapter 1 Introduction

Precision agriculture (PA), also known as digital agriculture, is the use of large data sources in combination with advanced crop and environmental analytical tools to assist farmers in adopting the right management strategies at the right rates, times, and places, to achieve both economic and environmental goals. In recent years, there has been increased global interest in PA as a potential step toward meeting an unprecedented demand to produce more high-quality food and energy sustainably. Agriculture provides humanity with food, feed, fibers, fuels, and raw materials for sheltering. However, agriculture's role must be balanced with environmental sustainability, climate change adaptation, and a growing human population. In this challenging agricultural context, crop growth and status must be monitored in various locations and environments with varying spatial and temporal resolutions. Near-real-time monitoring is required not only to respond to extreme events due to climate change, thereby minimizing their impact on the global food system (Wheeler & Von Braun, 2013), but also to optimize management practices sustainably by optimizing externalities (Areal et al., 2018). One of the PA technologies that allow growers to measure, evaluate, and visualize crop and soil health conditions at various stages of crop production conveniently and cost-effectively is remote sensing (RS).

1.1 Components of remote sensing technology and its working mechanism

Remote sensing is an important tool as it provides a non-destructive means of providing information from the local to the global scale in a systematic manner, allowing for the characterization of spatiotemporal variability within a given area. Extracting agricultural information from remote sensing involves instruments or a sensor mounted on a platform

such as a UAV (Unmanned Aerial Vehicle), aircraft, or satellite, and a probe that helps the acquisition of information about an object from distance. These sensors measure electromagnetic radiation of certain wavelengths that are either reflected or emitted by the target. Table 1.1 shows different components of remote sensing technology including sensors, platforms, and the software that is commonly used to process the acquired data. Similarly, among various applications of remote sensing technology, some of the agriculture applications are: plant phenotyping, land use monitoring, yield forecasting, precision farming and ecosystems services.

Table 1.1. Different components of remote sensing technology (Shamshiri et al., 2017)

Sensors, Camera	Platforms	Software applied
RGB Hyperspectral Multispectral NIR LiDAR Thermal Sonar	Airborne • Satellite • Piloted UAV Flexible UAV • Fixed wing • Multi rotor Ground-based • Vehicle mount • handheld	ArcGIS PIX4D Agisoft Quantum GIS

The basic working principle of remote sensing technology with UAVs, satellites, and other platforms is almost the same. Energy, in the form of light, will travel from the Sun to the Earth. Light waves travel virtually like ocean waves – the distance between the peak of one wave to the peak of the next is known as wavelength. The energy emitted

from the Sun is known as electromagnetic energy and is part of the electromagnetic spectrum. The wavelengths that are commonly employed for agricultural applications cover a small fraction of the electromagnetic spectrum. When electromagnetic energy hits the plants during remote sensing in agriculture, one of three things can occur. The energy can be reflected, absorbed, or transmitted, depending on the wavelength of the energy and the characteristics of the plant itself. The reflected, absorbed, and transmitted energy can be detected by remote sensing technology. The relationship between the three components determines the spectral signature of the plants. This signature is unique to the biochemical and physiological properties of the plants. Each section of the electromagnetic spectrum has characteristics of energy level, wavelength, and frequencies associated with its photons Figure 1.1). Gamma rays have the highest energy and shortest wavelength whereas radio waves have the lowest energy and longest wavelength. The visible region ranges from 400-700 nm, the near-infrared region ranges from 700-1000 nm, and the shortwave-infrared ranges from 1000-2500 nm. For plant phenotyping research, visible, near-infrared, shortwave-infrared, and thermal infrared are the regions of the electromagnetic spectrum that are commonly used. The interaction of electromagnetic radiation with plants varies according to the wavelength of the radiation. The spectral reflectance of typical healthy vegetation is low at the visible region due to strong absorption by photoactive pigments (chlorophyll, anthocyanins, and carotenoids). In the near-infrared region, the reflectance is higher due to multiple scattering in the internal leaf tissue. Similarly, in the short-wave infrared region, the reflectance is lower

due to absorption by water, protein, and other leaf biochemical compositions. The typical spectral reflectance of a corn leaf in three different wavebands is shown in Figure 1.2.

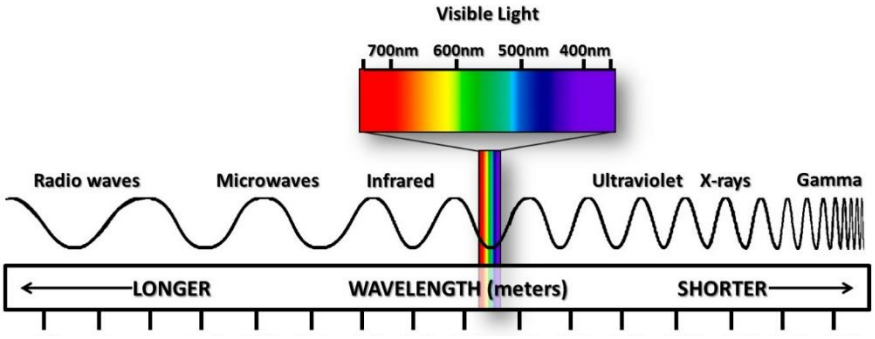


Figure 1.1 Electromagnetic spectrum with different bands. (Energy: The Driver of Climate, 2019) http://www.ces.fau.edu/nasa/module-2/radiation-sun.php

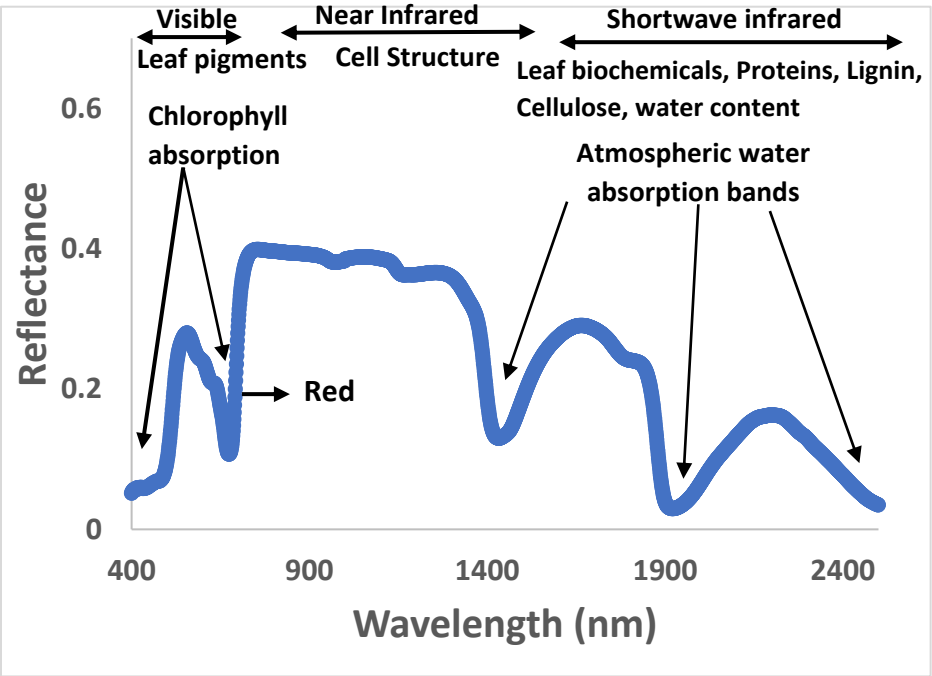


Figure 1.2 Spectral reflectance of a corn leaf from 400 to 2500 nm in the visible, near infrared, and shortwave infrared regions.

1.2 Frequently used imaging sensors

There are a variety of sensors available for obtaining images of crops at different wavelengths and resolutions, and with active or passive modes of operation. Active sensors are transducers that generates electric current or voltage directly in response to environment stimulation whereas passive sensors are transducers that produces a change in passive electrical quantity such as capacitance, resistance, or inductance as a result of the stimulation. Many phenotyping studies consider both passive RGB (Red, Green, Blue) and/or multispectral images, which are sometimes combined with an active LIDAR or sonar sensor. A study by (Bai et al., 2016), used a multi-sensor system for high throughput field phenotyping in soybean and wheat crops. Five different sensors; ultrasonic distance sensors, thermal infrared radiometers, NDVI sensors, portable spectrometers, and RGB web cameras were used to measure crop canopy at the plot scale. The result demonstrated a strong correlation between sensor-based plant traits and a significant correlation between sensor-based traits and grain yield at both early and late seasons with a person's correlation coefficient (r) from 0.41-0.55 and 0.55-0.70 respectively. Similarly, (Thapa et al., 2018), studied LiDAR-based instruments for high-throughput,3D measurement of morphological traits in maize and sorghum. A novel LiDAR-based instrument was developed to create 3D point clouds of a single plant. The morphological traits of plants including total leaf area, individual leaf inclination angle, and leaf angular distribution were derived and tested on maize and sorghum plants. The result from the study demonstrated a high correlation with R^2 of 0.91 between individual leaf area and leaf area measured from the 3D model and R^2 of 0.95 between total leaf area and trait measured from the 3D model. Other sensing techniques, such as using

fluorescence to estimate chlorophyll and nitrogen content, have potential but are currently limited to handheld sensors due to technological issues (Tremblay et al., 2012). Furthermore, the thermal infrared spectral domain is underutilized because the signal varies in time depending on the plant microclimate (Munns et al., 2010).

1.3 Different modeling approaches for crop traits prediction

There are different ways to connect physical measurements of crops obtained through remote sensing with agronomic traits such as crop yield, either using empirical or mechanistic approaches, or a combination of both. Empirical methods use statistical techniques to directly link inputs and outputs, while mechanistic approaches focus on the underlying mechanisms and relationships between inputs and outputs (Baker et al., 2018). Each method has its advantages and limitations, and the best approach usually depends on the specific context and application. In practice, the main distinction is that mechanistic approaches rely on assumptions and models, whereas empirical approaches necessitate data collection. Crop yield can be estimated using simple vegetation indices empirically based on satellite reflectance or using a combination of remotely sensed green area index (GAI) and process-based crop growth modeling in a mechanistic way. The empirical approach is simpler, but it may not be as accurate and may not be able to be extrapolated to other times and locations. The mechanistic approach is more detailed and provides more explanation, but it relies on assumptions that may not always be accurate, leading to higher uncertainty.

In the past, several studies have been done to predict chlorophyll content using imaging sensors. The study by (H. Zhang et al., 2022) studied three imaging modules; RGB, hyperspectral, and fluorescence imaging separately and in a combination of three to

predict chlorophyll content in Sorghum plants. The model predicted R^2 of 0.67-0.88 from RGB images, R^2 of 0.77-0.78 using vegetation index from hyperspectral images, and R^2 of 0.79 from fluorescence images whereas features combined from three modules using PLSR modeling have R^2 of 0.90. The study suggested a fusion of image features from different imaging modules using PLSR modeling significantly predicts chlorophyll contents in plants. Similarly, (Miao et al., 2021) studied automation of leaf counting in maize and sorghum plants using deep learning techniques. About 150,000 maize and sorghum plant images were generated and a subset of 17,783 was annotated of the position of the individual leaf tip. Two deep learning-based approaches were used. The first approach counting by regression based on CNN showed lower accuracy and increased biased for plants with large leaf numbers whereas detection based on faster R-CNN demonstrated near-human accuracy. Similarly, (Jöreskog et al., 2016) studied leaf counting in monocot plants using deep regression models to overcome the need for substantial training datasets and annotations. A deep neural network that does not require label leaf structures even with serious leaf overlapping in images was explored. The leaf skeleton was extracted, and the regression model was fed with original images, derived skeleton, and augmentation. The result demonstrated improved accuracy in leaf-counting with many overlaps and occlusions and reduces training costs compared to state-of-the-art methods. LIDAR acquisitions, stereo RGB images, and sonar are primarily used to measure canopy height (Andrade-Sanchez et al., 2014), but much more information is expected from the generated 3D point clouds to describe the plant stand architecture, including the area, density, and orientation of the various plant parts (e.g., leaf, stem, ear, or flower). Because of the limitation of underground sensing, root zone measurements

were also investigated in field phenotyping experiments. However, a few attempts have been made to characterize the root architecture using camera or scanner technologies in clear rhizotron tubes or using electrical resistivity tomography by linking the measured signal to root water absorbance rates (Postic et al., 2019).

High Throughput Plant Phenotyping (HTPP) experiments that use images to study plant traits do not take full advantage of the information they obtain. Even though multispectral cameras are used, many experiments only establish statistical relationships or use machine learning algorithms to correlate vegetation indices with traits like chlorophyll, leaf area index, and nitrogen content. (Araus & Cairns, 2014) highlighted this limitation in HTPP experiments. However, relatively few studies have applied radiative transfer models to simulate the radiative behavior of leaves or plant stands or have used model inversion techniques to infer physiological traits from multispectral images. Furthermore, in-field calibration of thermal infrared (TIR) sensors is more complex and delicate. Therefore, RGB images have been primarily used for classification or segmentation tasks (for example, plant and organ counting), while multispectral images are used to determine architectural traits (mostly green area index) or physiological traits (e.g., chlorophyll, water, nitrogen) through radiative transfer model inversion. While statistical and model inversion approaches can perform similarly in the context of phenotyping, the results may be limited if the dataset used to calibrate the statistical relationships is not large enough to represent a wide range of possibilities, including new cultivars (Jay et al., 2017). In addition, using a generic 1D radiative transfer model may not be appropriate for describing crops at the considered spatial resolutions, particularly for row crops. To address these limitations and develop operational pipelines for specific species, there is a

need for further advancements in 3D modeling and inversion techniques, which can be supported by field phenotyping experiments using 3D point clouds generated through Structure from Motion (SfM) or LIDAR techniques (Liu et al., 2017). Typically, images from different sensors have been used separately to study specific traits. However, some recent studies have combined traits from different spectral domains to analyze more complex traits, often by establishing statistical relationships between them. For example, (Tilly et al., 2015) used canopy height from LIDAR or SfM applied to RGB images, along with vegetation indices, to estimate above-ground biomass. (Jay et al., 2017) used machine learning on high-resolution RGB images to estimate the green fraction of vegetation and applied a threshold value to a lower-resolution multispectral image to distinguish green pixels from the background and correct the bias in the estimation. However, there is potential to use sensor synergy in even more ways, such as combining plant height from LIDAR or stereo images and cover fraction from RGB images with thermal infrared (TIR) images to better understand the water balance of a plant stand.

In a similar way, high throughput plant phenotyping (HTPP) allows for the collection of large training datasets that can be used to develop deep-learning models (Pound et al., 2017). This has led to significant progress in image classification, segmentation, and object identification in plant stands through high-throughput phenotyping experiments (Singh et al., 2016).However, these techniques are supervised and require manual annotation to identify different classes, which is a time-consuming and labor-intensive process that requires expert knowledge (Kamali & Nazari, 2018). They are also sensitive to the composition of the learning dataset, the type of crop, the spatial and spectral resolution of the sensor, and the acquisition conditions (Mohanty et al., 2016). As a

result, they are only slightly transferable from one species to another or from one phenotyping platform to another.

1.4 Plant breeding and phenotyping

Plant breeding is the application of genetic principles to produce the most desirable traits in plants. After the discovery of Gregor Mendel's theory in 1900 AD, modern plant breeding based on genetics came into being. After that many discoveries were made based on his theory which led to the establishment of the field of genetics. Genetics is the study of the DNA of organisms, how DNA manifests as genes, and how those genes are inherited by offspring. Genes are passed to offspring in both sexual and asexual reproduction, and over time natural selection can accumulate variations amongst individuals on the group level, in the process known as evolution. The interaction of the genetic makeup of an organism and environmental factors which is termed as G*E interaction leads to the expression of traits which is known as phenotype. Plant breeders can utilize genomic data more effectively if it is easier to quantitatively measure plant anatomical, ontogenetical, physiological, and biochemical properties (Walter et al., 2015).

