4-30-2021

Multispectral in-field sensors observations to estimate corn leaf nitrogen concentration and grain yield using machine learning

Razieh Barzin
razieh.barzin@gmail.com

Multispectral in-field sensors observations to estimate corn leaf nitrogen concentration and grain yield using machine learning

By

Razieh Barzin

Approved by:

S. D. Filip To (Major Professor)
Ganesh C. Bora
Jac Varco
Padmanava Dash
Steven H. Elder (Graduate Coordinator)
Scott Willard (Dean, College of Agriculture and Life Sciences)

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Biological Engineering
in the Department of Agricultural and Biological Engineering

Mississippi State, Mississippi

April 2021

Name: Razieh Barzin

Date of Degree: April 30, 2021

Institution: Mississippi State University

Major Field: Biological Engineering

Select Appropriate Title: S. D. Filip To

Title of Study: Multispectral in-field sensors observations to estimate corn leaf nitrogen concentration and grain yield using machine learning

Pages in Study: 92

Candidate for Degree of Doctor of Philosophy

Nitrogen (N) is the most critical fertilizer applied nutrient for supporting plant growth. It is a critical part of photosynthesis as a component of chlorophyl, hence it is a key indicator of plant health. In recent years, rapid development of multispectral sensing technology and machine learning (ML) methods make it possible to estimate leaf chemical components such as N for predicting yield spatially and temporally. The objectives of this study were to compare the relationships between canopy reflectance and corn (*Zea mays* L.) leaf N concentration acquired by two multispectral sensors: red-edge multispectral camera mounted on the Unmanned Aerial Vehicle (UAV) and crop circle ACS-430. Four fertilizer N rates were applied, ranging from deficient to excessivein order to have a broad rangein plant N status. Spectral information was collected at different phenological stages of corn to calculate vegetation indices (VIs) for each stage. Moreover, leaf samples were taken simultaneously to determine N concentration. Different ML methods (Multi-Layer Perceptron (MLP), Support Vector Machines (SVMs), Random Forest regression, Regularized regression models, and Gradient Boosting) were used to estimate leaf N% from VIs and predict yield from VIs. Random Forest Regression was utilized as a feature selection method to choose the best combination of variables for different stages and to interpret the relationships between VIs and corn leaf N concentration and grain yield. The Canopy Chlorophyll Content Index (SCCCI) and Red-edge Ratio Vegetation Index (RERVI) were selected as the most efficient VIs in leaf N estimation and SCCCI, Red-edge chlorophyll index ($\mathrm{CI}_{\mathrm{RE}}$), RERVI, Soil Adjusted Vegetation Index (SAVI), and Normalized Difference Vegetation Index (NDVI) were chosen as the most effective VIs in predicting corn grain yield. The results derived from using a red-edge multispectral camera showed that the SCCCI was the most proper index for predicting yield at most of the phenological stages and Gradient Boosting was the best-fitted model to estimate leaf N% with an 80% coefficient of determination. Using a Crop Circle ACS-430 showed that the Support Vector Regression (SVR) model achieved the best performance measures than other models tested in the prediction of leaf N concentration.

ACKNOWLEDGEMENTS

The author would like to express her deepest appreciation to Dr. Bora for providing the opportunity to learn and work under his guidance. Gratitude is extended to the supporting Dr. To to accept to continue guiding this project and also other committee members including Dr. Varco and Dr. Dash. Thank you all for your individual time and efforts invested in my education both inside and out of the classroom. I would also like to acknowledge my husband, family, and friends who encouraged me in my walk with the Lord and my pursuit of higher education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The United States Department of Agriculture (USDA) reported that approximately 36 million hectares of corn were planted and 11.3 ton/ha grain yield was harvested in the US in 2019 (USDA/NASS, 2019). Agricultural lands expanded about 10 million ha every year from 1980 to 2007 (West et al., 2010) to meet the needs of a growing population, changing diets, and increasing biofuel requirement. Harvested land under corn has expanded from 27 million to 35 million ha in the US from 1936 to 2016 (Kakkar, 2017). Therefore, corn has a remarkable impact on feeding the world’s population.

Nitrogen (N) is one of the essential nutrients required for plant growth, development, and reproduction. It is a critical input in boosting yields and economic return to agriculturalists and is often considered as the limiting nutrient for crop production (Bender et al., 2012). Nitrogen deficiency causes stunted growth, low chlorophyll contents, and yellowing of older leaves in plants. The readily available N in soils is mainly in the inorganic forms of ammonium and nitrate. Nitrogen is abundant in the atmosphere as a format of N_2 gas, but this form is not easily available to plants. A mismatch between applied N and crop N demand can potentially prevent crop growth or damage the environment when N fertilizer is more or less used, respectively. Both situations may lead to a decrease in N use efficiency (NUE) resulting in economical production losses and possibly environemental hazards. Excess N gets into the groundwater and contaminates it as a result of NO_3^--N leaching (West et al., 2010). Although worldwide usage of synthetic fertilizer N is increasing, NUE is around 50% for corn (Siqueira, 2016).

One of the main aims of agricultural production is attaining the highest crop yield at the lowest cost. Early detection and managing the problems associated with crop yield indicators can improve yield providing econimc and environemental benefits. Visible (blue, green, red) and thermal wavelengths (near-infrared (NIR) and red-edge) have been utilized successfully to monitor spatial variability of crop health, crop cover, soil moisture, N stress, and crop yield (Baez-gonzalez et al., 2005; Báez-González et al., 2002; Barzin et al., 2020; Doraiswamy et al., 2003; Lillesand et al., 2015; Lobell et al., 2005; Magri et al., 2005; Pathak et al., 2018; Sun, 2000; P. Yang et al., 2004). In recent years, aerial imagery using drones has been broadly utilized for crop yield prediction before harvest (Barzin et al., 2020; GopalaPillai and Tian, 1999; Senay et al., 1998). These spectral images can provide high spatial resolution and cloud-free information on the crop's characteristics and it allows continual analysis of crop vegetation conditions during the growing season. In the past, normalized difference vegetation index (NDVI) analysis has been widely used for analyzing plant growth to support precision farming (Báez-González et al., 2002; Butchee et al., 2011; Funk and Budde, 2009; GopalaPillai and Tian, 1999; Matsushita et al., 2007; Reyniers et al., 2006; Santamar, 2017; Sharma et al., 2015; C. Yang and Anderson, 2000). Other research has shown strong relationships between other spectral bands and yield (Aquino et al., 2018; Barzin et al., 2020; Ferencz et al., 2010; Fox, 2015; He et al., 2018; Johnson, 2014; Shi et al., 2013; Sumner, 2019; Tadesse et al., 2015). Senay et al., 1998 found a 0.99 coefficient of determination (R^2) between corn yield values and NIR wavelength of an aerial image under controlled conditions (Senay et al., 1998). Sumner, 2019 reported that green or red-edge bands have a strong relationship, however, the combined index such as Normalized Difference Red-edge (NDRE) that incorporates the red-edge bands outperformed the other indices (Sumner, 2019).

Various in-field sensors utilized to collect the spectral bands for plant reflectance can be used to calculate multiplevaried VIs. A common commercially availabel sensor such as the handheld GreenSeeker from the Trimble company which measures the red and NIR bands and provides the NDVI. Crop Circle ACS-211 from Holland Scientific company is another sensor that is like GreenSeeker and collects the red and NIR wavelength and computes the NDVI. Another sensor used in this research is Crop Circle ACS-430 measures three specific bands (red, red-edge, and NIR) and prepares the NDVI and NDRE. Since Crop Circle ACS-430 records the wavelengths and geolocation of each measured point on an SD flashcard, the other VIs related to these three bands can be calculated as well. Vehicle-mounted sensors can also be used to collect plant canopy information over large study areas or agricultural landscapes rapidly. One of the advantages of this sensor in comparison to other active sensors on the market is that it can measure spectral reflectance independent of its height above a target. The other sensor utilized in this research is the RedEdge™ multispectral band sensor from the MicaSense® company which can be mounted on Unmanned Aerial Vehicles (UAVs) and collect spectral bands from different heights. One of the advantages of these in-field sensors in comparison to satellite imagery is that these collections can be scheduled for a different time during the growing season considering the weather conditions.

This study used remotely sensed canopy reflectance data collected from a RedEdge™ multispectral band sensor mounted on a UAV and a handheld Crop Circle ACS-430.

The objectives of this study were to:

1- Develop a yield prediction model at specific corn growth stages using spectral data and VIs,

2- Develop and compare ML-based models to estimate the leaf N content of corn using different spectral bands and VIs acquired from a red-edge multispectral band sensor mounted on a UAV, and

3- Estimate leaf N content and predict corn yield using the multispectral handheld sensor's observations (crop circle ACS-430 field sensor) and develop and compare ML algorithms to find the best prediction model.

CHAPTER II

USE OF UAS TO TAKE MULTISPECTRAL IMAGERY AT DIFFERENT PHYSIOLOGICAL STAGES FOR YIELD PREDICTION AND INPUT RESOURCE OPTIMIZATION IN CORN

Introduction

The objectives of this study were to track five different spectral bands obtained through sensors mounted on a UAS at five different phenological stages of corn and evaluate 26 calculated VIs at each specific stage of growth. Feature selection approaches were applied to reduce the number of predictors. Consequently, relationships between the spectral bands and 26 VIs (as predictors), and corn yield (as a response variable) were investigated to determine more correlated covariates with the response variables. Finally, machine-learning techniques were used to develop models for corn yield prediction at each phenological stage. Estimation of corn yield during the crop-growing season is essential for efficient management at strategic phenological stages. Agricultural surveys and field sampling of standing crops are useful and reliable approaches to estimate corn production. However, the spatiotemporal variability of biophysical characteristics of the crops due to inconsistency in soil nutrients and water availability, as well as other environmental parameters affecting plant growth present challenges in estimating yield accurately on a large spatial scale. The Normalized Difference Vegetation Index (NDVI) can be used to quantify biomass production (Santamar, 2017) by measuring the difference between near-infrared (NIR) and red wavelengths and is widely used in agricultural crop studies (Bronson et al., 2005; J. M. Chen, 1996; Hogrefe et al., 2017). In this study, vegetation indices (VIs) including Normalized

Difference Red-Edge (NDRE), Optimized Soil Adjusted Vegetation Index (OSAVI), Simplified Canopy Chlorophyll Content Index (SCCCI), and Visible Atmospherically Resistant Index ($VARI_{green}$), were used to determine their importance in predicting corn grain yield.

Remote sensing technologies have been used across a wide range of application in agriculture to detect and monitor the biophysical characteristics of plants. The spectral information collected in pixel scale is used to compute VIs, which are algorithms derived from the spectral transformation of reflectance at two or more specified wavelengths and are used to evaluate vegetative cover or biomass and plant growth or health status. Differencing, Rationing, Rationing Sums and Differences, and Linear Combinations of different spectral wavelengths are standard methods used to calculate different VIs. One of the advantages of using a remotely sensed VI products is that they are computed in a uniform manner and comparable during time and location(Jackson and Huete, 1991).

Unmanned aerial systems (UAS) equipped with the RedEdge™ multispectral camera can be used to detect spatial and temporal variability in biophysical characteristics of corn, such as spectral reflectance for the specified wavelengths, which can be used to compute multiple VIs. Satellite imagery is routinely used to estimate the yield of different crops (Báez-González et al., 2002; Shiu and Chuang, 2019; Silvestro et al., 2017). Ongoing and past examinations show that the red-edge waveband is useful for estimating the chlorophyll content and N status of plants. NDVI-Red-edge is increasingly profitable and helpful for later stages when contrasted with the early V6 stage for in-season N application (Sharma et al., 2015).

A UAS can acquire data from low altitude, where interference by clouds is not an obstacle between the sensor and land surface (Bondi et al., 2016; Zhang and Kovacs, 2012). Although shadows created by a UAS can still be an issue, data from a UAS is more readily available as compared to satellite imaging because flights can be scheduled at key periods of designated phenological stages considering weather conditions, and they can provide greater spatial resolution. From an altitude of 60 m or less, a camera mounted on a UAS can collect more detailed and important local landscape information at around 4 cm spatial resolution. The spatial resolution can be improved by using a more accurate camera sensor or decreasing the UAS flight altitude (Iizuka et al., 2018). Therefore, a UAS can provide images with greater pixel resolution, and it is possible to acquire spectral images as far as required for research objectives more frequently.

The emergence of new statistical learning models such as Ensemble methods based on decision trees can estimate yield before harvesting. The decision tree approach is increasingly being used for different purposes such as corn optimal fertilizer estimation (Qin et al., 2018) and corn yield estimation (Khanal et al., 2018). Gradient boosting machines (GBMs) are ensemble learning models to empower the weaker models such as decision trees by combining the results from them. GBMs are widely used in a broad range of practical applications (Natekin and Knoll, 2013) and have demonstrated remarkable success for regression and classification applications.

Materials and Methods

The study was undertaken on an experimental plot at Mississippi State University.

Geographical Area of Study

The geographic area of the study was at the W.B. Andrews Agriculture Systems Research Farm at Mississippi State, MS, USA (33°28′13.5″N, 88°45′48.0″W) (Figure 2.1). The total area of the field was 0.8 ha mapped as a Marietta fine sandy loam (fine-loamy, mixed, siliceous, thermic, Aquic Fluventic Eutrochept). The imagery data were collected during the corn growing season of years 2017, 2018, and 2019.

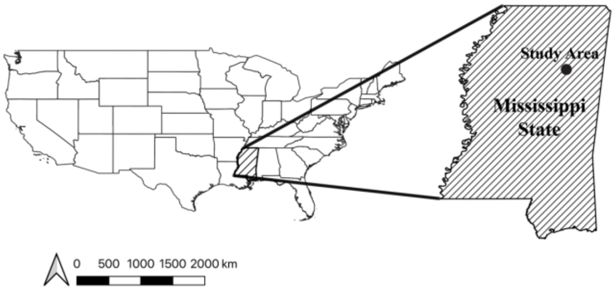


Figure 2.1 Geolocation of the study area monitored during the growing season.

Growing season precipitation totals measured at the experimental field varying between 58 cm in 2017, 42 cm in 2018, and 76 cm in 2019, as shown in Figure 2.2. The data were retrieved from the Mississippi Delta weather information of the Delta Agricultural Weather Center at the Delta Research and Extension Center, which is located at a distance of 1 km from the research field. The precipitation in 2019 was the greatest of the three years, whereas 2018 was the lowest year. Because of the low precipitation, signs of water stress were observed in the plants. The water deficiency issue was addressed through furrow-irrigation in early June 2018. The mean temperature was almost similar throughout all three growing seasons, which was 23 degrees Celsius.

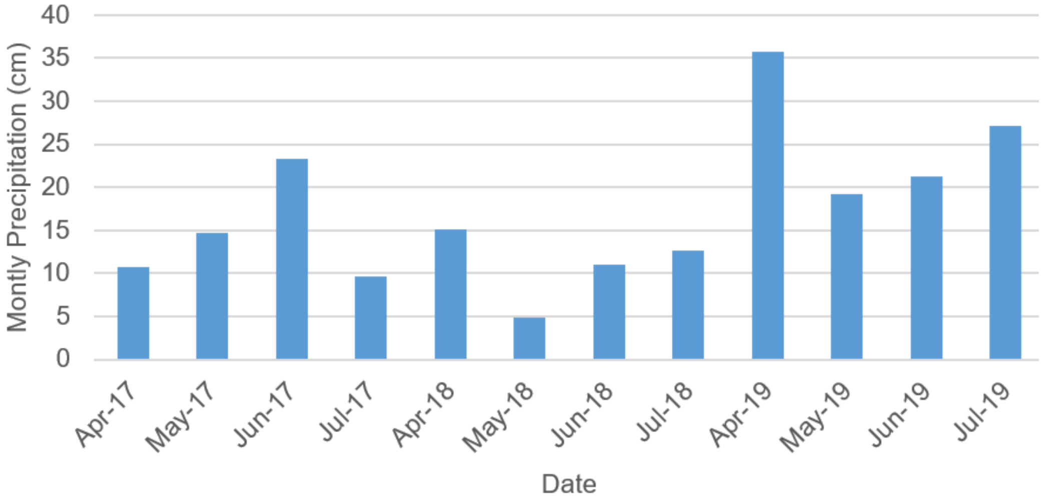


Figure 2.2 Monthly precipitation chart for the growing seasons: 2017, 2018, and 2019

Experimental Design

The experimental field was divided into 16 plots, including 12 rows of corn, which were planted at a row spacing of 97 cm, and plot length was 38 m, and there was a 3 m alleyway in between each plot. The experimental design was a randomized complete block. Corn (DeKalb Brand-DKC67-72 variety) was planted on 13 April 2017, 19 April 2018, and 23 April 2019. There were four treatments of N, including 0, 90, 180, and 270 kg/ha, applied randomly with four replicates. Figure 2.3. illustrates the spatial distribution of each treatment and associated replication in the study field. The goal of the spatially varied N application was to identify the optimal N requirements for the corn crop and to address the spatial variability of the soil. Treatments were randomly assigned to the plots and have been repeated each year on the same experimental units.