Plant phenotyping is the assessment of complex plant traits such as growth, development, resistance to stresses, architecture, physiology, ecology, yield, and the basic measurement of individual quantitative parameters that form the basis for complex traits (L. Li et al., 2014). In addition, crop phenomics can be defined as a multidisciplinary study of high-throughput accurate analysis of multidimensional phenotypes on an organism-wide scale through crop development (Yang et al., 2020).It is an accurate and precise estimation of plant traits like growth rate, biomass, architecture, chemical contents, etc., ranging from a

cellular level to the whole plant canopy level (Rahaman et al., 2015).But the lack of cost effective and less labor intensive phenotyping is a limiting factor for genome-assisted crop improvement. Therefore, it is imperative to expand the use of high throughput plant phenotyping to optimize crop improvements.

1.4.1 Importance of image-based phenotyping for plant breeding and commonly used phenotyping platforms

People around the world on all continents used to live as hunter-gathers until the Pleistocene era (Diamond & Bellwood, 2003). With the advancement of time, the domestication of relatively few wild plants and animals started between 8500-2500 BC (Sforza et al., 1995). Hunters and gatherers included agriculture as their main path of expansion due to the major advantages gained from food production. The first advantage was higher food yield acre-which can support higher population densities. The second advantage was the sedentary nature of most food-producing societies, which could store surplus food and was essential for social stratification and centralized states. The third one was the development of new technology and resistance to epidemic infectious diseases (Diamond & Bellwood, 2003). The Selection of superior traits has been prevailing since the first plant was domesticated in 8500 BC (Damania, 1998). According to the United Nations Department of Economic and Social Affairs' World Population Prospects 2020, the global population is projected to reach 9.7 billion by 2050, up from an estimated 7.9 billion in 2021 and to meet the predicted food demand for a growing population, crop production needs to be doubled. Land clearing and more intensive use of existing cropland could contribute to meeting the demand, but the environmental cost would be higher (Godfray et al., 2010). Now the question arises: how to increase food

production sustainably. There are two divided opinions: the first recommends modern techniques like mechanization, irrigation, fertilizer application, and improved genetics, and the other recommends techniques that improve productivity without the application of synthetic fertilizer and pesticides. Both approaches offer a needed solution, and it would be more impactful to blend the best ideas of both. However, improving the genetic makeup of crops by breeding is one of the way to maximize progress in both organic and conventional systems (Mir et al., 2019).

Traditional methods of phenotyping are often time-consuming, labor intensive, and involve destructive methods. They represent a major bottleneck for studies dealing with large numbers of genotypes and sample sizes (e.g., forward genetics and breeding experiments). Thus, this phenotyping bottleneck can be alleviated by image-based phenotyping which has the potential to assess plant performance in a quantitative, time-series, and cost-effective manner (Tausen et al., 2020). Although it is desirable to assess the growth rate of plants, it is challenging for breeders because it requires multiple measurements from planting until harvest and such manual and daily measurements may be impractical in large-scale studies (Walter et al., 2015). The existing methods for plant phenotyping are not sufficient and require further development and improvements to accurately capture physical, physiological, and biochemical traits related to plant mechanisms which is important to better understand the genetic mechanisms and biology underlying these traits. Therefore, image-based phenotyping could play a great role in relieving such phenotyping bottlenecks.

Many phenotyping platforms in research or commercial settings have been developed (Granier & Vile, 2014). Image-based, high-throughput phenotyping platforms are the

systems that allow monitoring of plant growth, development, and yield-related traits in a non-destructive, fast, and high-throughput manner. According to (D. Li et al., 2021), high throughput phenotyping platforms are those that can collect massive amounts of phenotypic data from hundreds of plants every day with a high degree of automation. A wide range of phenotyping platforms either customized or self-developed have emerged, which can be divided into 1) greenhouses and growth chambers under a controlled environment, 2) ground-based proximal phenotyping in the field, and 3) drone, aerial, or satellite remote sensing. In research by (Bai et al., 2019), a large-scale, cable-driven integrated sensing and robotic system was developed at the University of Nebraska-Lincoln field phenotyping facility. Four different sensors including a multispectral camera, thermal infrared camera and 3D scanning LiDAR, and visible near-infrared spectrometer were integrated into a system for plant measurement. The system is fully automated, robust, and anticipated to collect large-scale phenotypic data. Another study by (Bai et al., 2018), further explored soybean iron deficiency chlorosis (IDC) using RGB imaging with field-based scoring. The results from the study anticipated RGB images derived from field phenotyping platforms with real time IDC scoring models allows for more robust, fast and cost-effective scoring for soybean IDC screening and breeding purposes. Similarly (Virlet et al., 2017), developed an automated robotic field phenotyping platform with a dedicated sensor array. The sensors mounted were visible, fluorescence, thermal, and infrared, two hyperspectral imagers, and dual 3D laser scanners. These sensors in combination produce a detailed description of the canopy across the entire life cycle of plants. Similarly, (J. Li et al., 2019) studied the grain yield variation in winter wheat extracted from UAV imagery and investigated principal

variable selection to explain grain yield variation . 172 variables were extracted from each plot from vegetation index and plant height maps. A parametric (LASSO) and non-parametric algorithm random forest (RF) were applied for variable selection. The result concluded high-resolution imagery derived from UAS can derive more features within a plot level which is very important for breeding purposes.

Recently, high-throughput phenotyping platforms have been installed and operated in greenhouses or growth chambers (Australian Plant Phenotyping Facility). Greenhouse-based screening of plants offers various advantages over costly, time-consuming field trials, which typically do not evaluate entire plants. Additionally, the environment can be controlled which minimizes the effects of confounding biotic and abiotic factors and has good repeatability, auto-operation, and good resolution imagery. Furthermore, microclimatic conditions are also modified to suit the crop needs inside the greenhouse. Many of the complex phenotyping systems use conveyor belts to move plants to high-quality imaging chambers (Perez-Sanz et al., 2017). In this study, controlled environment phenotyping is particularly described in detail as shown in Table 1.2

Table 1.2.Overview of high throughput phenotyping platforms used in the greenhouse (D. Li et al., 2021)

Indoor HTPP	Model	Sensors	Throughput (pots)	Plants	Traits	Location
Conveyor type	Lemna Tec Scanalyzer 3D	RGB, NIR, FLUO	312	Barley	Biomass, plant height, width, compactness, drought stress	Germany

	Lemna Tec Scanalyzer 3D	RGB, NIR, FLUO, Hyperspectral	672	Sorghum, maize, barley	Biomass, leaf water content	USA
	Lemna Tec Scanalyzer 3D	RGB, NIR, FLUO, Hyperspectral	2,400	Chickpea, wheat	Nutrient stress, salt stress, water content, nitrogen content	Australia
	Bellwether	RGB, NIR, FLUO	1,140	Setaria	Plant height, biomass, water-use efficiency, water content	USA
Bench type	Phenovator	Monochrome	1,440	Arabidopsis thaliana	PLA, PSII efficiency	Netherlands
	Phenoscope		735	Arabidopsis thaliana	Rosette size, expansion rate, evaporation	France
	Phenoarch	RGB	----	Maize	The growth rate of ear and silk	France
	Glyph	RGB	120	Soybean	Water use efficiency, drought stress	Argentina
	Lemna Tec Scanalyzer	RGB, FLUO, NIR	------	Arabidopsis thaliana	Water stress	USA

1.5 Digital imaging module

A digital image is an image stored in digital form and divided into a matrix of pixels.

Each pixel consists of a digital value of one or more bits and may represent, but is not

limited to, energy, brightness, color, intensity, sound, or a classified value derived through image processing. A digital image is stored as a raster and may contain one or more bands. Image acquisition is the process of obtaining a digital representation of a scene and the devices that are used to capture a scene are known as imaging sensors. Different types of imaging modules are practiced based on the objective and acquisition of specific traits of plants. Imaging at different wavelengths is used for different aspects of plant phenotyping. For example, RGB (visible) imaging is used to measure the physical characteristics of the plant's projected area, biomass, color, root architecture, yield, disease severity, seed morphology, seedling vigor, fruit number and distribution, etc. Thermal infrared imaging could be useful to characterize the plant temperature to detect plant response to the amount of water and transpiration rate for water stress assessment. Similarly, hyperspectral images have been used for the assessment of biochemical traits in plants like water contents, macronutrients including Nitrogen (N), Phosphorus (P) and Potassium (K) as well as several micronutrients and other chemical contents. Digital image analysis in a controlled environment offers high resolution, high throughput, and the precise evaluation of crop traits of interest comparable with human data collection. With current technological advancements (camera resolution, processing speed, storage, availability of image processing tools, etc.), the application of image analysis is anticipated to grow for practical applications in evaluating crop traits in an accurate and high-throughput manner. RGB and hyperspectral imaging modulus are discussed here in detail because of the focus of this study on these two imaging modules in both wheat and corn plants.

1.5.1 Different imaging techniques used for high throughput estimation of wheat and corn traits

Wheat is the third most-produced cereal crop in the world, after maize (corn) and rice. In 2021, world wheat production was 779.9 million metric tons, and it is expected to reach 783.92 million metric tons in 2022/2023, which is 4.32 million tons more than the previous year (Wheat Production by Country | World Agricultural Production 2022/2023). In the United States (USA), the total wheat production in 2020 was 49.7 million metric tons, the average wheat yield was 3.34 metric tons per hectare (ha) and the total area planted was 17.99 million ha. Similarly, in 2021 the total production was 44.80 million metric tons, the yield was 2.98 metric tons per ha and the total area planted was 18.9 million ha (Crop Production 2021 Summary, 2022). Comparing the production from these two years, the production seems less in 2021 than in 2020 even though the total area planted was higher in 2021.

To meet the demand of the growing population and to maintain constant yield, it is important to develop wheat varieties that are adaptable to various environmental conditions. With the constant breeding in wheat, breeders have developed numerous semi-dwarf cultivars that give higher yield, reduced lodging, and have an efficient response to applied nitrogen. Due to increased population growth and higher consumption of wheat, emphasis has been given to improving yield by applying various inorganic fertilizers. According to (the US Department of Agriculture National Agricultural Statistics Service, 2022), the estimated amount of nitrogen fertilizer applied to wheat in the US for the 2021 crop year was 2.99 million metric tons (USDA Crop Production 2021 Summary). The physical traits such as height, biomass, leaf area, and

the biochemical traits such as macronutrients (N, P, K), micronutrients, chlorophyll content, and water content affect yield components that will eventually contribute to yield differences in wheat. Phenotyping of these traits is crucial and considered a major bottleneck for wheat improvement due to the labor-intensiveness, high-cost, time consuming, and different architecture of plant (Zhao et al., 2019). Literatures have explored applications of phenotyping in the assessment of plant architecture, seed characteristics, canopy growth, and root morphology in wheat (C. Zhang et al., 2018). In research from (Correia et al., 2022) thermal and multispectral images were explored to determine evapotranspiration and biomass in wheat using image-extracted parameters. High-throughput phenotyping methods proved to be an efficient approach to quantitatively characterize the G*E interaction of complex traits like stress tolerance in wheat. Similarly, a study reported by (Rahaman et al., 2015), predicted yield in wheat using secondary traits; spectral reflectance and canopy temperature. The grain yield heritability ranged from 0.30 to -0.72 while the correlation between grain yield and spectral reflectance and canopy temperature ranged from -0.5 to 0.5. This study suggests secondary traits derived from phenotyping can accurately predict wheat grain yield allowing breeding programs for making a robust and rapid selection. The study by (C. Zhang et al., 2018) investigated seed/seedlings using digital image analysis. Images were collected using commercial digital cameras and analysis was done using custom-developed algorithms. The data from image-based phenotyping and manual measurement of coleoptile were found significantly correlated ($p<0.0001$) with a correlation coefficient of (r) 0.69-0.91. (Morgounov et al., 2014) has suggested the change in NDVI among the genotypes from tillering to the flowering phase has a significant difference in yield.

Maize (Zea, mays L.) is a major staple food crop across the globe that plays a key role in agri-food systems. The maize production area globally is approximately 197 million ha with significant areas in sub-Sahara Africa, Asia, and Latin America (FAOSTAT, 2021). It is a versatile and multi-purpose crop that is used as feed and an important food crop in different parts of the world. The USA is the biggest producer of maize globally and each year, US farmers plant about 90 million acres of maize, the majority of which is grown in the mid-west region with Iowa and Illinois as the top one-third producer of US maize (Economic Research Service, 2022). In the United States, the total area planted, and total area harvested in 2021 was 37.78 million ha and 34.56 million ha, respectively. Similarly, the yield was 11.11 metric tons per ha and production was 38.39 million metric tons in 2021 (Crop Production 2021 Summary, 2022).

Sudden extreme climatic events such as irregular rainfall, temperature fluctuations, floods, drought, hailstorms, and thunderstorms affect the growth and development of maize plant. The pressure of increasing maize production in the future is high due to the increase in the global population and its important role in food security. Water and nitrogen are considered as two important factors that account for the majority of growth and production in maize as they have a positive impact on biomass and grain yield. However, the excessive amount of nitrogen fertilizer applied to maize has impacted groundwater, air pollution, and nitrogen use efficiency (NUE). Therefore, further study and research should be done on the traits related to NUE for a sustainable increase in production, environmental protection, and resource constraints. The traits related to NUE are biomass, height, leaf area, chlorophyll content, etc. The assessment of these traits on large scale is difficult due to the destructive nature of phenotyping, high labor cost, more

time consumption, and low throughput data collection. Therefore, studies have investigated the HTPP of maize using imaging technologies to investigate diverse corn traits.

A study by (Ge et al., 2016) investigated two different maize genotypes under two water treatments. RGB and hyperspectral images of maize genotypes were obtained from an automated imaging system at the University of Nebraska-Lincoln. An excessive green pixel extraction algorithm was used for RGB image analysis and leaf reflectance was obtained from hyperspectral images which were further analyzed using the partial least squares regression (PLSR) model. A strong correlation was found between plant projected area and destructively measured parameters such as Fresh Weight (FW), Dry Weight (DW), and Leaf area (LA). Hyperspectral imaging was found effective in estimating water content in maize using NDVI index (670nm as red and 770nm as NIR) with the coefficient of determination of 0.81- 0.92. Similarly, (Pandey et al., 2017) further investigated the potential of hyperspectral imaging to quantify biochemical properties of maize like water content, macronutrients (N, P, K) as well as micronutrients (Na, Fe, Mn, B, Cu, Zn) using pairwise score plot to compare first three principle component (PC1, PC2, PC3) of the plant spectra. Most of the spectral differences were due to water treatment followed by nutrient treatment, which had important implications in PLSR modeling. The prediction accuracy was found higher for water content (R^2: 0.93) followed by macronutrients (R^2: 0.69-0.92). However, this study did not include different levels of nitrogen and their respective effect on crop growth, development, and morphology of corn plants. The research by (Liang et al., 2018) examined classical and hyperspectral time-series imaging of maize lines mainly used in field trials. This study

also supports the strong correlation between manual measurement and image-based measurements. However, there might be non-random measurement errors while measuring traits like biomass due to genotypic variation, which was not accounted for in this study. Another study by (Ge et al., 2019) investigated maize chlorophyll content, leaf water content (LWC) and specific leaf area (SLA), and nitrogen (N), phosphorous (P), and potassium (K) using visible, near-infrared, and short wave infrared (VIS-NIR-SWIR) spectroscopy in field condition. Partial least square regression (PLSR) and support vector regression (SVR) modeling was used to estimate leaf properties and several vegetation indices (GNDVI, RENDI, and NDVI) were also calculated from hyperspectral data. The chlorophyll content was estimated most accurately with R^2 of > 0.94, and N was predicted satisfactory with R^2 of 0.85 but some VIs were not able to estimate other four-leaf properties (LWC, SLA,N,P). The study suggested VIS-NIR-SWIR spectroscopy can be a promising tool to determine leaf physiological and chemical properties with low-cost and non-destructive approaches which can be a great benefit for the plant phenotyping community.

However, there are still some knowledge gaps that need to be fulfilled to establish image-based phenotyping as a robust method for trait assessment. The measurement protocols used for image-based plant phenotyping are currently not standardized, making it difficult to compare results across multiple studies. Similarly, more robust image segmentation algorithms that can accurately extract meaningful data from digital images are required particularly for colors like yellow, and brown produced by nutrient deficiency or other diseases. While image-based plant phenotyping has the potential to be a valuable tool, researchers and industry have yet to embrace it because of the high cost of the

phenotyping platforms and need for advanced and more expertise in high throughput data analysis. Because of the costly phenotyping platforms, it can limit their adoption by smaller research groups or organizations. However, with ongoing research and development, these gaps will gradually be filled by effectively monitoring crops from a newer perspective.