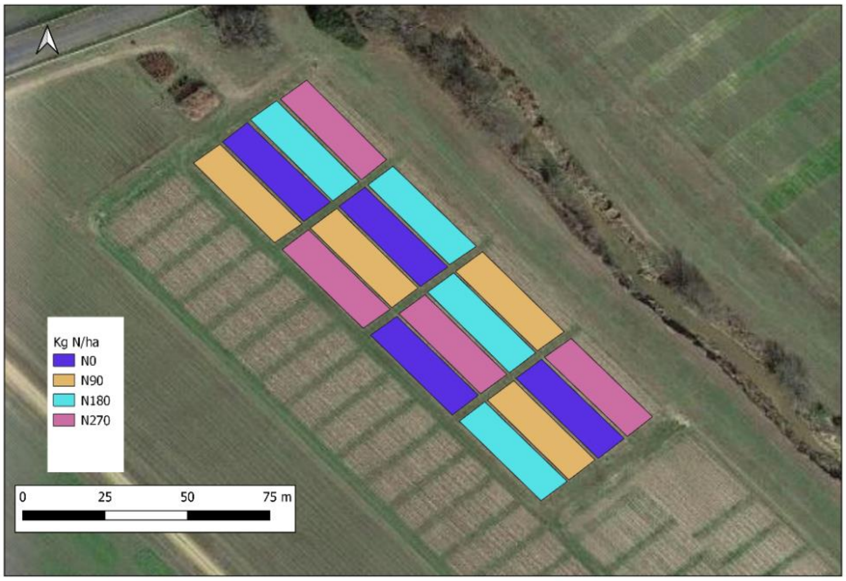


Figure 2.3 Nitrogen treatments and four replicates in 2017-2019 at the Research Farm, Mississippi State, US.

The first N application was made just after emergence each year at V2-3 (2–3 leaves with visible leaf collar), followed by a second application at V6-7 (6–7 leaves with visible leaf collar). Figure 2.4 illustrates fertilizer N applications, planting/harvesting, and flight dates during the different phenological stages of corn for 2017, 2018, and 2019. Fertilizer N was side dressed as liquid urea ammonium nitrate (UAN) (32-0-0) with an applicator equipped with coulters, and liquid knives spaced 23 cm from one side of each corn row and 7.62 cm deep. Limited furrow irrigation (50.8 mm) in 2018 was supplied to the experimental area in early June because of the low rainfall received with resulting visible signs of water stress.. Strip tillage was utilized for these years, although plots were initially disked, and beds were formed following the 2017 growing season. Following the 2017 corn harvest, a Persian clover (Trifolium resupinatum L.) cover crop was planted at a percent live seeding rate of 6.74 kg seed ha-1 across the whole experimental area using a no-till grain drill to provide winter cover and possible soil improvements. Plots were fertilized based on soil test results and received uniform applications of P-K-Mg-S before planting for all parcels. The fertilizer blend consisted of two parts muriate of potash (0-0-60), one part concentrated super phosphate (0-46-0), and one part sulfate of potash-magnesia (0-0-22-11Mg-22S) and was applied at a material rate of 224 kg ha–1. Weeds and pests were controlled based on Mississippi State University Extension recommendations. The field under study has been used for corn cultivation since 2012 with the same fertilize N rates applied each year. Corn grain was harvested using a two-row plot combine, which collected the yield from rows 2 and 3 and rows 10 and 11 (Figure 2.5).

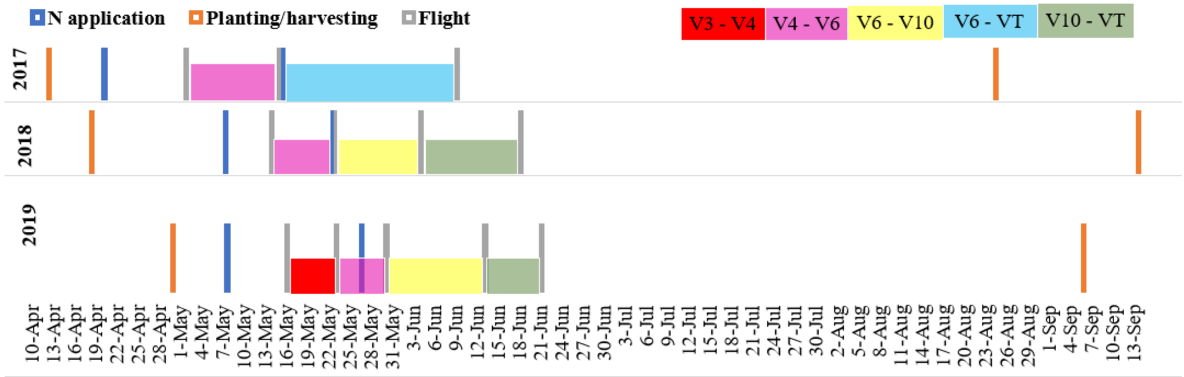


Figure 2.4 N applications, planting/harvesting, and flight dates during the different phenological stages of corn for 2017, 2018, and 2019.

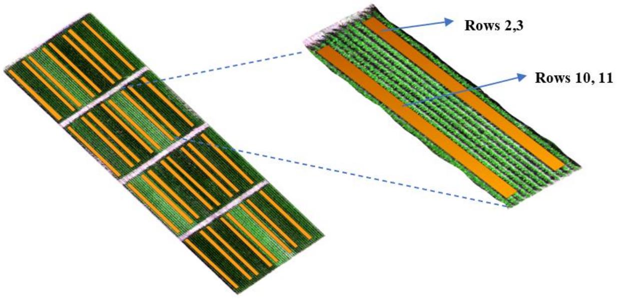


Figure 2.5 The spatial location of rows 2, 3, 10, and 11

Data Collection

The MicaSense RedEdge™ multispectral band sensor mounted on a UAS was used to capture images in five different spectral bands simultaneously. The unit weight was 150 grams with a dimension of 12.1 cm × 6.6 cm × 4.6 cm. The UAS was flown at an altitude of 60 m in 2017 and 2018, whereas in 2019 it was flown at an altitude of 30 m. Decreasing the altitude from 60 m to 30 m provided better images with approximately four times greater resolution. The enhanced resolution was beneficial in the separation of soil and vegetation. The sensor was mounted on the bottom of the UAS with a viewing angle not exceeding 10 degrees from nadir. The multispectral sensor measured the wavelength at five different spectral bands, including blue (475 nm center, 32 nm bandwidth), green (560 nm center, 27 nm bandwidth), red (668 nm center, 16 nm bandwidth), red-edge (717 nm center, 12 nm bandwidth), and near-infrared (842 nm center, 57 nm bandwidth). All five bands were collected simultaneously at a rate of one capture per second. Optimal image acquisition time is within plus or minus two and a half hours of local solar noon (Bronson et al., 2005; Erdle et al., 2011; Gitelson, 2004; Gitelson et al., 2002; Shanahan et al., 2001), therefore, all the flights were performed around 10:30 am under cloud-free conditions. The length of the flight was around 10 minutes for the 0.8 ha field area; therefore, environmental conditions such as solar radiation, temperature, and humidity were nearly constant during the data acquisition process. The UAS images were acquired with a horizontal overlap of at least 75%. Images were stitched and mosaicked with the Pix4D mapper software (Pix4D SA, Lausanne, Switzerland) to obtain unique and compiled images for the study area. Superimposed images obtained through the stacking of images were disoriented during the first flight. This may be due to the errant movement of the camera. To address this issue a co-registration process was adopted. A calibrated reflectance panel (CRP) was used for the radiometric calibration of the acquired images. The CRP offers calibration information associated with the acquired images across the visible and near-infrared images. Images of the CRP that had been taken before and after the flight were used to convert raw pixel values into reflectance. The initial processing of the raw images was done at the Geosystems Research Institute (GRI) at Mississippi State University.

Vegetation Indices

Vegetation indices are mathematical combinations of wavelength-specific spectral reflectance developed to detect and monitor vegetation's phenological conditions remotely. For vegetation, reflectance by itself is low in both the blue and red bands of the spectrum due to maximum chlorophyll absorption in those bands, while reflectance has a peak in the green band. Because of the cellular structures of leaves, the reflectance is much more significant in the NIR bands compared to visible bands. In this study, several VIs were derived from a 5-band multispectral sensor. Multispectral bands are visually and numerically similar; on the other hand, they are often highly correlated (Barzin et al., 2017a; González-audícana et al., 2004). To avoid the issue associated with VI calculation, row data was converted to percentage reflectance to signify the quantitative data. The name of the indices and associated spectral bands are listed in Table 2.2. The choices of indices by the researchers may vary according to their need but for biomass content, most indices involving red, infrared, and red-edge bands were preferred. These bands are supposed to explain even the subtle changes in biomass content.

Table 2.2 Mathematical representation of vegetation indices and ratios calculated from spectral reflectance.

	Vegetation Indices (VI)	Name	Formula	Study Groups (Reference)
1	DVI	Difference Vegetation Index	NIR-Red	(Tucker, 1979)
2	GDVI	Green Difference Vegetation Index	NIR-Green	(Wu, 2014)
3	RDVI	Renormalized Difference Vegetation Index	(NIR-Red)/$\sqrt{NIR + Red}$	(Roujean and Breon, 1995)
4	TDVI	Transformed Difference Vegetation Index	1.5(NIR-Red)/$\sqrt{\mathrm{NIR}^2 + \mathrm{Red} + 0.5}$	(Bannari et al., 2002)
5	NDVI	Normalized Difference Vegetation Index	(NIR-Red)/(NIR+Red)	(Pathak et al., 2018; Rouse et al., 1974)
6	GNDVI	Green Normalized Difference Vegetation Index	(NIR-Green)/(NIR+Green)	(Gitelson and Merzlyak, 1998a)
7	NDRE	Normalized Difference Red-edge	(NIR-Red-edge) / (NIR+Red-edge)	(Gitelson and Merzlyak, 1994; T. B. Raper and Varco, 2015)
8	SCCCI	Simplified Canopy Chlorophyll Content Index	NDRE / NDVI	(T. B. Raper and Varco, 2015)
9	EVI	Enhanced Vegetation Index	2.5*(NIR-Red)/ (NIR+6Red-7.5Blue+1)	(Matsushita et al., 2007)
10	TVI	Triangular Vegetation Index	0.5 [120 (NIR-Green)] -200 (Red – Green)	(Broge and Leblanc, 2001)
11	VARIgreen	Visible Atmospherically Resistant Index	(Green-Red)/(Green + Red-Blue)	(Gitelson et al., 2002)

Table 2.2 (continued)

	Vegetation Indices (VI)	Name	Formula	Study Groups (Reference)
12	TGI	Triangular Greenness Index	(Red-Blue) (Red-Green) - (Red - Green) (Red - Blue))/2	(Hunt et al., 2011)
13	NLI	Non-Linear Index	(NIR^2-Red)/(NIR^2+Red)	(Vescovo and Gianelle, 2008)
14	MNLI	Modified Non-Linear Index	(NIR^2-Red) *(1+0.5)/ (NIR^2+Red+0.5)	(Feng et al., 2019; Gong et al., 2003)
15	SAVI	Soil-Adjusted Vegetation Index	1.5*(NIR-Red))/(NIR+Red+0.5)	(Rondeaux et al., 1996)
16	GSAVI	Green Soil-Adjusted Vegetation Index	1.5 * (NIR-Green)/(NIR+Green+0.5)	(Sripada et al., 2008)
17	OSAVI	Optimized Soil-Adjusted Vegetation Index	(NIR-Red)/(NIR+Red+0.16)	(Rondeaux et al., 1996)
18	GOSAVI	Green Optimized Soil-Adjusted Vegetation Index	(NIR-Green)/(NIR+Green+0.16)	(Sripada et al., 2008)
19	MSAVI2	Modified Soil-Adjusted Vegetation Index 2	(2NIR+1-$\sqrt{(2\text{NIR}+1)^2-8(\text{NIR}-\text{Red})}$)/2	(Qi et al., 1994)
20	MSR	Modified Simple Ratio	(NIR/Red)-1/$\sqrt{\text{NIR}/\text{Red}}$ +1	(J. M. Chen, 1996)
21	GRVI	Green Ratio Vegetation Index	NIR / Green	(Tucker, 1979)
22	WDRVI	Wide Dynamic Range Vegetation Index	(0.1 NIR-Red) / (0.1 NIR + red)	(Gitelson, 2004)
23	SR	Simple Ratio	NIR/Red	(Fraser and Latifovic, 2005)
24	GARI	Green Atmospherically Resistant Index	NIR-Green - (1.7 (Blue-Red))/(NIR+Green-(1.7 (Blue-Red))	(Gitelson et al., 1996)
25	GCI	Green Chlorophyll Index	(NIR/Green) - 1	(Gitelson et al., 2003)
26	GLI	Green Leaf Index	(Green-Red-Blue)/(2Green+Red+Blue)	(Louhaichi et al., 2001)

Masking Soil Pixels

To estimate the spatial average of VIs for each corn row, it was essential to mask bare soil pixels located between corn rows. After the VIs calculation, the bare soil pixels were removed since these pixels do not provide further information in the yield estimation modeling. Eliminating these pixels reduced the image processing time and attributed to better estimate the spatial average of VIs for each row. Moreover, reflectance data from the corn rows contain information associated with the corn leaves and the scattered wavelength from the background soil within the leaves. The background soil reflectance potentially decreases the effectiveness of the leaves in VI values (Morales et al., 2019). The occurrence of such a phenomenon is explicitly noticed when the leaves are in the primary phenological stages. To reduce this effect, different VIs such as the Soil-Adjusted Vegetation Index (SAVI), Optimized Soil-Adjusted Vegetation Index (OSAVI), Green Soil-Adjusted Vegetation Index (GSAVI), Green Optimized Soil-Adjusted Vegetation Index (GOSAVI), and Modified Soil-Adjusted Vegetation Index 2 (MSAVI2) have been used. These VI's takes care of the contribution of the soil reflectance in VIs calculation, specifically in the leaf edge pixels which may have soil and vegetation information together, therefore, the pure soil pixels were removed. As a result, an empirical equation (Equation 1.1) was used to mask the unshaded and shaded bare soil pixels.

$$G_{index} = 2 * Green - Red - Blue \qquad (2.1)$$

The G_{index} values greater than 0.06 were selected as vegetation pixels based on trial and error. Although NDVI has been used to remove soil pixels (Gallo et al., 2018; Sader and Winne, 1992), it does not detect and remove shaded pixels. Therefore, the proposed G_{index} filter can remove all shaded and unshaded bare soil pixels precisely.

Harvesting Process

The Corn grain was harvested by a two-row plot combine for the whole plot length. Rows 2 and 3 and rows 10 and 11 of each plot were combined. Some of the plots suffered extensive raccoon damage in 2018; hence, ears were harvested by hand from uniformly standing undamaged rows. Hand harvesting was performed by pulling ears from two 6.1-m row lengths of each harvest row pairs (rows 2, 3, and rows 10, 11). The regions of damage were skipped during the hand harvesting. All grain yield data were adjusted to 15.5% moisture content. Since yield data were collected for rows 2–3 and rows 10–11, VIs derived from pixels reciprocating the rows were calculated. The bare soil pixels between rows were eliminated by applying Equation 1.1, and then the spatial average of pixels was taken for each row and applied as variables in company with yield for each row. Figure 2.5 shows the spatial location of the rows within each treatment.

Outlier Detection

After calculating VIs for each phenological stage, outliers for each VI and at each stage were removed from the data set by utilizing a well-known z-score ($z = (x - \overline{x})/\sigma x$) (Schubert and Kriegel, 2014; Torres et al., 2011). Here, the observations with z-score greater than 2.5 were considered as outlier data. The threshold number flexible between 2.5 and 3 were used to remove outliers (Nurunnabi et al., 2015). A smaller threshold number of results in a greater selection of outliers. All in all, approximately 3–8% of the data for each growth stage were removed as outliers.

Feature Selection

The process of identifying the most important features is called “feature selection”. The random forest method is one of the most popular machine-learning methods used in data science workflows. This method is a combination of tree predictors used commonly as a tool for classification, regression, and ranking of candidate predictors (Breiman, 2001; Janitza et al., 2016). In this research, the most important variables for each phenological stage were identified by the Random Forest feature selection method. The Random Forest method uses a training dataset and creates multiple subsets of the data randomly. Then, trees (samples) are used to create a ranking of classifiers and perform a vote for each predicted result. Finally, prediction results are selected which have the most votes (Breiman, 2001). Random Forest is considered a highly accurate and robust method (Janitza et al., 2016) because of the number of decision trees participating in the process. This method has acceptable predictive performance, low overfitting, and simple interpretability.

Statistical Analysis

In this research, density plot and Shapiro–Wilk's test (Akbarzadeh Baghban et al., 2013; Kox et al., 2016; Razali and Wah, 2011; Vellidis et al., 2013) were used to evaluate whether the data follow Gaussian distribution or not. Although some statisticians suggest that in case of the large sample size (n > 30), we can ignore the distribution of the data and use parametric tests (Razali and Wah, 2011)(Akbarzadeh Baghban et al., 2013; Ghasemi and Zahediasl, 2012; Vellidis et al., 2013), the observed yield data were not large enough (32 samples size for each year) and were not following the normal distribution; therefore, two approaches were used to make yield prediction models: 1) the data were normalized and then multiple regression models were fitted using the important features, and 2) a gradient boosting decision tree model was hired as a non-parametric method to estimate corn yield. Gradient boosting machines (GBMs) is a method of converting weak learners into strong learners like the Random Forest, however, in GBM the kth tree is trained from the first k-1 trees and updated the residual for the ith example of the difference between prediction and observations (Friedman, 2002). In other words, the predictors were sequentially trained and tried to correct the predecessors. One of the advantages of using GBM is that this method is highly customizable to the specific necessity of the application, such as being learned with regard to various loss functions (Natekin and Knoll, 2013).

For the multiple regression models, some of the input variables were not associated with the response variables that triggered excessive complexity in the final model. Therefore, possible combinations of the essential variables were used to fit different multiple regression models. Linear model selection was used to determine the number of significant variables that improve the model by maximizing the adjusted R^2, minimizing Bayesian information criterion (BIC), and minimizing the cross-validated prediction error (Cp) (Hastie et al., 2004). Furthermore, cross-validation (CV) (Kutner et al., 2005) was used as a backup method to ensure the predictors were correctly determined. Cross-validation is a method used in the selection of models to test the ability of different models in their accuracy in the prediction of results. In this research, the data were split into two subsets. Eighty percent of the data were used as training samples or the model-building set, and 20% of the data were used for prediction or as a validation set. Each variable was included in the model, and then the average of the cross-validation error was estimated. Overall, after removing outliers, the Random Forest and cross-validation methods were used to find the number of influential variables for predicting the corn yield. Random Forest feature selection illustrates the importance of variables at each stage. Different variables were selected for each phenological stage of corn.