1.6 Role of image-based plant phenotyping and computer vision during Covid-19

The COVID-19 pandemic had a significant impact on cereal crop production and the global food system. The lockdowns and restrictions on movement put in place to contain the spread of the virus disrupted supply chains and affected the availability of inputs such as seeds, fertilizers, and labor. This led to delays in planting and harvesting, and in some cases, reduced crop yields. The FAO estimates that the pandemic resulted in a decrease of up to 10% in global cereal production in 2020. However, the impact on cereal production has been uneven, with some countries and regions experiencing larger declines than others. The FAO has also warned of the potential for food insecurity and malnutrition to increase in vulnerable populations because of the pandemic's effects on agriculture and food systems. It is important to note that the impact of the pandemic on cereal production and the global food system is ongoing and dynamic, and the full extent of its effects is still being assessed. During the COVID-19 pandemic, image-based phenotyping has played a critical role in supporting plant and crop research and development, as it allows researchers to continue their work while maintaining social distancing and other safety measures. It allows researchers to collect data on plant traits such as growth, development, and stress responses without the need for manual measurements or destructive sampling. For example, image-based phenotyping can be used to study plant

responses to different environmental conditions, such as changes in temperature, light, or water availability, which can inform the development of more resilient crop varieties. In addition, image-based phenotyping can be used to monitor the health and growth of crops remotely, which can be especially useful in the current context where fieldwork and travel may be restricted. This can help researchers to identify potential problems or stressors affecting crop performance and make informed decisions about management practices.

Overall, image-based phenotyping can be a valuable tool for plant and crop researchers to continue their work and advance our understanding of plant biology and crop production during the COVID-19 pandemic. As a result, more research and development are required to establish image-based phenotyping as a frontier for crop improvement and gene discovery, ultimately contributing to food security.

1.7 Objectives

There are two broad objectives of this study. The first objective is to investigate the relationship between plant pixel information derived from RGB image analysis and manually measured physical traits for wheat and corn plants. Specifically, this study aims to determine the degree of correlation between plant pixels or total plant projected area extracted from RGB image analysis and physical traits such as leaf area (LA) for corn, total biomass area for wheat, fresh weight (FW), and dry weight (DW). The study will utilize statistical analyses and various image segmentation algorithms to determine the strength of the correlation between the two sets of data and evaluate the feasibility of using RGB image analysis as a non-invasive and efficient method for measuring plant physical traits.

The second objective of this study is to evaluate the accuracy of predicting biochemical traits using reflectance data obtained from hyperspectral images and a handheld non-imaging spectrometer (ASD). Specifically, the study will focus on predicting nitrogen (N), phosphorous (P), and potassium (K) levels in wheat and corn plants using three different machine learning models; partial least squares regression (PLSR), random forest (RF), and support vector regression (SVR) and the performance of each model will be compared.

Overall, this study aims to contribute to the field of plant science by evaluating the feasibility and accuracy of using RGB and hyperspectral imaging to measure and predict plant physical and biochemical traits, respectively. The outcomes of this study will provide insights into the development of non-invasive and efficient methods for plant trait measurement, ultimately contributing to crop improvement, advancement of plant breeding, and agriculture research.

Chapter 2 Materials and Methods

2.1 Plant materials and growing condition

The wheat and corn phenotyping experiments were carried out at the Greenhouse Innovation Center, University of Nebraska-Lincoln, using the LemnaTec Scanalyzer3D system (LemnaTec GmbH, Aachen, Germany). Peatmoss mix (18 bags total) was used as the growth media for wheat, which was divided into two preparations of nine bags each. Two osmocote formulations, 2184.1 g/m of lime, and 103.131 liters/m of water were mixed for every nine bags. The two Osmocote formulations were based on fertilizer release, with one having a 2–3 month release of 15-9-12 (N-P-K) fertilizer and the other having a 5–6 month release of 15-9-12 (N-P-K) fertilizer. A total number of 200 pots were used, 192 with wheat plants and 8 empty pots with growth media only to estimate the evaporative loss. 1000 grams of pot mix was used per pot. At first, wheat genotypes were chosen and sown on December 16th, 2021, and vernalized for three months to stimulate flowering and seed production. On March 17th, 2022, seedlings were inspected, and vigorous seedlings were transplanted into pots. Pots were kept in a separate greenhouse for a few days after transplantation to allow seedling establishment before transporting them to the conveyor belt for imaging. To monitor the growth and development of the wheat plants, the pots were transported to a conveyor belt and automated imaging was performed. The greenhouse's temperature was kept between 20-25 °C during the day and 15-17 °C at night, with relative humidity (RH) of ~ 60%. The daytime photosynthetic active radiation (PAR) was supplemented with 200 mol m^{-2} s^{-1}

LED red/blue light lamps. The total photoperiod was kept at 17 hours. Throughout the study period, all pots were kept well-watered until the wheat plants reached full maturity.

The maize phenotyping experiment used a single inbred, corn line B-73. Premier tech horticulture pro-mix plus media was used, with 2184.1 grams of lime added per meter. The pots used had a diameter of 241.3 mm, a height of 259.08 mm, and a capacity of 9.082 liters. They were filled and weighed to the nearest gram. The greenhouse temperature, like the wheat crop, was kept between 17 and 25°C. RH of nearly 60%, PAR of 200 µmol m^{-2} $s^{-1}$and a total photoperiod of 17 hours were maintained, which was comparable to the wheat crop.

2.2 Experimental Design

Twelve pairs of winter wheat genotypes were chosen for the wheat phenotyping study. Nine pairs differed in NDVI and three pairs did not differ in NDVI, from a previous field screening experiment. The experimental design was a complete randomized design (CRD) with eight replications. Extra wheat genotypes were also grown outside the belt for destructive sampling.

One commercial inbred line, B-73, was used for the maize phenotyping study. A completely randomized design (CRD) was used, with four treatments: two nitrogen levels (High and Low), two water levels (well-watered (WW), and drought (D). High N + WW, High N+D, low N + WW, and low N + D were the four different treatment combinations that were used. Each treatment was applied to 25 plants. The Hoagland formulation was used to create the 100% ammonium solution and the 0% ammonium solution. In total, 120 corn plants were grown, 100 on the conveyor belt and 20 in a separate greenhouse chamber outside the belt. From mid-July, 2022 to the 16th of September 2022, imaging

was performed by transporting 100 pots to the conveyor belt. The corn plants were monitored every week and destructive/manual measurement was taken at various growth stages ranging from V6 to R5 stages. Following imaging, an equal number of plants from each treatment were chosen for destructive sampling from the conveyor belt.

2.3 Data collection

In total, sixty wheat plants were destructively measured at five growth stages starting from stem elongation, booting, anthesis, early grain filling, and physiological maturity. The plants were monitored based on Zadok's scale. At first, imaging was done by passing the wheat plants through automated imaging chambers before the destructive sampling. Genotypes were selected randomly, and above ground portion of the plants was cut, and five traits were recorded: Leaf area (LA), total fresh weight (FW), total dry weight (DW), chlorophyll content, and spectral reflectance as a ground truth measurement. Total plant leaf area was measured using a leaf area meter LI-3100C (LI-COR Inc., Lincoln, NE) and total shoot fresh weight was measured using a digital balance. The chlorophyll content was measured using an MC100 chlorophyll meter (Apogee Instruments, Inc., Logan, UT) on three leaves from each plant and three positions on each leaf starting from the base to the apex (nine measurements from each plant in total). Leaf level spectral properties of the plants were determined using a handheld spectrometer (Analytical Spectral Device, or ASD) with a spectral range of 350-2500 nm. A total of nine measurements were taken from each plant (three positions on each leaf) following the same protocol as chlorophyll content. After every nine measurements, the spectrometer was calibrated with a spectral panel (Labsphere, Inc., North Sutton, NH) with >99% reflectance for white referencing. Finally, all nine scans

were averaged and taken as an average spectral reflectance per plant for further statistical analysis. After all these measurements, samples were kept in a walk-in oven for drying at 60°C for 72 hours and dry weight was taken immediately. All the samples were sent to Ward Laboratories (Kearney, NE) for routine plant tissue analysis to determine nitrogen, phosphorus, potassium, and other micronutrient content in plants. Similarly, images were also captured for the same plants just before the destructive sampling.

The same set of traits was recorded for the corn experiment as wheat; leaf area, fresh weight, dry weight, chlorophyll content, and spectral reflectance. For the leaf area, leaves were separated from the stem and passed through the leaf area meter and a similar protocol was followed for measuring fresh weight, and dry weight. For chlorophyll content and spectral reflectance, the fifth leaf was selected counting from the top, and measurement was taken on the middle portion, three on one side of mid-rib and three on another side. In total, six measurements were taken and averaged for further analysis. Both black and white reference panels were used while scanning leaves from the ASD spectrometer to see the difference in reflectance. In total, 74 corn plants were destructively sampled weekly starting from V6 to R5 stages. The treatments were applied weekly with 250 ml of 100% ammonium solution as high nitrogen and 250 ml of 0% ammonium solution as low nitrogen in each pot. All the pots were automatically watered/weighed at the station and imaged daily. Similar to wheat, after drying the leaf samples in a walk-in oven for 72 hours, dry weight was measured, and all the leaf samples were sent to Ward Laboratories located at Kearney, Nebraska for N, P, and K and other micronutrient determination.

2.4 High throughput imaging

The high-throughput plant phenotyping facility at Greenhouse Innovation Center, University of Nebraska-Lincoln has four; RGB, hyperspectral/NIR, thermal, and fluorescence imaging chambers. The system also incorporates an automated weighing and watering station where the change in pot weight because of water evaporation and transpiration can be quantified, and prescribed amounts of water can be precisely applied. A specific number of plants were selected at each growth stage randomly and loaded manually on a conveyor belt to image on the same day before destructive sampling. Image acquisition was done every day from 10 am to 2 pm to minimize the effect of temporal variation. Pots move through the conveyor belt and the imaging process starts as they enter the chambers. This study focuses on images acquired from the RGB (Basler AG, Ahrensburg, Germany) and hyperspectral (Headwall Photonics, Fitchburg, MA, USA) cameras. The RGB camera is responsive to the wavelength range from 400-700 nm and the image has a resolution of 6576*4384 pixels. Images were taken from 10 different views (0°, 36°, 72°, 90°, 108°, 144°, 216°, 252°, 288°, and 324°) and one top view (90°) to get the 3D view of the plant. The hyperspectral camera has a push broom type of imaging spectrometer with halogen bulbs as a source of illumination. A total of 243 images of each plant were captured, each image with a spectral resolution of ~4.5-5nm, and the camera has a wavelength ranging from 545-1700nm (Pandey et al., 2017). Upon exiting the hyperspectral chamber, pots (including reference pots, with no plants) were automatically weighed and water was applied at 75% of field capacity to well-watered treated plants and 50% of field capacity to drought-treated plants.

Image analysis for both RGB and hyperspectral images was done by using MATLAB R2022a. The major objective of image segmentation is to extract the foreground from the background. RGB images were transformed to a single band using excessive green pixel index, $ExG = 2 * g - r - b$, where g, r, and b are the green, red, and blue components. After excessive green pixel extraction, the single band image was converted to a binary image by applying a specific threshold value. For wheat, a threshold of 0.05, and for corn a threshold of 1.1, was found to segment plant pixels effectively from the background. The resulting binary image may still contain noise. Therefore, morphological opening was performed to remove noise and any other non-plant material. The morphological opening function in MATLAB is a basic image processing operation which is composed of two steps: erosion of the image followed by dilation of the same image using the structuring element. This operation is useful to remove small details such as noise and thin lines while preserving the larger structure in an image. The excessive green index algorithm was applied to all the images from ten side views and the extracted plant pixel from each side was averaged as the total plant pixel of a given plant. The top view image was not included in the analysis due to errors in some images. Figure 2.1 shows the original image and the resultant binary image after RGB image segmentation in wheat plants.

Figure 2.1 Original and binary image after RGB image segmentation in wheat plants

For corn plants, RGB image segmentation was done by using two methods: Excessive green pixel extraction algorithms and HSV (Hue, Saturation, Value) color segmentation algorithms. At first, the correlation was performed by using green pixels only (Figure 2.2); second, correlation was performed by adding green pixels derived from excessive green pixel algorithm and yellow and brown pixels from HSV color segmentation algorithm (Figure 2.3). Third, correlation was performed using plant pixels extracted from HSV color segmentation algorithm only including green, yellow, and brown pixels (Figure 2.4).

The hue refers to the actual color of the image and its value ranges from 0 to 360 degrees. The Saturation indicates the intensity or purity of color, and the value ranges from 0 to 100 percent where 0 indicates a gray color and 100 indicates a fully saturated color. Similarly, the value indicates the brightness or lightness of the color represented by a percentage. A lower value indicates a darker color whereas a higher value indicates a brighter color. For HSV segmentation, at first, the RGB image was converted to HSV color space. The range of HSV values for the specific color was determined and the

resultant image was converted to a binary image by applying a specific threshold value. The resultant image has some noise, therefore, morphological operations were performed.

Figure 2.2 Original and binary image after RGB image segmentation in corn

Figure 2.3 Yellow and brown pixel segmentation using HSV color segmentation algorithm in corn plants

Figure 2.4 Green, yellow and brown pixel segmentation using HSV color segmentation algorithm in corn plants

For hyperspectral image processing, a customized algorithm was built in MATLAB R2022a to convert 243 images per plant into a 3D image cube that represents the full range of spectra covered. At first, an NDVI image was calculated using the image band at 27 (670nm) and band 48(770nm), $(670nm - 770nm) \div (670nm + 770nm)$. A mask was developed to recognize the plant materials in all images. A threshold value of 0.25 for wheat and threshold of 0.27 for corn was found effective to separate foreground and background from the plants. The wavelengths 670nm and 770nm were chosen due to their effectiveness, popularity, and low level of noise. The resulting NDVI image consists of the pixels of plant materials including both stem and leaves. Therefore, another image at band 132(1160 nm) was chosen in which the leaves were brighter than the stem. This image along with the NDVI image was used together to further classify stem and leaf

pixels. For this specific study, pixels belonging to plant leaves were used to extract leaf-level spectral reflectance from all hyperspectral bands. Figure 2.5 and Figure 2.6 shows the hyperspectral image segmentation process for wheat and corn plants, respectively.

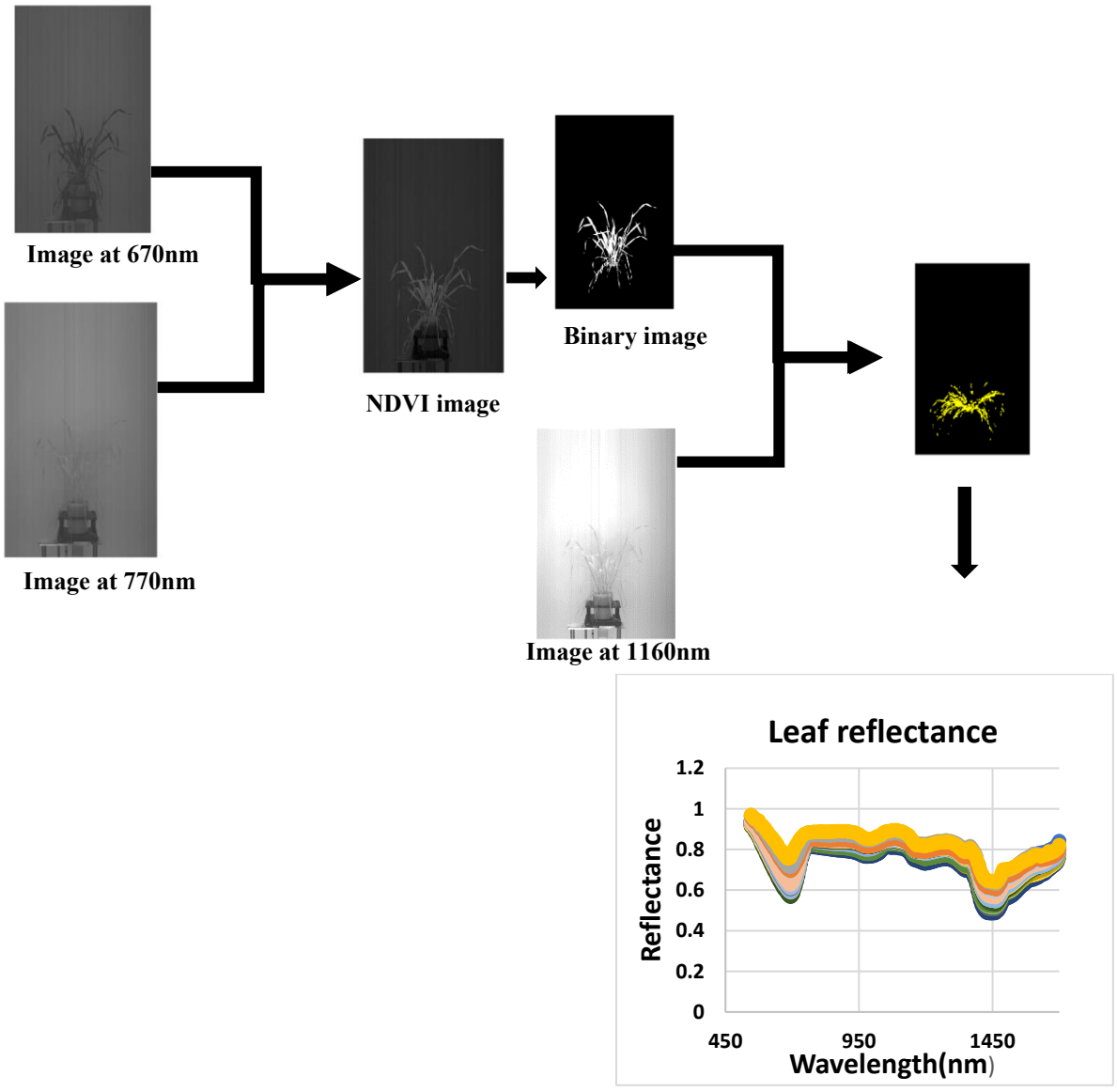


Figure 2.5 Hyperspectral image segmentation process of wheat plants

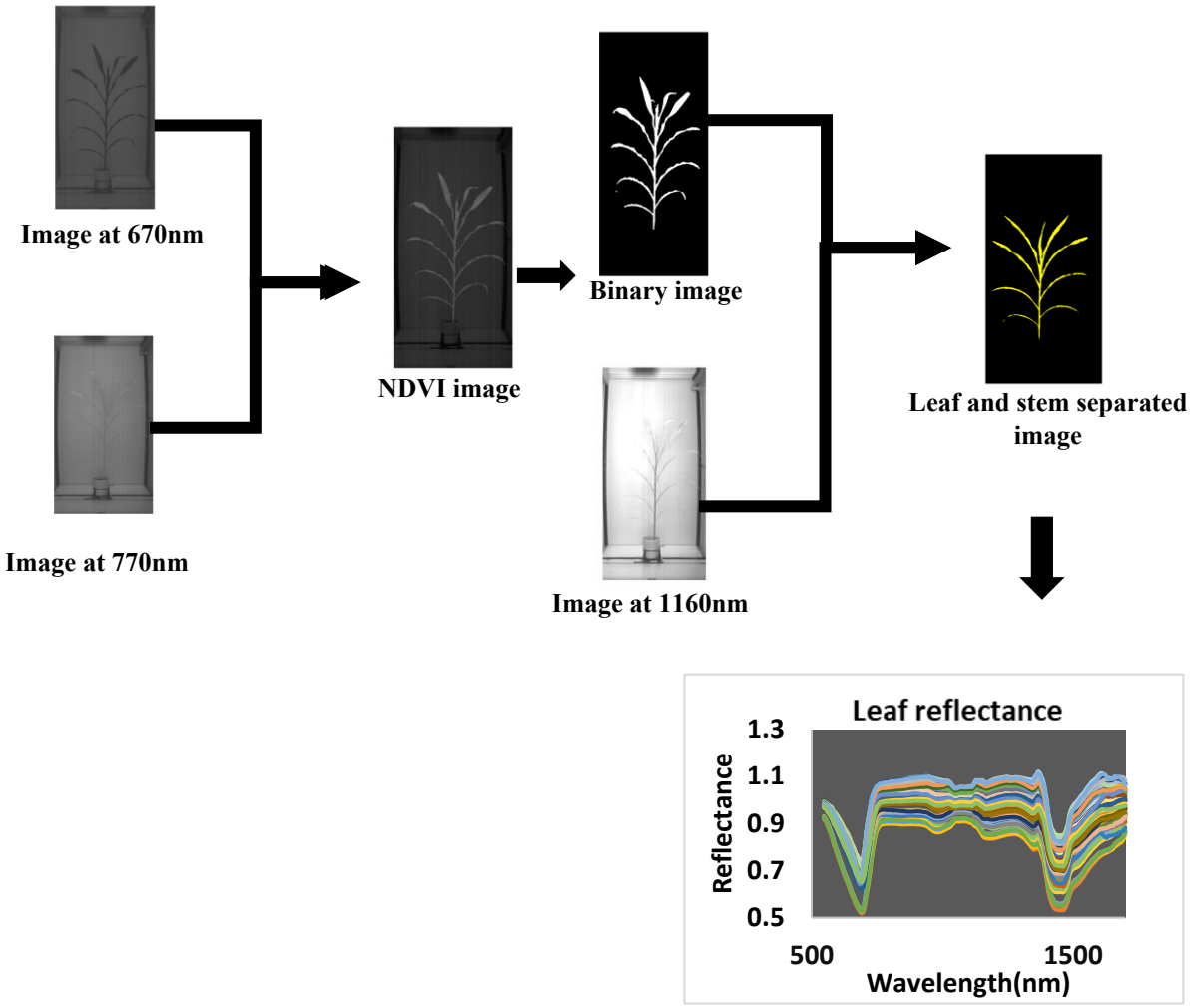


Figure 2.6 Hyperspectral image segmentation process of corn plants

2.5 Statistical methods

Partial least squares regression (PLSR) is a generalized form of multiple linear regression model that is suitable for highly correlated spectral data due to robustness in the presence of noise. This method reduces the available variables to a smaller set of predictors and performs regression which is mostly used when there is multicollinearity among the variables. The model was applied to predict biochemical content in wheat and corn using spectral data derived from two methods: the spectral data collected from the ASD

spectrometer and the reflectance extracted from hyperspectral images. The dataset was divided into 70% training and 30% testing. The size of the PLSR model (number of latent variables) was selected based on the number of latent variables giving the first lowest mean squared error of cross-validation and the selected variables capture the maximum covariance between predictor and response variables.

Similarly, random forest (RF) and support vector regression (SVR) models were also tested on the same datasets to compare the performance of the three models. RF model uses multiple decision trees, called number of estimators, which are merged for a more accurate prediction. The basic logic behind the random forest model is multiple uncorrelated models (the individual decision trees) perform better as a group than alone. The decision trees in random forest models were trained using bagging, bootstrapping, and aggregation. Bagging is an ensemble method that combines predictions from multiple machine learning algorithms to make more accurate predictions than an individual model. Bootstrapping randomizes the input data; the row of original data is picked randomly with replacement i.e. the same row can be presented in the dataset more than once. Aggregation reduces the sample datasets into summary statistics based on the observation and combines them.

SVR model finds a best-fit hyperplane that has a maximum number of points. The hyperparameters used in this model are kernel, C, and gamma. The kernel used was the radial basis function (rbf) which converts lower dimensional datasets to a higher dimension and helps to search a hyperplane in higher dimensional space. The C parameter helps to choose a decision boundary and controls how much importance the model places on minimizing errors in training data. A low C value produces a large

margin, more errors and underfit the training data but may better fit the unseen data whereas a high C value produces a small margin, less error but may overfit the unseen data. The gamma parameter decides how much curvature we want in a decision boundary and controls the distance of influence of a single data sample.

The chlorophyll content prediction in corn plants was done by calculating the modified chlorophyll absorption ratio index (MCARI) from the reflectance data obtained by using the black reference panel and white reference panel. The simple linear regression model was followed using the MCARI index as an independent variable and chlorophyll content as a dependent variable. The dataset was divided into a calibration set of 70% and a validation set of 30%. The total number of data points was 72 and the result was presented for the validation set only. The equation of MCARI index is given as:

$$MCARI\ index = (R_{700} - R_{670}) - 0.2 \times (R_{700} - R_{550})\left(\frac{R_{700}}{R_{670}}\right)$$

2.5.1 Model performance

Coefficient of determination (R^2) is used to evaluate the overall goodness of fit of the model. It represents the proportion of variance in the dependent variables that is explained by the independent variables in the model. The value usually ranges from 0 to 1 and the higher value indicates that the model fits better. The equation of Pearson's correlation coefficient is given as:

$$r = \frac{\Sigma(x_i - \bar{x_i})(y_i - \bar{y_i})}{\sqrt{\{\Sigma\{(x_i - \bar{x_i})^2 \Sigma(y_i - \bar{y_i})\}^2}}$$

Similarly, RMSE measures the average squared distance between the predicted and actual values in a dataset, and then takes the square root of that value. It indicates the average model prediction error or the average magnitude of the error or residuals between predicted and actual values. The lower value of RMSE indicates a better model fit.

$$RMSE = \sqrt{\sum_{i}^{n} \frac{(y_i - x_i)^2}{n}}$$

Where:

x_i is the measured value

y_i is the predicted value

$\widehat{x_i}$ is the mean of measured values

$\widehat{y_i}$ is the mean of predicted values

n is the number of observations

Overall, R^2 was used to evaluate the overall goodness of fit of the model, RMSE was used to evaluate the accuracy of the model and bias measured the systematic error that the model makes when it tries to approximate a target value. All three in combination were used to evaluate the performance of the three models.

2.5.2 Linear mixed effect model

The linear mixed effect (LME) model was performed among manually measured wheat datasets and using lme (Comprehensive R Archive Network (CRAN), 2023) package in Rstudio (R: The R Project for Statistical Computing, n.d.). LME is a statistical technique that accounts for both fixed and random effects. Twelve pairs of wheat genotypes were treated as fixed effects and replication within each genotype was treated as a random effect. The equation of the model is given as:

$$y = \beta + \beta_1 \times x_i + \gamma + \epsilon$$

where,

y is the response variable; Leaf area (LA), Fresh weight (FW), and Dry weight (DW)

β is the global intercept

x_i is the fixed effect of 24 different (12 pairs) wheat genotypes

γ is the random effect of replication

ε is a normal distribution of residuals

Again, a linear mixed model with yield as the response and replication as a random model effect was used to determine if there is significant difference in yield among the genotypes.

$$y_{ij} | \gamma_j \sim N(\mu_{ij}, \sigma^2)$$

$$\gamma_j \sim N(0, \sigma_j^2)$$

Where,

y_{ij} is the yield of i^{th} genotypes at j^{th} replication

μ is the overall mean

τ_i is the fixed effect of 24 different (12 pairs) wheat genotypes

γ_j is the random effect of replication

ϵ_{ij} is a normal distribution of residuals

2.5.3 Repeated measure model

RGB image analysis was performed for 192 wheat plants which were imaged daily in the greenhouse and plant pixels were extracted. 24 wheat genotypes were replicated eight times and the pixel count was extracted at four different time points; booting, stem elongation, anthesis and physiological maturity and the yield of each genotype was also recorded at the end of the experiment. The repeated measure model was used to determine if pixel count of each genotype is significantly different over time. The covariance structure used was an autoregressive parameter and the degrees of freedom was adjusted using a Kenward Rodgers adjustment. The same plant (i.e., same genotype) was measured at multiple time points. The equation of the model is given as:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + s_{ik} + w_{ijk}$$

$$s_{ik} \sim N(0, \sigma_k^2)$$

$$\boldsymbol{w}_{ik} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{ik})$$

$$\boldsymbol{w}_{ik} = [w_{i1k}, w_{i2k}, w_{i3k}, w_{i4k}]^T$$

$$\boldsymbol{\Sigma}_{ik} = \sigma_w^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

Where ρ is the autoregressive parameter.

Where,

y_{ijk} is the pixel count for the kth replication of the ith genotype at the jth time point

μ is the overall mean of the pixel count

α_i is the effect of the ith genotype

β_j is the effect of the jth time point

$(\alpha\beta)_{ij}$ is the interaction effect between the jth time point and the ith genotype

s_{ik} is the effect of the kth replication at ith genotype (random effect)

w_{ijk} is the residual error term

Chapter 3 Results

3.1 RGB image analysis of wheat plants

The linear regression analysis was performed on 39 wheat plants from three different growth stages: booting, stem elongation, and anthesis. One plant was removed from the analysis as an outlier. Figure 3.1 shows the correlation between destructively measured physical traits: total leaf area, fresh weight, and dry weight with plant pixel count. The correlation was higher for total leaf area with pixel count, with the coefficient of determination (R^2) of 0.756 followed by fresh weight ($R^2 = 0.740$) and dry weight ($R^2 =0.597$) as demonstrated in Figure 3.1(a), Figure 3.1(b), Figure 3.1(c) respectively. The R^2 of 0.756 for leaf area represents about 75.6% of variation in leaf area can be explained by pixel count. For total leaf area, the estimated coefficient for intercept was 15.195, and for average pixel count it was 143.64. The p values for all three traits indicate pixel count is statistically significant ($p<0.05$) in predicting total biomass area, fresh weight, and dry weight in wheat plants.