In this research, the software programs ArcMap 10.7.1 (Environmental Systems Research

Institute, Inc. (ESRI), Redlands, CA, USA) (Esri, Redlands, Ca, 2019), QGIS v.3.12.0 (Böschacherstrasse 10a CH-8624 Grüt (Gossau ZH), Zurich, Switzerland) ("QGIS Development Team," 2020), and R version 3.6 (R Core Team: Vienna, Austria) (R Core Team: Vienna, 2019)were used to manipulate and analyze the data.

Results and Discussion

Graphical (density plot) and numerical (Shapiro–Wilk's test) assessment of the normality of the data illustrated that the data does not follow a normal distribution (Figure 2.6). It can be observed that the corn yield data distribution shape does not match the normal distribution (dashed lines). Since the normality test is sensitive to sample size, therefore, it is important to combine visual inspection and significance tests in order to make the right decision. The Shapiro–Wilk's test confirmed the same result and therefore, the data was normalized. The correlation between yield and 31 independent variables is shown in Figure 2.7 at V4-5 and VT stages.

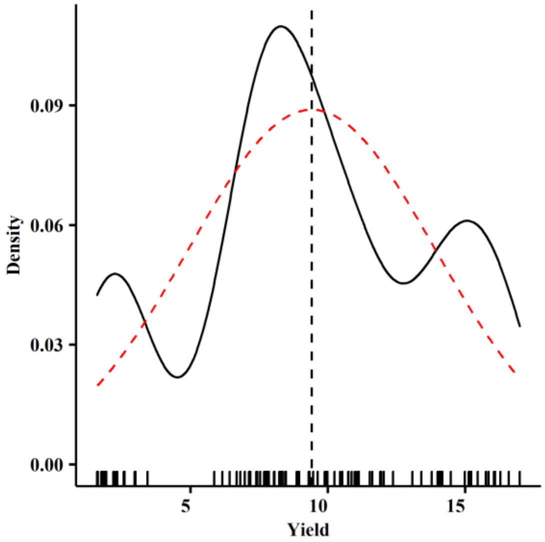
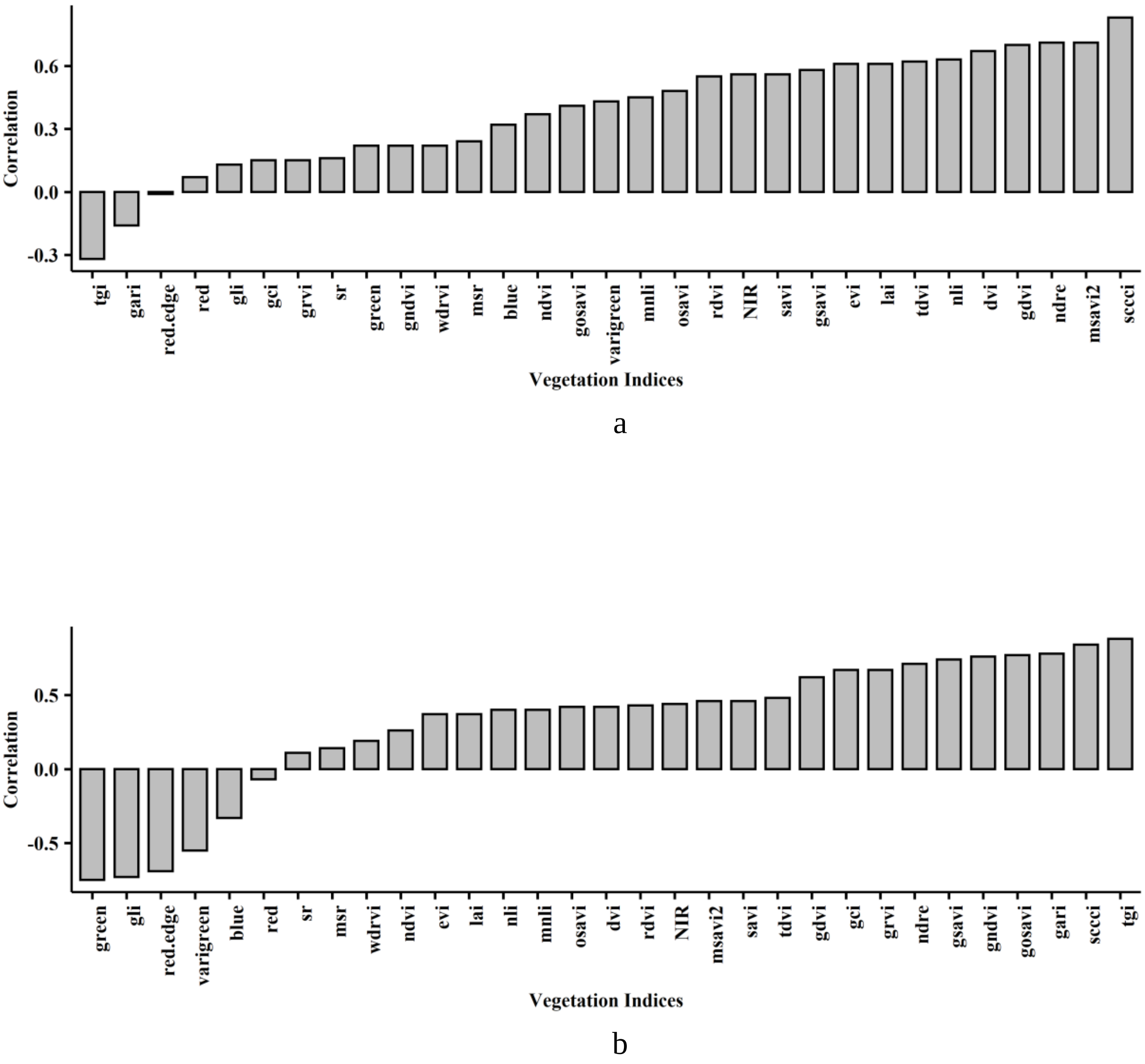


Figure 2.6 Yield density plot (solid line) and corresponding normal distribution (dashed line).



a

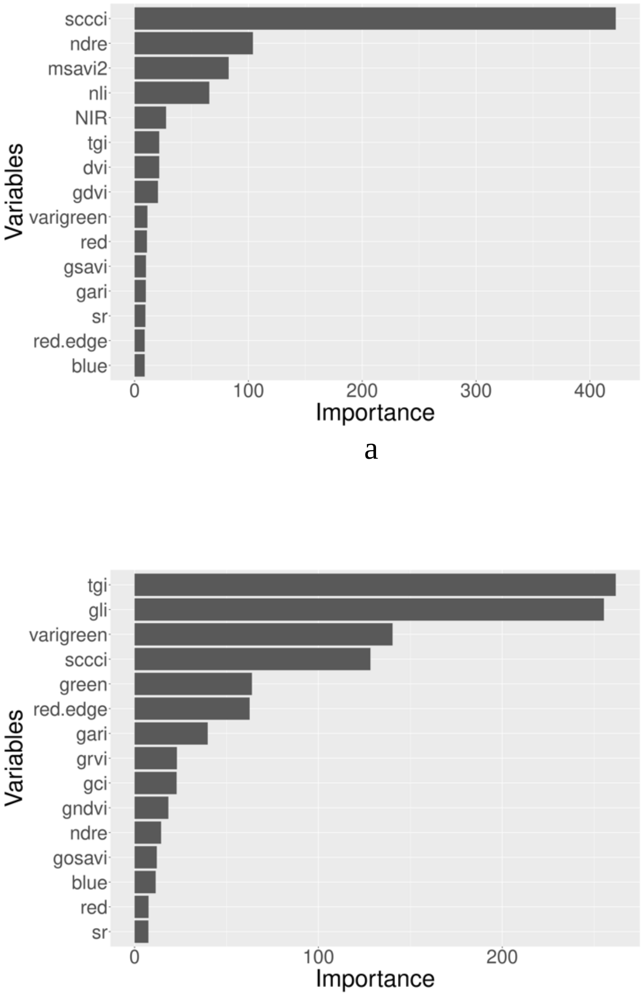
b

Figure 2.7 Correlation between corn yield and 31 variables (5 spectral bands and 26 vegetation indices (VIs): (a) at V4-5 (4-5 leaves with visible leaf collar); (b) at tasseling stage (VT).

The taller the data bar, the greater is the correlation between each variable and yield. As shown in Figure 2.7, the SCCCI, NDRE, MSAVI2, and Green Difference Vegetation Index (GDVI) at V4-5 stage and Triangular Greenness Index (TGI), SCCCI, Green Atmospherically Resistant Index (GARI), and GOSAVI at VT were more correlated with yield as a response variable. However, all of these variables were not included in the final model due to their interaction between the independent variables.