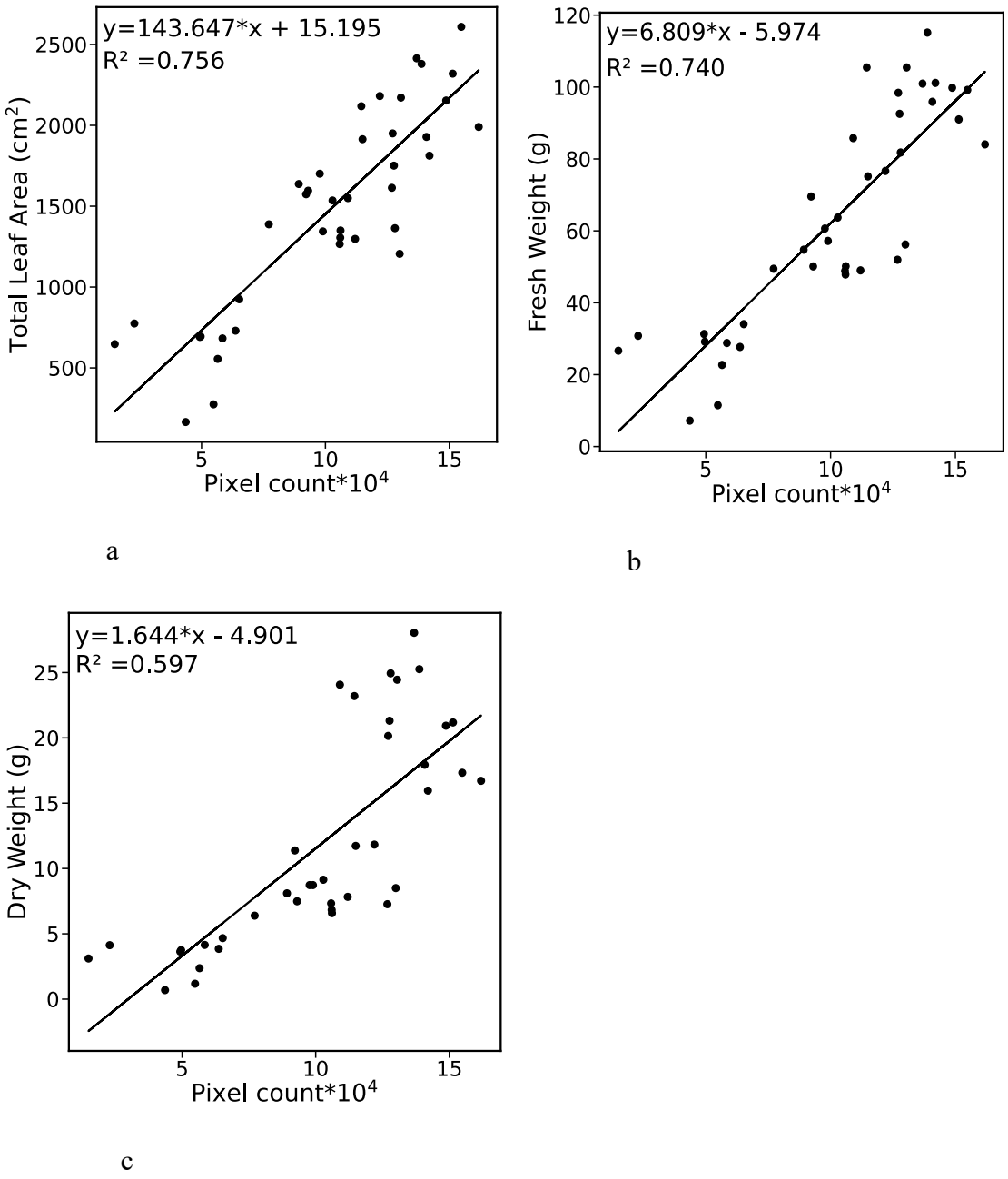


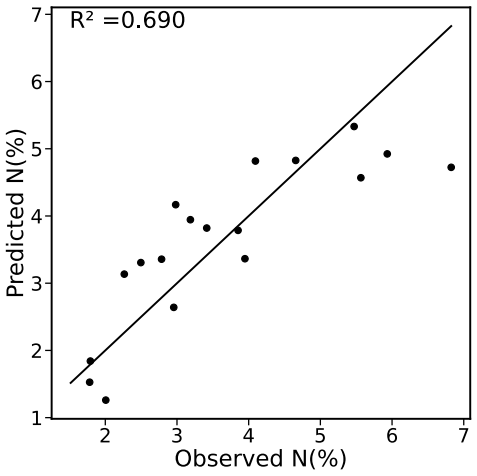
Figure 3.1.Correlation between total leaf area (a), fresh weight (b), and dry weight (c) with plant pixels derived from RGB image analysis of wheat plants

3.2 N, P, and K prediction using spectral reflectance from hyperspectral images and ASD in wheat plants

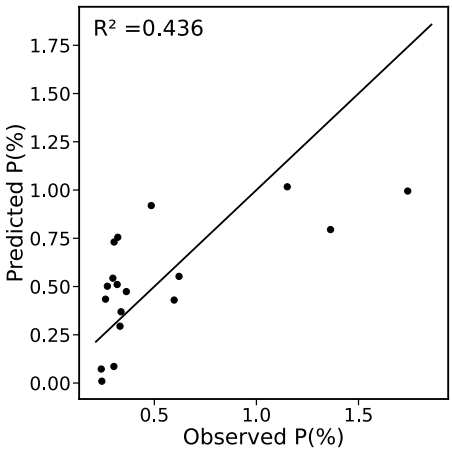
The hyperspectral image processing of wheat plants was done to extract leaf reflectance to predict macronutrient (N, P, K) content in plants. Figure 3.2 shows N, P, and K prediction among wheat genotypes using PLSR modeling. The leaf reflectance was extracted from hyperspectral images with wavelengths ranging from 546 to 1700 nm, which served as an input for the PLSR model. The number of samples used for the PLSR model was 57 wheat plant images taken on the same day before their destructive sampling. The wheat genotypes were selected randomly for manual measurement. Data from hyperspectral images showed the prediction accuracy was highest for nitrogen with R^2 of 0.690 followed by potassium ($R^2 = 0.684$) and phosphorous ($R^2 = 0.436$) as shown in Figure 3.2.The number of latent variables used were 5, 7, and 4 for the N, P, and K, respectively. Similarly, for N, P and K prediction using ASD, the result was shown in Figure 3.3. The prediction accuracy for K was found highest with R^2 of 0.772, followed by P ($R^2 = 0.702$) and N ($R^2 = 0.673$) as shown in Figure 3.3(c), Figure 3.3(b) and Figure 3.3(a) respectively.

The prediction accuracy for the N, P, K prediction from both datasets were compared. The N prediction accuracy was comparable among both datasets using PLSR modeling with an R^2 of 0.673 and RMSE of 0.807% from ASD and R^2 of 0.690 and RMSE of 0.814% from the hyperspectral image dataset. The correlation was higher from hyperspectral image dataset whereas the accuracy was higher from ASD with lower RMSE. But there was large variation in P and K prediction among the two datasets. The P prediction was higher from ASD datasets with R^2 of 0.702 and RMSE of 0.229% and K

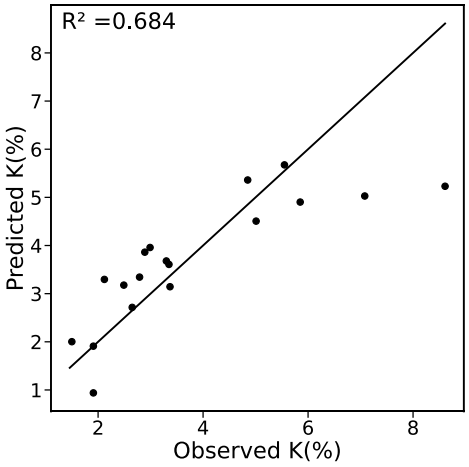
prediction was also higher from ASD datasets with R^2 of 0.772 and RMSE of 0.898%. Similarly, the reflectance from hyperspectral images shows negative bias of -0.037%, -0.003% and -0.104% for N, P and K prediction respectively. This shows the model underestimates the true value of N,P, and K and the predicted values are slightly lower than the actual values. Also, the ASD data shows positive bias of 0.019% for N, 0.046% for P and negative bias of -0.072% for K as shown in Table 3.1 with their respective R^2, RMSE, bias, and the numbers of latent variables used.



a

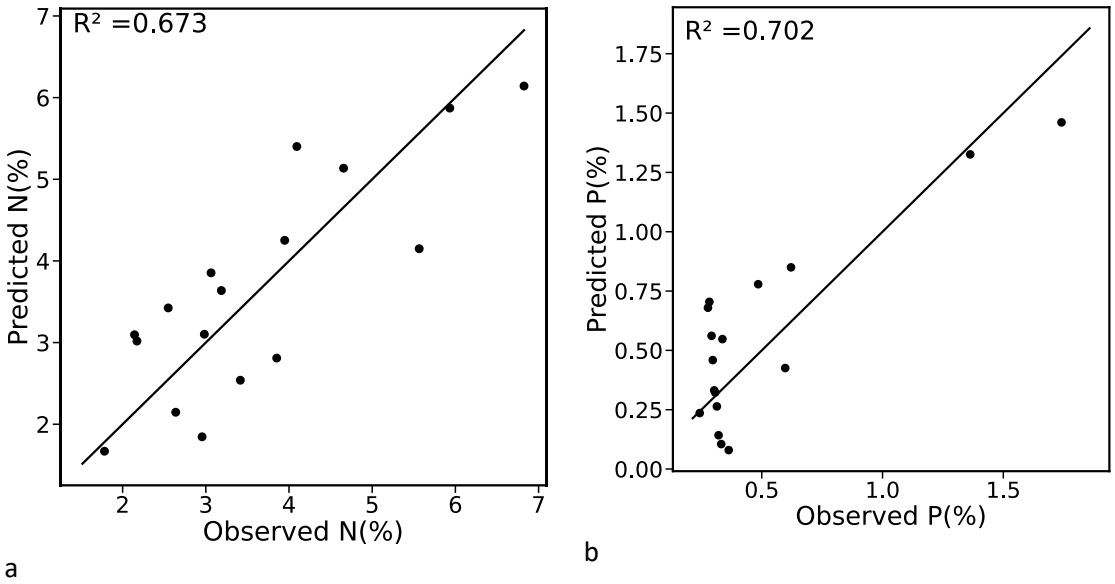


b



c

Figure 3.2.N, P and K prediction using PLSR model from hyperspectral images of wheat genotypes



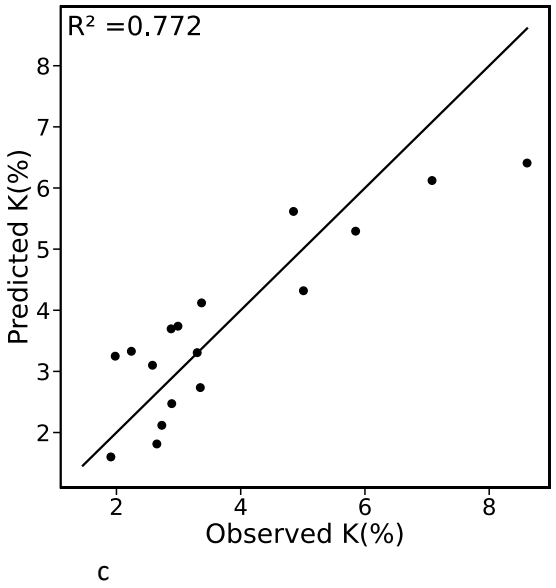


Figure 3.3.N, P and K prediction using PLSR model from ASD reflectance data of wheat genotypes.

Table 3.1 N, P and K prediction using both hyperspectral images and ASD in wheat genotypes using PLSR modeling

		R^2	RMSE (%)	Bias	Number of latent variables
N	Hyperspectral images	0.690	0.814	- 0.037	5
	ASD Spectrometer	0.673	0.807	0.019	5
P	Hyperspectral images	0.436	0.317	- 0.003	7
	ASD Spectrometer	0.702	0.229	0.046	5
K	Hyperspectral images	0.684	1.119	- 0.104	4
	ASD Spectrometer	0.772	0.898	- 0.072	5

3.3 Linear mixed effect model among wheat genotypes

A linear mixed effect model was performed to determine if there was any significant difference between wheat genotypes for manually measured physical traits. The genotypes for manual measurement were randomly selected. The result showed there was significant difference in dry weight, fresh weight, and biomass among the selected wheat genotypes for destructive sampling as the p-value was 0.0121, 0.0549, and 0.0045 respectively for all three traits.

Table 3.2 showed genotypes 67 and 83 have significantly different fresh weights with p-value of 0.0165 and 0.019 respectively. Similarly, genotype 67 has a coefficient of

59.403 and a standard error of 23.54 whereas genotype 83 has a coefficient of -44.18 with a standard error of 17.93.

Table 3.3 shows that genotypes 51, 67 and 83 have significantly different fresh weight as all of their p-values was less than 0.05. Similarly, Table 3.3 shows genotypes 69, 86, 88, and 114 have significantly different dry weights with their respective p-value of <0.05.

Table 3.2.Wheat genotypes with significantly different fresh weight

	Coefficient	Standard error	DF	t-value	p-value
Intercept	75.3771	11.82	34	6.37	0.000
Genotype 67	59.403	23.54	34	2.522	0.0165
Genotype 83	-44.18	17.937	34	-2.46	0.0190

Table 3.3.Wheat genotypes with significantly different leaf area

	Coefficient	Standard error	DF	t-value	p-value
Intercept	1503.59	225.59	34	6.665	0.000
Genotype 51	707.149	315.357	34	2.2423	0.0316
Genotype 67	1627.4062	476.74	34	3.413	0.0017
Genotype 83	-958.0856	363.047	34	-2.639	0.0125

Table 3.4.Wheat genotypes with significantly different dry weight

	Coefficient	Standard error	DF	t-value	p-value
Intercept	20.29	6.49	34	3.12	0.0036
Genotype 69	19.294	6.66	34	2.893	0.0066
Genotype 86	16.82	7.306	34	2.302	0.0276
Genotype 88	13.98	6.669	34	2.097	0.0434
Genotype 114	39.066	11.035	34	3.539	0.0012

Similarly, a linear mixed model with yield as the response and replication as a random model effect was used. The result showed there was significant difference in yield among the genotypes with p value <0.0001 as shown in Table 3.5. A plot of the least squares means yield and the confidence intervals is included in the figure. A pairwise comparison was done to see the yield difference among the genotypes. Only significantly different results were included in the table. Similarly, a plot of the least squares means yield and the confidence intervals was included in the Figure 3.4.

Table 3.5 Type III test of fixed effects

Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Genotype	23	161	4.78	<.0001

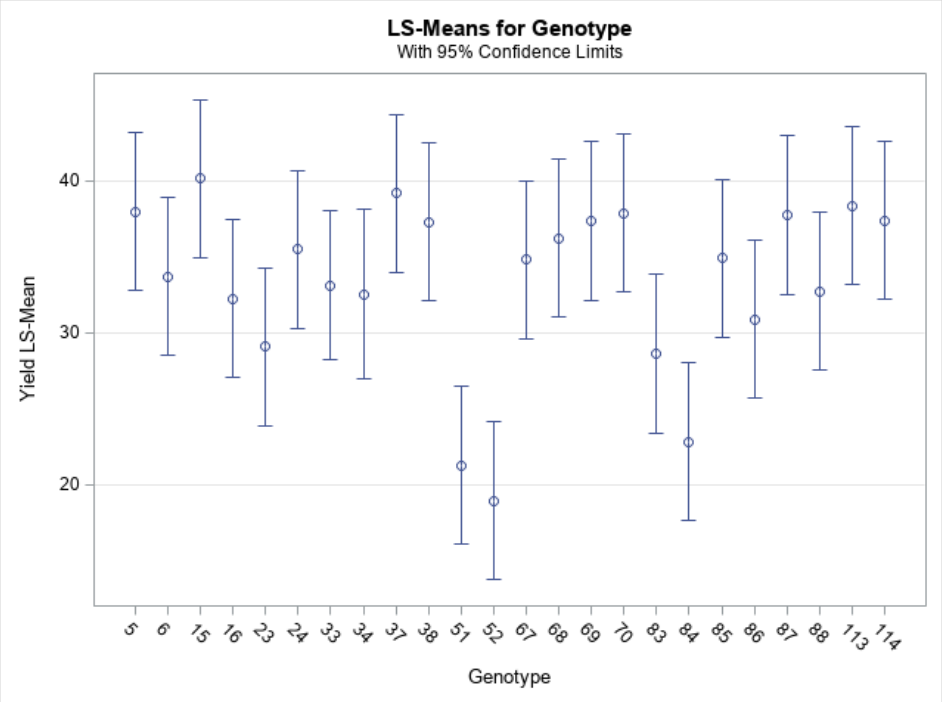


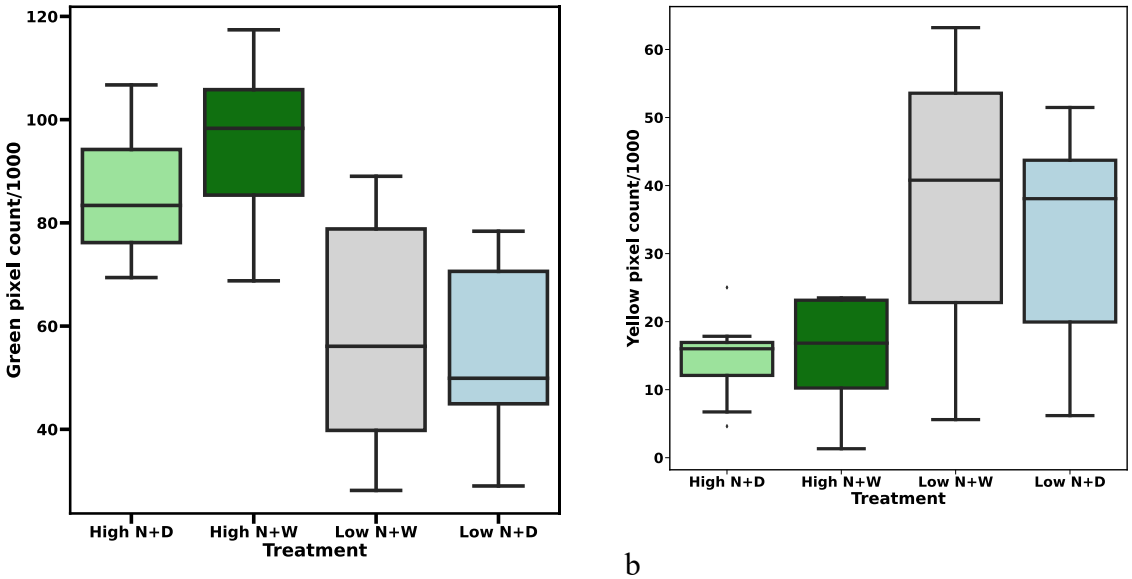
Figure 3.4 LS-Means of yield of each wheat genotypes

A pairwise comparison was done to see the yield difference among the wheat genotypes. Only significantly different results are included in the table.

3.5 RGB image analysis of corn plants

RGB image analysis of corn plants was done in three different ways as explained in chapter 2 of this study. Figure 3.5 shows the distribution of green pixel count and yellow and brown pixel count after RGB image segmentation. Corn plants that received water and high nitrogen have the highest green pixels and the plants with drought and low nitrogen have the lowest green pixels count as shown in Figure 3.5 (a). Similarly, plants with drought and high nitrogen have higher green pixels compared to plants with water and low nitrogen. As shown in Figure 3.5 (b) plants with drought and low nitrogen and water and low nitrogen treatment has the highest yellow pixels with a slightly larger range for water and nitrogen treatment as expected. Nitrogen and water are considered the most important components for supporting the growth and development of plants. Nitrogen is a part of the chlorophyll molecule and gives green color to plants that are involved in creating food through the photosynthesis process. Therefore, lack of enough

nitrogen in plants shows yellowing (chlorosis) of the plants which is shown clearly in the figure.



a b

Figure 3.5 Distribution of green pixels (a) and yellow and brown pixels (b) among different treatment combinations; D+0N (Drought and low nitrogen), D+N (Drought and high nitrogen), W+0N (well-watered and low nitrogen), and W+N (well-watered and high nitrogen)

The linear regression model was applied to 72 corn plant samples.

Figure 3.6 shows the correlation between physical traits and plant pixel count, the correlation was higher between leaf area and plant pixels with R^2 of 0.679, followed by dry weight (R^2 = 0.398) and fresh weight (R^2 = 0.382) respectively. Overall, the plants that received water and high nitrogen are seen higher in the scatter plots and the plants with drought and low nitrogen treatment were at the lower end of the scatter plot.

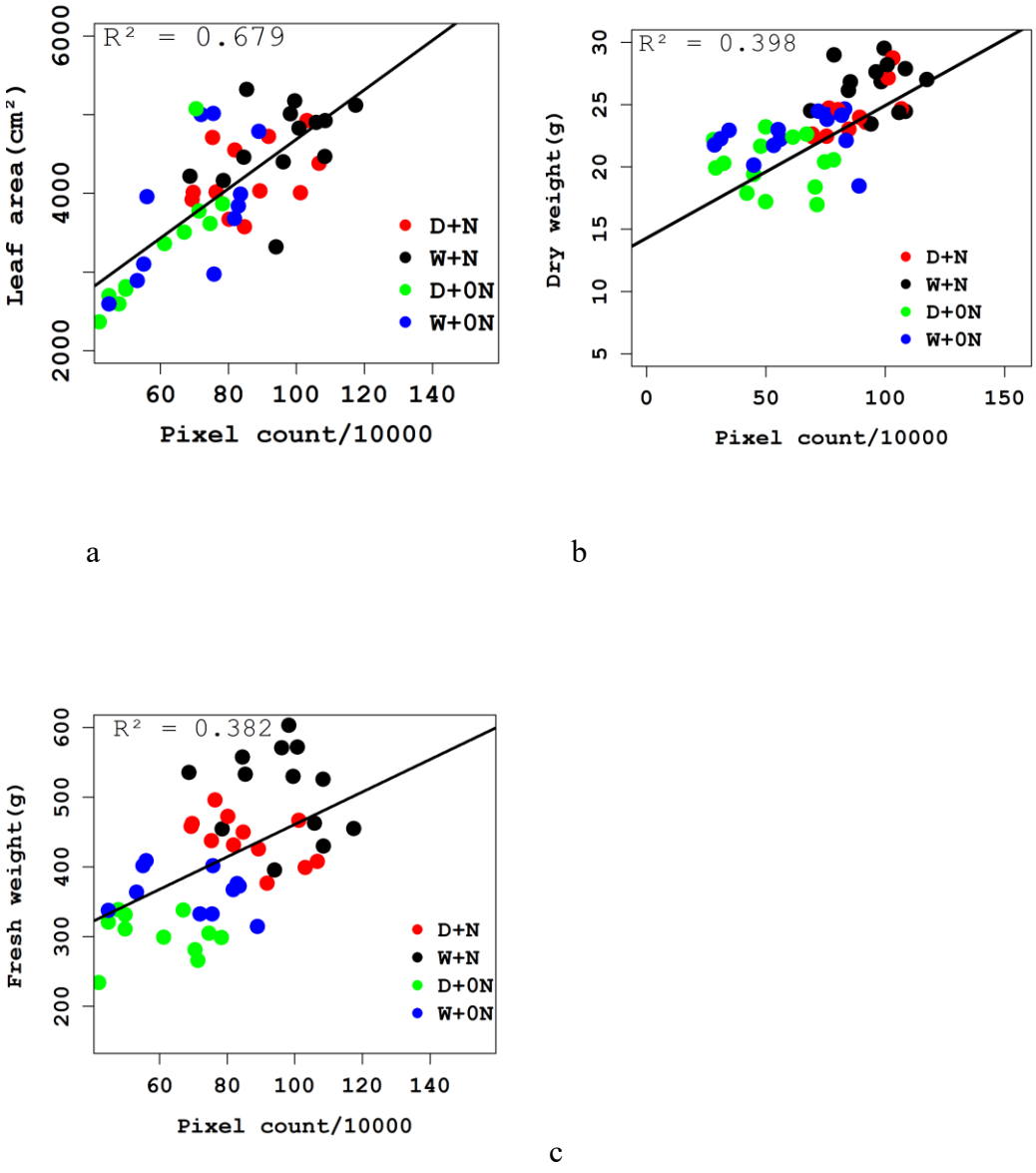


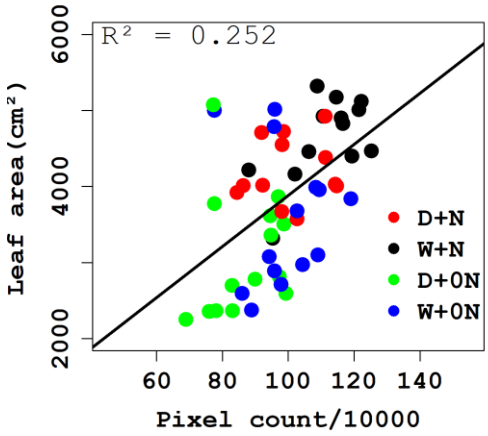
Figure 3.6 Correlation between leaf area, dry weight, and fresh weight with plant pixels from excessive green pixel extraction algorithm in corn plants.

The plants with water and low nitrogen as well as the plants with drought and low nitrogen treatment were seen at the lower end, which shows nitrogen has a significant role in the growth and development of corn plants. However, the plants that received high nitrogen and well-watered has the points at the upper end of the regression line. Even

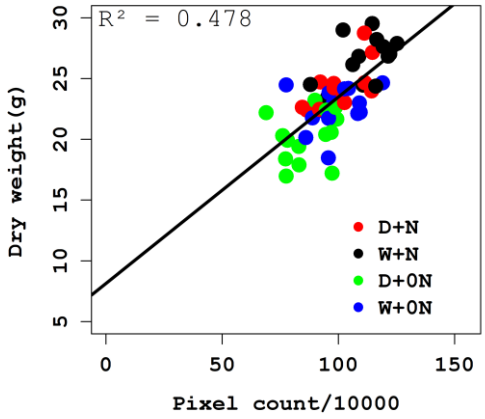
though the result from the RGB image analysis using the green pixel extraction algorithm was lower compared to other studies, it might have implications with different treatment combinations of nitrogen and water. Overall, the correlation was found to be stronger for plants that received nitrogen as well as water treatment.

Another RGB image segmentation approach was followed by extracting yellow and brown pixels using the HSV color segmentation algorithm. The green plant pixels extracted from the green pixel extraction algorithm and the yellow and brown pixels extracted from HSV color segmentation algorithms were added and counted as a total plant pixel count of each plant. A linear regression model was applied to see the difference in correlation, but the correlation was lower compared to the excessive green pixel extraction algorithm as shown in Figure 3.7.The highest correlation from this method was obtained for dry weight with R^2 of 0.478, followed by fresh weight (R^2 = 0.442) and leaf area (R^2 = 0.252) as shown in Figure 3.7.Overall, the correlation for plants with water and nitrogen treatment was higher which was consistent with the excessive green pixel extraction algorithm.

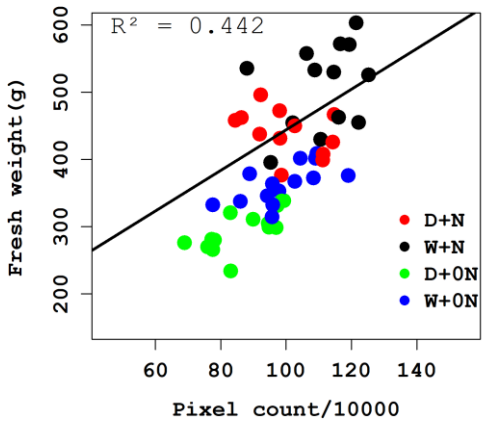
Similarly, the result from the HSV segmentation algorithm extracting all green, yellow, and brown pixels is presented in Figure 3.8. The correlation was higher for dry weight with pixel count with an R^2 of 0.214 followed by fresh weigh (R^2=0.192) and leaf area (R^2 =0.190) as shown in Figure 3.8. The correlation between leaf area and pixel count was comparable with the previous algorithm presented in Figure 3.7 but the correlation of pixel count with fresh weight and dry weight than the previous algorithm.



a

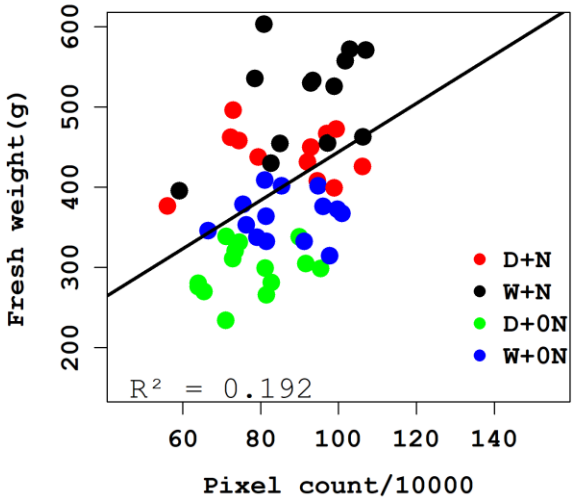


b

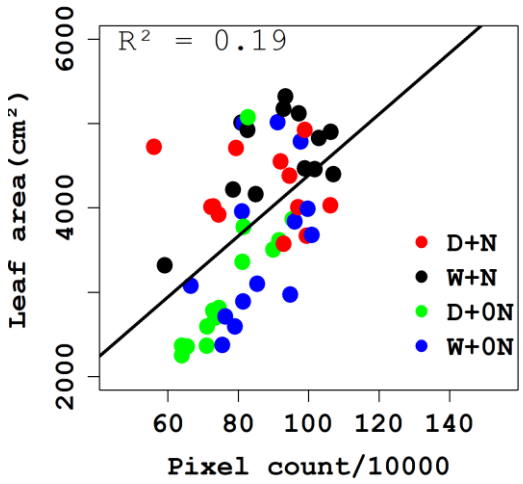


c

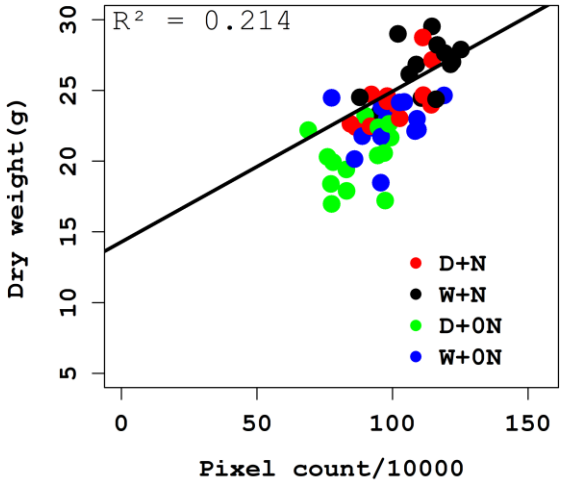
Figure 3.7 Correlation between leaf area, dry weight, and fresh weight with plant pixels from the excessive green index and HSV color segmentation algorithm



a



b



c

Figure 3.8 Correlation between leaf area, dry weight, and fresh weight with plant pixels extracted from HSV color segmentation algorithm including green, yellow, and brown

3.6 N, P, and K prediction using reflectance from hyperspectral images and ASD in corn plants using three different machine learning algorithms

Table 3.6 shows the comparison between three different machine learning models; PLSR, RF, and SVR to predict N, P, and K content using reflectance from hyperspectral image datasets and reflectance from ASD spectrometer in corn plants. For hyperspectral image datasets, N prediction was higher from the PLSR model with the R^2 value of 0.663, RMSE of 0.271%, and bias of 0.092% as shown in Figure 3.9 (a). Similarly, P was predicted higher from the RF model with R^2 of 0.740, RMSE of 0.019%, and negative bias of – 0.0023% and K was also predicted higher from the RF model with R^2 of 0.876, RMSE of 0.27%, with positive bias of 0.044%.as shown in Figure 3.9 (b) and Figure 3.9 (c) respectively.

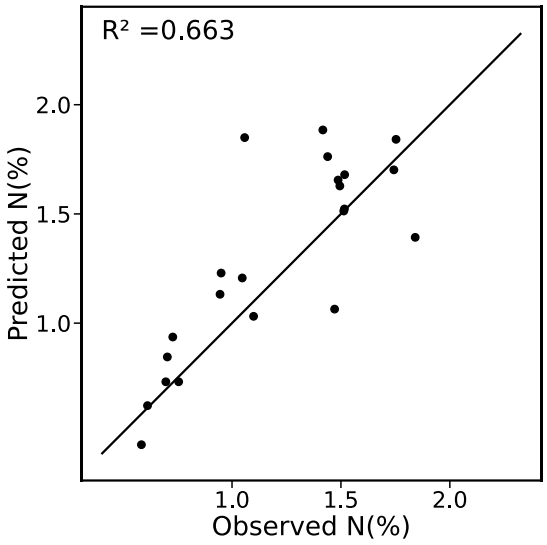
For the spectral reflectance data obtained from the ASD spectrometer, the result shows N, P, and K prediction was higher than the RF model. The R^2 for N prediction was 0.676 with an RMSE of 0.34% and negative bias of -0.075%. Similarly, the R^2 value for P prediction was 0.41 with an RMSE of 0.027% with negative bias of – 0.0033%. Among N, P, and K from spectral reflectance datasets, the K prediction accuracy was higher with R^2 of 0.69, RMSE of 0.382 %, and negative bias of -0.063 as shown in Figure 3.10.

Although the performance of the PLSR and RF models was comparable, the SVR model showed lower prediction for N, P, and K contents from either dataset. Comparing the N prediction from both datasets, the prediction accuracy was found to be similar from both PLSR and RF model with R^2 of 0.66 and 0.67 respectively. However, the RF model was effective in prediction P and K from both datasets with a coefficient of determination

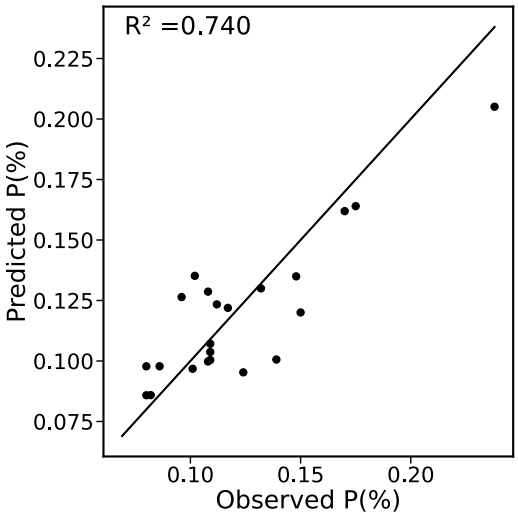
slightly higher in hyperspectral image datasets compared to reflectance data from ASD as demonstrated in Table 3.6.