In machine learning modelling, feature selection is primarily focused on eliminating non-informative or redundant predictors from the model. Some machine learning models contain built-in variable selection, meaning that the model will only include variables that improve accuracy. In these cases, the model can pick and choose which representation of the data is best. The most common techniques are using correlation coefficient for selecting the important features correlation. All in all, statistical-based feature selection methods involve evaluating the relationship between each input feature and the target variable using statistics and selecting those input variables that have the strongest relationship with the target variable. Random Forest selected different features for each corn phenological stage. For instance, the

SCCCI, NDRE, and MSAVI2 were the most striking features to predict the yield at the V4-5 stage (Figure 2.8 a). The TGI, Green Leaf Index (GLI), VARI$_{green}$, and SCCCI were the most significant VI’s to estimate yield at the VT stage (Figure 2.8 b).



a

b

Figure 2.8 Feature importance by Random Forest feature selection method: (a) at V4-5; (b) at VT.

Regarding the model selection method, for the VT stage, three predictors had a significant impact on increasing yield prediction accuracy (Figure 2.9). The three variables lead to almost the greatest adjusted-R^2 and the lowest BIC and Cp. Although adding a 4th variable increased the adjusted-R^2 or decreased the BIC and Cp, these improvements were not significant. Therefore, three predictors were used in the yield prediction model for the VT stage, which explained approximately 95% of the corn yield variation.

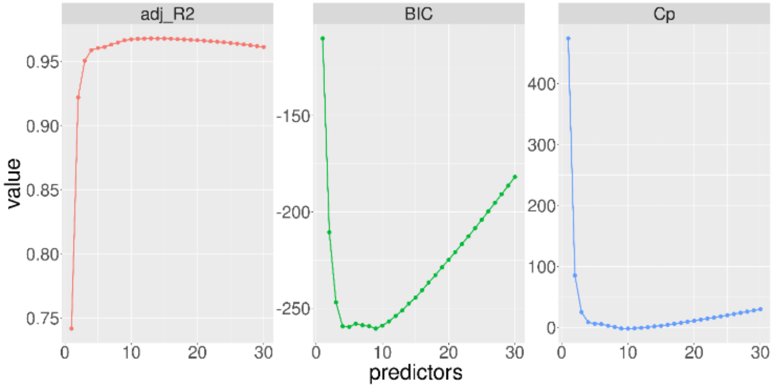


Figure 2.9 The linear model selection at the VT stage.

Moreover, the mean CV error confirmed that the same number of variables were needed for the final model (Figure 2.10). As a result, the best subset selection on the full dataset with the lowest mean square error (MSE) was the 3-variable model used to predict grain yield at the VT stage.

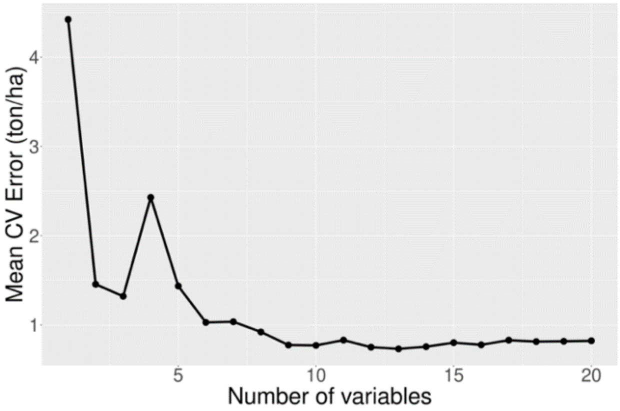


Figure 2.10 Mean cross-validation error versus the number of variables.

To achieve a mathematical yield prediction algorithm, multiple linear regression models were fitted for each phenological stage (Table 2.4).

Table 2.3 Regression models and performance for each model to predict yield at different phenological stages.

Phenological stage	Yield prediction models	R^2-adj
V3	Yield = - 23 + 144.4 OSAVI	0.63
V4-5	Yield = - 13.36 + 45.48 SCCCI	0.69
V6-7	Yield = - 161 + 590.3 GARI + 151.7 NDRE - 456.9 GNDVI	0.70
V10-11	Yield = - 22.64 + 68.93 SCCCI - 19.13 SAVI	0.90
VT	Yield = - 10.96 + 26.07 SCCCI - 68.25 GLI + 13.25 $\mathrm{VARI}_{\mathrm{green}}$	0.93

Although TGI was the most important feature at VT (Figure 2.8 b), it was not statistically significant among other selected variables. Similarly, all these processing methods were applied for each of the five growth stages in order to predict grain yield. Since plant leaf area and metabolism are different at phenological stages, the relationships between yield and spectral bands/VI are likely to differ; therefore, a model for each stage was developed. The coefficients of determination (R^2) for different models at each phenological stage are shown in Table 2.4. All the variables used in these models are statistically significant (at the α = 95% significance level). As the spectral bands and VIs are highly correlated, principal component analysis (PCA) may help, however in this research we used this subjective method to solve the proble.

Furthermore, the measured yield data were compared to the fitted models to evaluate the performance of the algorithms. Figures 2.11 a to 2.11 e illustrate the scatter plot of observed yield values versus predicted yield values at different stages using multiple linear regression. In these figures, the blue line is the 1:1 line, which has a slop of 1, and the red line is the modeled prediction regression line. The regression line indicates how a response variable changes as predictors change. The increasing similarity in the slope and intercepts of the regression line to the 1:1 indicates a more significant model predictive capacity.