Table 3.6.N, P and K prediction using three different machine learning models using hyperspectral image datasets and ASD datasets in corn plants

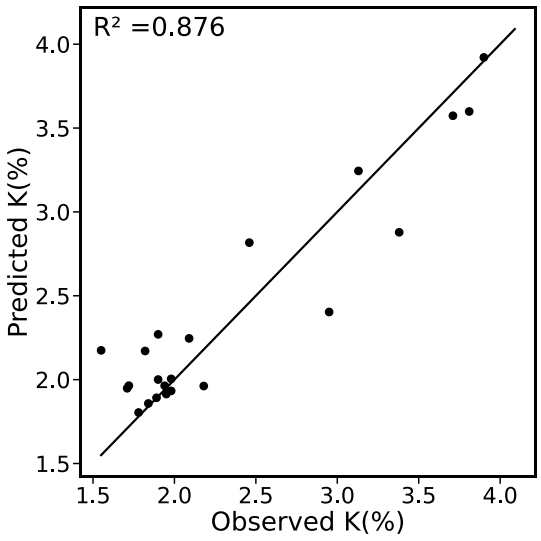
Macronutrients	**Datasets**	**Number of Sample size**	**Number of latent variables**	**Partial least square regression (PLSR)**			**Random forest regression (RF)**			**Support vector regression (SVR)**		
				R^2	RMSE (%)	Bias	R^2	RMSE (%)	Bias (%)	R^2	RMSE (%)	Bias
Nitrogen	Hyperspectral images	72	5	0.663	0.271	0.092	0.591	0.362	-0.024	0.124	0.441	0.0122
	ASD	72	4	0.596	0.333	0.119	0.676	0.344	-0.075	0.037	0.554	0.062
Phosphorous	Hyperspectral images	72	7	0.255	0.028	0.010	0.740	0.019	-0.0023	_	0.043	0.023
	ASD	72	4	0.055	0.0281	0.001	0.413	0.027	-0.0033	_	0.043	0.027
Potassium	Hyperspectral images	72	7	0.630	0.472	0.293	0.876	0.27	0.044	0.471	0.567	0.131
	ASD	72	6	0.680	0.381	0.202	0.690	0.382	-0.063	0.014	0.645	0.117



a

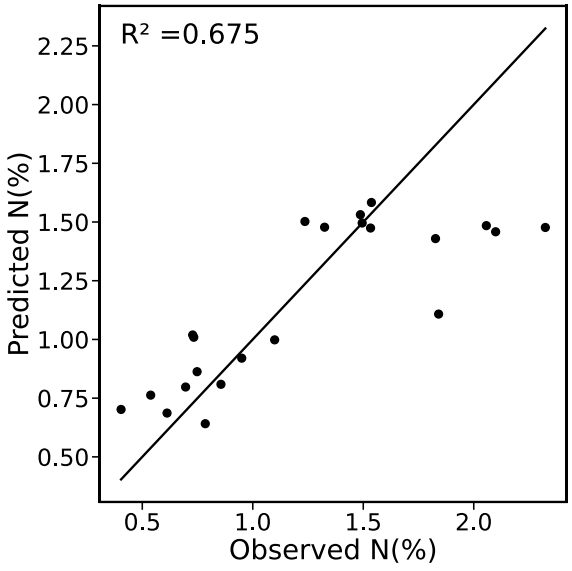


b

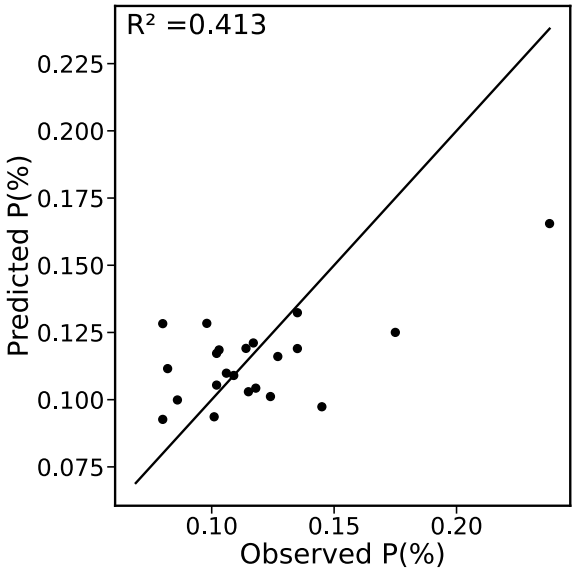


c

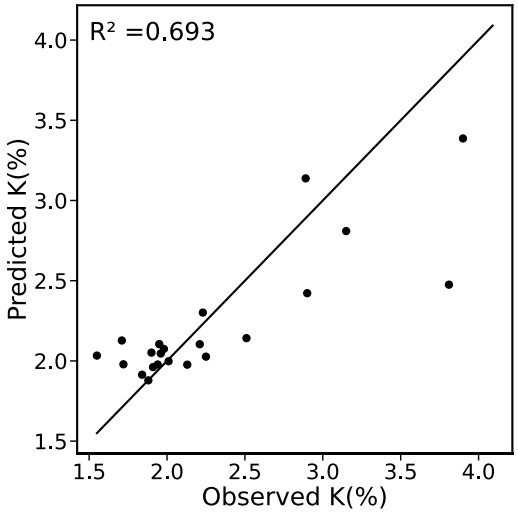
Figure 3.9 . Highest N (from PLSR model), P (from RF model), and K (from RF model) predicted with hyperspectral images reflectance among three different machine learning models



a



b

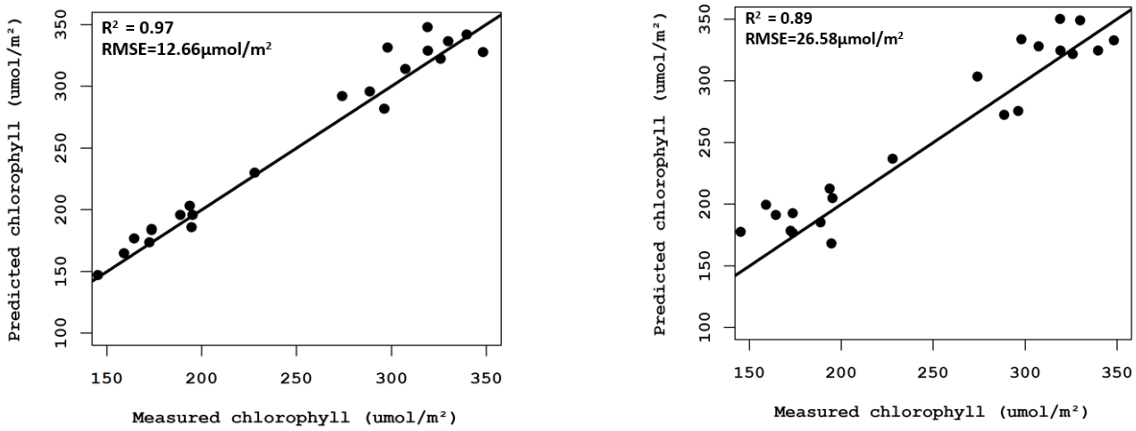


c

Figure 3.10 . Highest N (RF), P (RF), and K (RF) predicted from ASD reflectance among three different machine learning models

3.7 Chlorophyll content prediction

Table 3.7 shows there is a highly significant relationship between chlorophyll content and MCARI index as the p-value is <0.001 for both reflectance data using a black and white reference panel. The estimated effect of the MACARI index using a black reference panel was 0.97 with a standard error of 0.068 whereas the estimated effect of the MACARI index using a white reference panel was found 1.002 with a standard error of 0.031. Overall, the correlation was higher for chlorophyll content prediction using a white reference panel with R^2 of 0.97 and RMSE of 12.66 µmol/m^2 as indicated in Figure 3.11 (a). Similarly, R^2 of 0.89 and RMSE of 26.58 µmol/m^2 was obtained from the black reflectance panel, as shown in Figure 3.11 (b). This result is consistent with other studies that used hyperspectral spectral data to estimate leaf chlorophyll content, which generally yielded R^2 values > 0.90 (J. Li et al., 2023 and Wijewardane et al., 2023).



a

b

Figure 3.11 Correlation between predicted and measured chlorophyll content using MCARI index with reflectance from white reference panel (a) and reflectance from black reference panel (b)

Table 3.7.Chlorophyll content prediction using a linear regression model with ASD reflectance data taken from the black and white reference panel.

		Estimates	Standard Error	t-value	Pr(>|t|)
Black reference panel	Intercept	16.13	16.97	0.95	0.352
	MACARI index	0.97	0.068	14.17	1.54e-12***
White reference panel	Intercept	4.98	7.77	0.641	0.528
	MACARI index	1.002	0.031	31.96	<2.2e-16***

Chapter 4 Discussion

The main objective of this study was to characterize the physical and biochemical traits of wheat and corn plants using high throughput image analysis. To quantify this objective wheat and corn plants were grown on the LemnaTec greenhouse plant phenotyping facility at the University of Nebraska Lincoln. The RGB and hyperspectral images of the plants were captured regularly using an automated imaging system and the ground truth measurements of the plants were taken through destructive sampling. The whole study was based on the hypothesis that high throughput plant phenotyping using image analysis techniques, especially RGB and hyperspectral imaging, could accurately and effectively determine physical and biochemical traits in plants. To test this hypothesis, the experiment was performed, and different methodologies were followed as described in Chapter 2 of this thesis.

The first objective of the study was to determine the degree of correlation between plant pixel information derived from RGB image analysis with manually measured physical traits for wheat and corn plants. The findings from RGB image analysis demonstrate that the correlation was higher for leaf area followed by fresh weight and dry weight for both wheat and corn plants. The correlation was found higher when the plants were at earlier growth stages and the correlation started to decrease at the later developmental stage of plants. The overall correlation of plant pixels with manually measured physical traits was found to be higher for wheat plants compared to corn plants. The wheat plant architecture is different compared to the plants like maize and sorghum on which the studies were done before. Therefore, the correlation might be affected due to the difference in the

architecture of the wheat plants. Also, the distance from the camera to the object was also larger in RGB and hyperspectral imaging for wheat plants which could be the other factor for affecting correlation in wheat plants.

The second objective of the study was to determine biochemical traits prediction accuracy between hyperspectral imaging and non-imaging spectrometer (ASD). The result showed reflectance data obtained using hyperspectral image processing was as effective as reflectance data obtained using ASD to predict nitrogen content in both wheat and corn plants. However, the prediction accuracy of P and K differed largely in both crops. For wheat crops, N prediction from hyperspectral images and N prediction from ASD reflectance were found similar using PLSR modeling whereas P and K prediction was found higher from ASD reflectance data.

Similarly, the third specific objective was to determine the best predictive machine learning model to determine biochemical traits using both hyperspectral image datasets and ASD reflectance specifically on corn plants. Three different machine learning algorithms were performed and compared: PLSR, RF, and SVR. The result showed, for hyperspectral image datasets, nitrogen prediction accuracy was higher using PLSR modeling whereas P and K prediction was higher using the random forest model. For ASD datasets, the N, P, and K prediction accuracy was found higher from the random forest model.

This study highlighted the benefits of image analysis techniques to capture a large amount of data in a relatively short amount of time and provided a good insight for performing experiments on plants like wheat that has different architecture from corn. This study also explored more image segmentation algorithms for plants treated with

various nitrogen and water levels and could set a standard benchmark protocol for ground truth measurement. This study has provided insights regarding the time and labor required for plant phenotyping, allowing for more efficient and effective breeding programs where a large number of plants should be screened to identify desirable traits.

However, it is well noted that this study has some limitations and challenges that need to be addressed. More robust and standardized measurement protocols need to be established during imaging and ground truth data collection to further improve the results. Since the nitrogen dose applied was usually lower than the normal rate for corn plants, it was not sufficient for the effective growth and development of corn plants. Therefore, even high nitrogen-treated plants developed yellow color, and the plant with lower nitrogen treatment had worse symptoms in the early growth stages, which may have influenced the correlation with RGB imaging for corn plants. So, the appropriate amount of nitrogen application should be considered during future studies.

Chapter 5 Conclusion and Future work

In conclusion, this study has demonstrated the potential to characterize the physical, biochemical and physiological traits of wheat and corn plants using high throughput image analysis. Overall, the correlation was higher for leaf area/total biomass area with plant pixel followed by fresh weight and dry weight for both wheat and corn plants. The correlation between manually measured physical traits and data from RGB image analysis was found higher for wheat plants compared to corn plants. The nitrogen prediction was higher using PLSR modeling in both crops. The nitrogen prediction accuracy was comparable using data from hyperspectral images and ASD spectrometer, whereas the prediction accuracy of P and K differed largely in both crops.

It can be concluded that the use of image analysis techniques has the potential to revolutionize the field of plant phenotyping to measure desired plant traits more efficiently and accurately on a large scale. While there are still challenges to be addressed, such as developing standardized protocols for image acquisition and improving the accuracy of machine learning algorithms used for plant trait quantification. For future studies, the potential of other imaging techniques, such as fluorescence and thermal, can be further explored to determine other physiological traits.

In conclusion, the issues and challenges addressed in this thesis will help future researchers to replicate and improve the methods and show the extent and possibilities of using high throughput plant phenotyping using image analysis techniques. The findings from this study can also be useful in replicating similar experiments in other crop species having different architecture. These image analysis techniques have the potential to revolutionize the field of plant phenotyping by advancing our understanding of important

plant traits, which ultimately contribute to crop improvement addressing global challenges related to food security and climate change.

Chapter 6 References

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., Salvucci, M. E., & White, J. W. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology*, *41*(1), 68–79. https://doi.org/10.1071/FP13126

Approach, T., Kullar, R., Wenzler, E., Alexander, J., Goldstein, E. J. C., Chicago, I., Florida, A. C., Stewardship, E., & Beach, N. (2022). *ce pt e d us cr ip t Ac ce pt e d us cr*. 1–76.