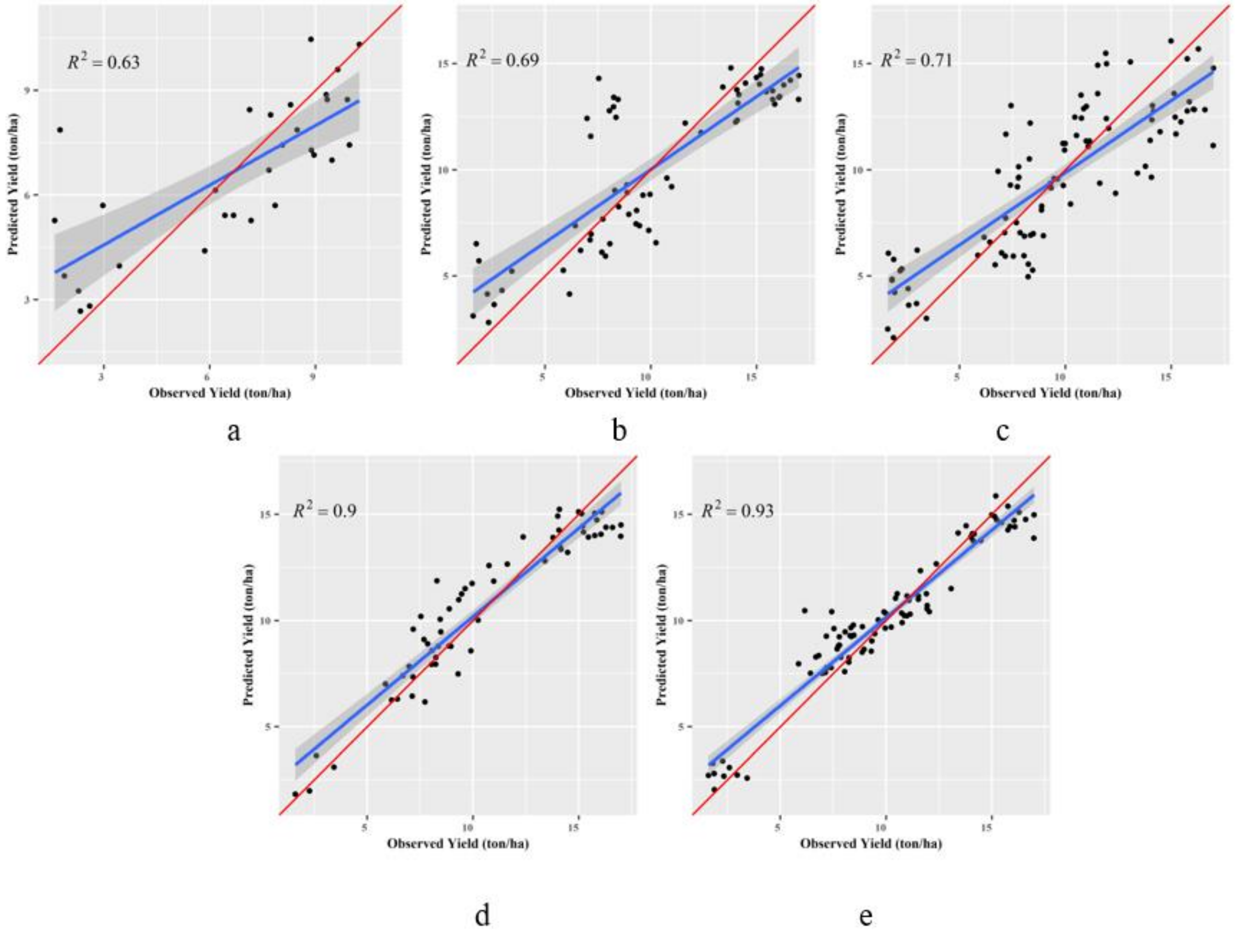
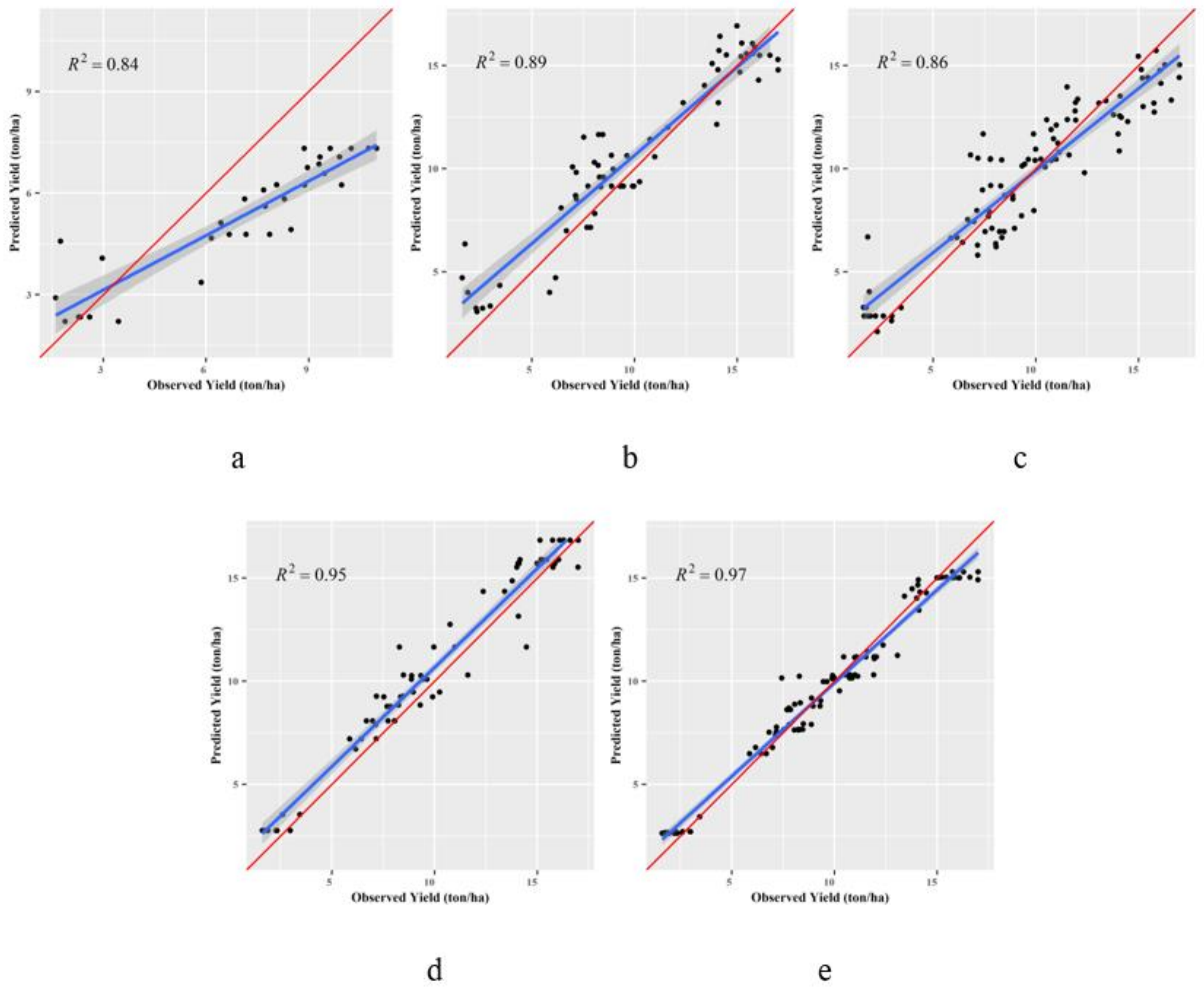


Figure 2.11 Scatter plot of observed versus predicted yield model of corn using multiple linear regression: (a) V3 (3 leaves with visible leaf collar); (b) V4-5; (c) V6-7; (d) V10; (e) VT stages.

The yield model at V3 with one single variable, OSAVI, resulted in the simplest yield estimation model, which has the advantage of simplicity and ease of calculation. As Rondeaux (1996) concluded, OSAVI excels in regions with sparse vegetation where the soil is visible through the plants (Hatfield and Prueger, 2010)(Rondeaux et al., 1996) and, thus, utilization of a VI that corrects for soil interference such as OSAVI worked the best when corn leaf area was at the lowest and bare soil was the greatest among all the sampling times. This model predicted corn grain yield using only one variable at the V3 stage, and the model explained the 63% variation in the corn yield. At the V4-5 stage, the SCCCI predicted grain yield with the highest level of accuracy. The variable selection method indicated that including more variables did not significantly improve the efficiency of the model at V3 and V4-5 stages. Adding a 2nd variable (VI) only increased the R^2 to 0.7 which was not significant. At the V6-7 stage, the GARI, NDRE, and Green Normalized Difference Vegetation Index (GNDVI) together resulted in the best prediction accuracy of grain yield. Although the NDVI is a commonly used index to predict yield, this index saturates when the leaf area index is greater than 1.5 (Trotter et al., 2008). The green NDVI (GNDVI) was most strongly related to grain yield because it has a broader dynamic range than NDVI and combines the green and NIR wavelengths, which are more strongly associated with leaf chlorophyll, N content, and grain yield than the red wavelength (Gitelson et al., 1996; Rattanakaew, 2015; Shanahan et al., 2001). The SAVI is a VI related to biomass to predict yield in highly vegetated areas (Trotter et al., 2008). At the V10-11 stage, SAVI was one of the most effective variables in predicting yield with an R^2 of 0.90. The SCCCI, GLI, and $VARI_{green}$ were found to provide the best predictive accuracy at the VT stage ($R^2 = 0.93$). Moderate to high R^2 values (0.63, 0.69, and 0.70) at V3, V4-5, and V6-7 were observed respectively, and high R^2 values (0.90 and 0.93) at V10 and VT were obtained, respectively. As a result, at V3 and V4-5, the models with a single index, and after V6-7, the predictive models with multiple indices produced the most solid relationships between observed and predicted grain yields. Similar results were reported for lettuce yield prediction (Kizil et al., 2012). To conclude, the VIs used in the predictive models (Table 2.4) were interchangeable with the next most important VIs (Figure 2.8); however, this replacement led to a reduction in the prediction accuracy of the model by about 5%.

Assessment of the gradient-boosting models used for prediction of corn yield for five phenological stages suggested that VI’s derived from MicaSense™ observations can predict corn yield with relatively greater accuracy (Figure 2.12 a to 2.12 e). The R^2 ranged from 0.84 at V3 to 0.97 at VT. Although GBM for the V3 stage hads a relatively high R^2, the root mean square error was high (RMSE = 2.1 ton/ha). Compared with multiple regression models, ensemble learning models resulted in better estimations. For instance, at the V4 to V5 stages, gradient boosting consistently performed better in the prediction of corn yield during both the cross-validation and validation phase. The advantage of the non-parametric ensemble learning models over the regression-based model could be associated with the existence of non-linear relationships between the yield and VIs that Random Forest-based algorithms can integrate during model development. Additionally, the flexibility in the hyperparameters configuration of the boosting methods can result in a greater performance in yield prediction.



a

b

c

d

e

Figure 2.12 Scatter plot of observed versus predicted yield model of corn using gradient boosting machines: (a) V3; (b) V4-5; (c) V6-7; (d) V10; (e) VT stages.

Two critical results were observed from this study: first, as shown in Table 2.4, the SCCCI as a combined index seems to be the most appropriate index for predicting yield (T. B. Raper and Varco, 2015). Second, as corn development progressed, the GBM and regression models predicted final grain yield more accurately, indicating that there is some relationship between the VI and the biomass content of the corn. It appears that the variation in fertilizer N rates provided a good basis for differentiating growth and, ultimately, grain yield, which provided adequate sensitivity for model building and testing.

Conclusions

This study utilized five distinct wavelengths and 26 calculated VIs as input variables to develop regression-based and tree-based learning models at different corn phenological stages to predict corn grain yield at V3, V4-5, V6-7, V10-11, and VT phenological stages. The influence of the variables was found to vary with the phenological stages. In general, the VI that contributed to the majority of the models was SCCCI, suggesting the importance of red-edge-based VIs during yield estimation. At V3 and V4-5 stages, OSAVI and SCCCI were the single dominant features in the yield-predicting models, respectively. The most suitable GBM models with the greatest R^2 values of 0.97 and 0.95 resulted at the V10 and VT stages, respectively. Similarly, the highest R^2 values were obtained at the same stages using regression-based models. Although the R^2 at V10-11 and VT stage are higher than previous stages, applying N fertilizer is not applicable by instrument intalled on a tractor. This cannot make happen because the height of corn is too high, so if we are going to add extra N application, V6-7 stage would be the most commercial stage. When the models' performances were compared for individual stages in both regression-based and tree-based models, however, the accuracies were greater as corn development progressed. One of the goals of this research was to find the models for each stage with minimum error (maximum R^2) by using the appropriate number of predictors. The methodology used in this research can be extended to predict yield for other crops or in other regions as well, where yield prediction is mainly reliant on weather and climatic conditions.