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, *19*(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Areal, F. J., Jones, P. J., Mortimer, S. R., & Wilson, P. (2018). Measuring sustainable intensification: Combining composite indicators and efficiency analysis to account for positive externalities in cereal production. *Land Use Policy*, *75*(February), 314–326. https://doi.org/10.1016/j.landusepol.2018.04.001

Bai, G., Ge, Y., Hussain, W., Baenziger, P. S., & Graef, G. (2016). A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. *Computers and Electronics in Agriculture*, *128*, 181–192. https://doi.org/10.1016/j.compag.2016.08.021

Bai, G., Ge, Y., Scoby, D., Leavitt, B., Stoerger, V., Kirchgessner, N., Irmak, S., Graef, G., Schnable, J., & Awada, T. (2019). NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing,

and agronomic research. *Computers and Electronics in Agriculture*, *160*(January), 71–81. https://doi.org/10.1016/j.compag.2019.03.009

Bai, G., Jenkins, S., Yuan, W., Graef, G. L., & Ge, Y. (2018). Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. *Frontiers in Plant Science*, *9*(July), 1–12. https://doi.org/10.3389/fpls.2018.01002

Baker, R. E., Peña, J. M., Jayamohan, J., & Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, *14*(5), 1–4. https://doi.org/10.1098/rsbl.2017.0660

Correia, P. M. P., Cairo Westergaard, J., Bernardes da Silva, A., Roitsch, T., Carmo-Silva, E., & Marques da Silva, J. (2022). High-throughput phenotyping of physiological traits for wheat resilience to high temperature and drought stress. *Journal of Experimental Botany*, *73*(15), 5235–5251. https://doi.org/10.1093/jxb/erac160

Ge, Y., Atefi, A., Zhang, H., Miao, C., Ramamurthy, R. K., Sigmon, B., Yang, J., & Schnable, J. C. (2019). High-throughput analysis of leaf physiological and chemical traits with VIS-NIR-SWIR spectroscopy: A case study with a maize diversity panel. *Plant Methods*, *15*(1), 1–12. https://doi.org/10.1186/s13007-019-0450-8

Ge, Y., Bai, G., Stoerger, V., & Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Computers and Electronics in Agriculture*, *127*, 625–632. https://doi.org/10.1016/j.compag.2016.07.028

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food Security: The

Challenge of Feeding 9 Billion People. Science 327:812-818. *Science*, *327*(February), 812–818. http://www.elgaronline.com/view/9780857939371.xml

Granier, C., & Vile, D. (2014). Phenotyping and beyond: Modelling the relationships between traits. *Current Opinion in Plant Biology*, *18*(1), 96–102. https://doi.org/10.1016/j.pbi.2014.02.009

Jay, S., Maupas, F., Bendoula, R., & Gorretta, N. (2017). Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research*, *210*(June), 33–46. https://doi.org/10.1016/j.fcr.2017.05.005

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Regression Models*. 35–133. https://doi.org/10.1007/978-3-319-33153-9_2

Kamali, M. I., & Nazari, R. (2018). Determination of maize water requirement using remote sensing data and SEBAL algorithm. *Agricultural Water Management*, *209*(January), 197–205. https://doi.org/10.1016/j.agwat.2018.07.035

Li, D., Quan, C., Song, Z., Li, X., Yu, G., Li, C., & Muhammad, A. (2021). High-Throughput Plant Phenotyping Platform (HT3P) as a Novel Tool for Estimating Agronomic Traits From the Lab to the Field. *Frontiers in Bioengineering and Biotechnology*, *8*(January), 1–24. https://doi.org/10.3389/fbioe.2020.623705

Li, J., Veeranampalayam-Sivakumar, A. N., Bhatta, M., Garst, N. D., Stoll, H., Stephen Baenziger, P., Belamkar, V., Howard, R., Ge, Y., & Shi, Y. (2019). Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. *Plant Methods*, *15*(1), 1–13. https://doi.org/10.1186/s13007-019-0508-7

Li, J., Wijewardane, N. K., Ge, Y., & Shi, Y. (2023). Improved chlorophyll and water content estimations at leaf level with a hybrid radiative transfer and machine learning model. *Computers and Electronics in Agriculture*, *206*(January), 107669. https://doi.org/10.1016/j.compag.2023.107669

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

Liang, Z., Pandey, P., Stoerger, V., Xu, Y., Qiu, Y., Ge, Y., & Schnable, J. C. (2018). Conventional and hyperspectral time-series imaging of maize lines widely used in field trials. *GigaScience*, *7*(2), 1–11. https://doi.org/10.1093/GIGASCIENCE/GIX117

Liu, S., Baret, F., Abichou, M., Boudon, F., Thomas, S., Zhao, K., Fournier, C., Andrieu, B., Irfan, K., Hemmerlé, M., & Solan, B. de. (2017). Estimating wheat green area index from ground-based LiDAR measurement using a 3D canopy structure model. *Agricultural and Forest Meteorology*, *247*(November 2016), 12–20. https://doi.org/10.1016/j.agrformet.2017.07.007

Miao, C., Guo, A., Thompson, A. M., Yang, J., Ge, Y., & Schnable, J. C. (2021). Automation of leaf counting in maize and sorghum using deep learning. *Plant Phenome Journal*, *4*(1), 1–15. https://doi.org/10.1002/ppj2.20022

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science*, *282*(January), 60–72. https://doi.org/10.1016/j.plantsci.2019.01.007

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-

based plant disease detection. *Frontiers in Plant Science*, *7*(September), 1–10. https://doi.org/10.3389/fpls.2016.01419

Morgounov, A., Gummadov, N., Belen, S., Kaya, Y., Keser, M., & Mursalova, J. (2014). Association of digital photo parameters and NDVI with winter wheat grain yield in variable environments. *Turkish Journal of Agriculture and Forestry*, *38*(5), 624–632. https://doi.org/10.3906/tar-1312-90

Munns, R., James, R. A., Sirault, X. R. R., Furbank, R. T., & Jones, H. G. (2010). New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. *Journal of Experimental Botany*, *61*(13), 3499–3507. https://doi.org/10.1093/jxb/erq199

Pandey, P., Ge, Y., Stoerger, V., & Schnable, J. C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Frontiers in Plant Science*, *8*(August), 1–12. https://doi.org/10.3389/fpls.2017.01348

Perez-Sanz, F., Navarro, P. J., & Egea-Cortines, M. (2017). Plant phenomics: An overview of image acquisition technologies and image data analysis algorithms. *GigaScience*, *6*(11), 1–18. https://doi.org/10.1093/gigascience/gix092

Postic, F., Beauchêne, K., Gouache, D., & Doussan, C. (2019). Scanner-based minirhizotrons help to highlight relations between deep roots and yield in various wheat cultivars under combined water and nitrogen deficit conditions. *Agronomy*, *9*(6). https://doi.org/10.3390/agronomy9060297

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., & French, A. P. (2017). Deep learning for multi-task plant phenotyping. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, *2018-*

Janua, 2055–2063. https://doi.org/10.1109/ICCVW.2017.241

Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., & Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Frontiers in Plant Science*, *6*(AUG), 1–15. https://doi.org/10.3389/fpls.2015.00619

Shamshiri, R., Mahadi, M. R., Ahmad, D., Bejo, S. K., Aziz, A., Ishak, W., Ismail, W., & Man, H. C. (2017). Controller Design for an Osprey Drone to Support Precision Agriculture Research in Oil Palm Plantations Written for presentation at the 2017 ASABE Annual International Meeting Sponsored by ASABE. *An ASABE Meeting Presentation*, *1700014*, 2–13.

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

Tausen, M., Clausen, M., Moeskjær, S., Shihavuddin, A. S. M., Dahl, A. B., Janss, L., & Andersen, S. U. (2020). Greenotyper: Image-Based Plant Phenotyping Using Distributed Computing and Deep Learning. *Frontiers in Plant Science*, *11*(August), 1–17. https://doi.org/10.3389/fpls.2020.01181

Thapa, S., Zhu, F., Walia, H., Yu, H., & Ge, Y. (2018). A novel LiDAR-Based instrument for high-throughput, 3D measurement of morphological traits in maize and sorghum. *Sensors (Switzerland)*, *18*(4). https://doi.org/10.3390/s18041187

Tilly, N., Aasen, H., & Bareth, G. (2015). Fusion of plant height and vegetation indices for the estimation of barley biomass. *Remote Sensing*, *7*(9), 11449–11480. https://doi.org/10.3390/rs70911449

Tremblay, N., Wang, Z., & Cerovic, Z. G. (2012). Sensing crop nitrogen status with fluorescence indicators. A review. *Agronomy for Sustainable Development*, *32*(2), 451–464. https://doi.org/10.1007/s13593-011-0041-1

US Department of Agriculture National Agricultural Statistics Service. (2022). *Crop Production 2021 Summary*. *January*, 83. https://www.agmrc.org/media/cms/CropProdSu01122017_C27B81FB8F64D.pdf

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., & Hawkesford, M. J. (2017). Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant Biology*, *44*(1), 143–153. https://doi.org/10.1071/FP16163

Walter, A., Liebisch, F., & Hund, A. (2015). Plant phenotyping: From bean weighing to image analysis. *Plant Methods*, *11*(1), 1–11. https://doi.org/10.1186/s13007-015-0056-8

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*, *341*(6145), 508–513. https://doi.org/10.1126/science.1239402

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant*, *13*(2), 187–214. https://doi.org/10.1016/j.molp.2020.01.008

Zhang, C., Si, Y., Lamkey, J., Boydston, R. A., Garland-Campbell, K. A., & Sankaran, S. (2018). High-throughput phenotyping of seed/seedling evaluation using digital image analysis. *Agronomy*, *8*(5), 1–14. https://doi.org/10.3390/agronomy8050063

Zhang, H., Ge, Y., Xie, X., Atefi, A., Wijewardane, N. K., & Thapa, S. (2022). High

throughput analysis of leaf chlorophyll content in sorghum using RGB, hyperspectral, and fluorescence imaging and sensor fusion. *Plant Methods*, *18*(1), 1–17. https://doi.org/10.1186/s13007-022-00892-0

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., Wang, J., & Fan, J. (2019). Crop phenomics: Current status and perspectives. *Frontiers in Plant Science*, *10*(June). https://doi.org/10.3389/fpls.2019.00714

APPENDIX

APPENDIX A. Some glimpses during experimental set up and ground truth data collection.

APPENDIX B. A pairwise comparison between genotypes at each time point for pixel count. (Only genotypes that have significantly different pixel count is included in the table).

Simple effect	**Genotype**	**_Genotype**	Estimate	Standard error	Adj.p	Adj.lower	Adj.upper
time 3	**6**	**88**	0.00379	0.001	0.0296	0.00016	0.00743
time 3	**16**	**23**	0.00448	0.001	0.002	0.00084	0.00812
time 3	**16**	**68**	0.00397	0.001	0.0156	0.00033	0.00761
time 3	**23**	**24**	-0.0038	0.001	0.0301	-0.0074	-0.0002
time 3	**24**	**88**	0.00464	0.001	0.001	0.001	0.00828
time 3	**33**	**88**	0.00402	0.00097	0.0082	0.00049	0.00756
time 3	**34**	**88**	0.00411	0.00103	0.0157	0.00034	0.00788
time 3	**37**	**88**	0.00417	0.001	0.0072	0.00053	0.00781
time 3	**38**	**88**	0.0042	0.001	0.0064	0.00056	0.00784
time 3	**84**	**88**	0.00367	0.001	0.0453	0.00003	0.00731
time 4	**5**	**23**	0.00466	0.001	0.0009	0.00102	0.00829
time 4	**5**	**86**	0.00366	0.001	0.0473	1.7E-05	0.0073
time 4	**5**	**88**	0.00394	0.001	0.0178	0.0003	0.00758
time 4	**6**	**23**	0.00502	0.001	0.0002	0.00138	0.00866
time 4	**6**	**86**	0.00402	0.001	0.0131	0.00038	0.00766
time 4	**6**	**88**	0.0043	0.001	0.0043	0.00066	0.00794
time 4	**15**	**68**	0.00413	0.001	0.0086	0.00049	0.00777

time 4	**15**	**69**	0.00377	0.001	0.0328	0.00013	0.0074
time 4	**15**	**70**	0.00424	0.001	0.0054	0.0006	0.00788
time 4	**15**	**85**	0.00391	0.001	0.0197	0.00027	0.00755
time 4	**15**	**87**	0.00391	0.001	0.0198	0.00027	0.00755
time 4	**15**	**113**	0.00442	0.001	0.0026	0.00078	0.00806
time 4	**16**	**68**	0.00367	0.001	0.0452	0.00003	0.00731
time 4	**16**	**70**	0.00379	0.001	0.0305	0.00015	0.00742
time 4	**16**	**113**	0.00396	0.001	0.0162	0.00032	0.0076
time 4	**23**	**24**	-0.0044	0.001	0.0028	-0.008	-0.0008
time 4	**23**	**34**	-0.0042	0.00103	0.0115	-0.008	-0.0004
time 4	**23**	**37**	-0.0043	0.001	0.0039	-0.008	-0.0007
time 4	**23**	**52**	-0.005	0.001	0.0002	-0.0086	-0.0013
time 4	**23**	**67**	-0.0041	0.001	0.0107	-0.0077	-0.0004
time 4	**24**	**88**	0.00368	0.001	0.0432	4.4E-05	0.00732
time 4	**37**	**88**	0.0036	0.001	0.0564	-4E-05	0.00724
time 4	**38**	**86**	0.00487	0.001	0.0003	0.00123	0.00851
time 4	**38**	**113**	0.00366	0.001	0.0473	1.7E-05	0.0073
time 4	**51**	**68**	0.00421	0.001	0.0062	0.00057	0.00785
time 4	**51**	**69**	0.00385	0.001	0.0244	0.00021	0.00749
time 4	**51**	**70**	0.00433	0.001	0.0038	0.00069	0.00796
time 4	**51**	**85**	0.00399	0.001	0.0145	0.00035	0.00763
time 4	**51**	**87**	0.00399	0.001	0.0146	0.00035	0.00763
time 4	**51**	**113**	0.0045	0.001	0.0018	0.00086	0.00814

time 4	**52**	**86**	0.00396	0.001	0.0162	0.00032	0.0076
time 4	**52**	**88**	0.00424	0.001	0.0054	0.0006	0.00788
time 4	**68**	**84**	-0.0039	0.001	0.0235	-0.0075	-0.0002
time 4	**70**	**84**	-0.004	0.001	0.0154	-0.0076	-0.0003
time 4	**83**	**86**	0.0043	0.001	0.0042	0.00066	0.00794
time 4	**83**	**88**	0.00458	0.001	0.0013	0.00094	0.00822
time 4	**84**	**85**	0.00364	0.001	0.0497	1.95E-06	0.00728
time 4	**84**	**87**	0.00364	0.001	0.05	2.13E-07	0.00728
time 4	**84**	**113**	0.00415	0.001	0.0078	0.00051	0.00779

APPENDIX C. A pairwise comparison between genotypes for yield difference. (Only genotypes that have significantly different yield is included in the table).

Genotypes	**_Genotypes**	Estimates	Standard error	Adj.p	Adj.upper	Adj.lower
5	**51**	16.7125	3.6797	0.0025	3.0917	30.3333
5	**52**	19.0125	3.6797	0.0002	5.3917	32.6333
5	**84**	15.1375	3.6797	0.0126	1.5167	28.7583
6	**52**	14.7375	3.6797	0.0185	1.1167	28.3583
15	**51**	18.875	3.6797	0.0002	5.2542	32.4958
15	**84**	17.3	3.6797	0.0013	3.6792	30.9208
24	**51**	14.2375	3.6797	0.0293	0.6167	27.8583
24	**52**	16.5375	3.6797	0.003	2.9167	30.1583
33	**52**	14.1674	3.5773	0.0216	0.9255	27.4094
37	**51**	17.9375	3.6797	0.0006	4.3167	31.5583
37	**84**	16.3625	3.6797	0.0036	2.7417	29.9833
38	**51**	16.025	3.6797	0.0052	2.4042	29.6458
38	**52**	18.325	3.6797	0.0004	4.7042	31.9458
38	**84**	14.45	3.6797	0.0241	0.8292	28.0708
51	**68**	-14.975	3.6797	0.0148	-28.596	-1.3542
51	**69**	-16.113	3.6797	0.0047	-29.733	-2.4917
51	**70**	-16.613	3.6797	0.0028	-30.233	-2.9917
51	**85**	-13.663	3.6797	0.0483	-27.283	-0.0417
51	**87**	-16.488	3.6797	0.0032	-30.108	-2.8667

51	**113**	-17.1	3.6797	0.0016	-30.721	-3.4792
51	**114**	-16.138	3.6797	0.0016	-29.758	-2.5167
52	**67**	-15.85	3.6797	0.0016	-29.471	-2.2292
52	**68**	-17.275	3.6797	0.0016	-30.896	-3.6542
52	**69**	-18.413	3.6797	0.0016	-32.033	-4.7917
52	**70**	-18.913	3.6797	0.0016	-32.533	-5.2917
52	**85**	-15.963	3.6797	0.0016	-29.583	-2.3417
52	**87**	-18.788	3.6797	0.0016	-32.408	-5.1667
52	**88**	-13.788	3.6797	0.0016	-27.408	-0.1667
52	**113**	-19.4	3.6797	0.0016	-33.021	-5.7792
52	**114**	-18.438	3.6797	0.0016	-32.058	-4.8167
69	**84**	14.5375	3.6797	0.0016	0.9167	28.1583
70	**84**	15.0375	3.6797	0.0016	1.4167	28.6583
84	**87**	-14.913	3.6797	0.0016	-28.533	-1.2917
84	**113**	-15.525	3.6797	0.0016	-29.146	-1.9042
84	**114**	-14.563	3.6797	0.0016	-28.183	-0.9417