The accuracy of the models in this research might have been affected by different variables, including a smaller number of yield samples collected for each year and the use of limited machine-learning algorithms. In this study, we observed considerable improvement in yield prediction with the use of the ensemble learning model rather than the linear regression algorithm. For corn yield prediction, spectral information, preprocessing, and preprocessing algorithms were important.

The results of this research demonstrated the use of the smallest number of predictive variables that are statistically significant which resulted in an improved explanation for corn yield prediction. Contributing to the accuracy in the predictive capacity of these models included the following: preprocessing of data, including removal of soil pixels, deletion of 3–8% of outliers before conducting the statistical analysis, evaluating for appropriate variables, and selecting appropriate machine-learning model.

CHAPTER III

MACHINE LEARNING APPROACHES TO ESTIMATE CORN LEAF NITROGEN USING MULTISPECTRAL IMAGERY

Introduction

Nitrogen (N) is considered the most critical component of healthy crops. It is one of the major structural elements of chlorophyll content that supports plant growth, development, and productivity. The primary purpose of precision cropping systems is to provide information to make better decisions to use the optimum amount of inputs on the right place and at the right time to enhance the production and optimize the inputs for economic benefit and to protect the environment (Pathak et al., 2018). Since N is a vital element in chlorophyll and it needs for photosynthesis process, helpful information about plants' physiological status can be obtained by the leaf chlorophyll content (Bojović and Marković, 2009; Clevers and Gitelson, 2013; Hunt et al., 2011) A strong correlation between leaf N content and chlorophyll concentration has been found for various plant species (Baret, F., Houlès, V., & Guerif, 2007; Oppelt and Mauser, 2010; Yoder and Pettigrew-Crosby, 1995). Nitrogen deficiency in maize leads to a decrease in chlorophyll concentration in turn changing the leaf color to pale (yellowish-green) with spindly stalks (Sawyer, 2004). As N is a mobile nutrient, this symptom starts from lower (older) leaves and continues to the upper leaves if the deficiency persists. With N stress, absorbance, transmittance, and reflectance are affected (Al-Abbas et al., 1972); canopy reflectance of red light rises, while near-infrared (NIR) reflectance declines. While N plays a critical role in developing phenological stages of corn, over-application of N fertilizer can lead to environmental degradation through the process of leaching away from root zone (Butchee et al., 2011).

Data based digital agriculture has generated interest in the variable rate (VR) applications, which can optimize input application for positive environmental effects and economic benefits, as well as attain efficient N use. Studies illustrated that optical sensing measurements based on VR-N applications may lead to higher N use efficiency in corn production (Stefanini et al., 2019). Farmers can utilize fertilizer more effectively using precise agriculture technologies like real-time on-the-go optical sensing measurements based on varying levels of N application (Stefanini et al., 2019). Zermas et al., (2015) described a methodology to detect N deficiencies in the corn field (Zermas et al., 2015).

Multispectral cameras mounted on Unmanned Aerial Vehicles (UAVs) can collect more detailed and useful local background information with higher spatial resolution. The ground resolution of UAV sensors is around 5 cm, which can be improved by using a high resolution camera sensor and decreasing flight altitude (Iizuka et al., 2018). Besides, UAVs have the advantage of collecting higher temporal resolution images to supplement data from satellites, even on cloudy days (Bondi et al., 2016; Gnädinger and Schmidhalter, 2017; Zhang and Kovacs, 2012). Moreover, the UAV flight altitude is less than clouds height (except in tropical climate) (Ruwaimana et al., 2018), which makes it possible to detect and analyze the land surface data under the clouds (T. Wang et al., 2019). Calibration Reflectance Panels (CRP) are utilized to convert sensor radiance into target reflectance to compensate incident light conditions and generation of quantitative data (Bondi et al., 2016). Therefore, UAV's provide images that are reasonably independent of time and weather conditions. It is also time-saving and provides lower-priced imagery (Gnädinger and Schmidhalter, 2017). Zermas et al., (2015) collected data by sensors mounted on an UAV flying over stressed areas at a low height and captured high-resolution RGB images. Using supervised learning methods to distinguish N deficiency in crop leaves Zermas et al. (2015) reported a performance of 84.2% even if the leaves were covered by other leaves. Machine Learning (ML) is a branch of artificial intelligence for identifying natural patterns based on past data and automates probabilistic analytical models to make better decisions and predictions. As artificial intelligence has become one of the proven technologies to deal with digital information, it could restore significant outcomes while taking into consideration the crop types (Brinkhoff et al., 2019; He et al., 2018; Miyoshi et al., 2020). One of the significant advantages of ML techniques is the capability of autonomously solving huge non-linear problems utilizing multiple sources of data (Chlingaryan et al., 2018). Using ML techniques in multispectral imaging data can reveal physiological and structural characteristics in crops. It also tracks physiological dynamics due to environmental impacts and supports better decision making and decisive actions in real-world situations with or without minimum human intervention. Moreover, it can be applied in field spectroscopy in both the offline and online prediction of parameters in the field (Morellos et al., 2016). These techniques can work with derived spectral indices such as Vegetation Indices (VIs) and whole spectral response traces (Van Wittenberghe et al., 2014). Spectral indices depend on a small number of available spectral bands and therefore do not use the entire information conveyed by the spectral trace. Therefore, it is crucial to find a specific VI suitable for a given task.

Identifying suitable VIs to estimate leaf or plant N concentration requires screening of most useful band combinations from a large range of digital information. Most of the mathematical methods are computationally sluggish or may result in irrational output. This may be due to multiple input parameters, multicollinearity or multiple interactive terms. There are some ML techniques such as Multi-Layer Perceptron (MLP), Support Vector Machines (SVMs), Random Forest regression, Regularized regression models, and Gradient Boosting, which are broadly applicable in precision agriculture. Some of the statistical and ML methods that have been explored for fertilizer N recommendations are: Ridge Regression, Lasso, Elastic Net, Principal Component Analysis (PCA), and Random Forest. All of the above-mentioned methods have their own benefits and drawbacks. Ridge regression, Lasso, and Elastic Net are straightforward and less complicated, but require optimization of hyperparameters and yield considerably poorer results with non-linear relationships (Arruda et al., 2015; Xing et al., 2018). Whereas, PCA accommodates multicollinearity at the cost of scaling and normalization, which may ensure relationships with apparent variables (J. Wang et al., 2017). Compared to Ridge regression, Lasso, Elastic Net and PCA, Random Forest performance is considered to be more accurate and profound. The method can manage non-linear relationships. Random Forest is resilient to outliers and compensates for the overfitting observed during the process (Abdel-rahman et al., 2013). Osco et al., 2020 used machine learning approach to predict leaf N concentration and plant height based on the NDVI, NDRE, GNDVI and SAVI indices, identifying Random Forest as a beneficial predictor for both the N-concentration and plant height (Osco et al., 2020). The feature of inherent regularization and low sensitivity to the data dimensionality makes the technique self-effacing for solving imaging problems (Tuia et al., 2018).

Machine Learning provides a robust and adaptable framework for data-driven decision making and incorporation of professional knowledge into the system. Hyperspectral images used combined with PCA and nonparametric regression algorithms can formulate biophysical and biochemical traits of the crops. Random Forest methods are less time consuming in both training and prediction whereas the kernel methods provide more amenable and manageable in context to hyperspectral imaging (Berger et al., 2020). The objectives of this research were to develop and compare different ML-based models to estimate the leaf N content of corn using different VIs and spectral bands acquired from a red-edge multispectral band sensor mounted on a UAV. Besides, the performance of VIs and spectral bands were compared individually and collectively to see if the combinations of VIs substantially improve results as compared to the original spectral data.

Data and Methods

Site description

The rain-fed corn crop was located in the research field (33°28’13.5” N, 88°45’48.0” W) MS, USA. The total study area was 0.8 ha with a Marietta fine sandy loam (fine-loamy, mixed, siliceous, Aquic Fluventic Eutrochept, thermic) soil. The nutrients K-P-Mg-S were applied according to soil test recommendation and crop maintenance rates, while weeds and pests were controlled based on Mississippi Cooperative Extension Service recommendations. The crop received furrow-irrigation in June 2018 and 2019 due to low total rainfall and visible water stress signs. No additional fertilizer N was supplied to the field throughout irrigation.

The experimental design was a randomized complete block divided into 16 plots with four replicates. Four N fixed fertilizer N rates of 0, 90, 180, and 270 kg/ha were applied on 16 plots randomly. Each plot was comprised of 12 rows of corn planted with a 97 cm spacing between rows with an average row length of 38.1 m (**Error! Reference source not found.**).

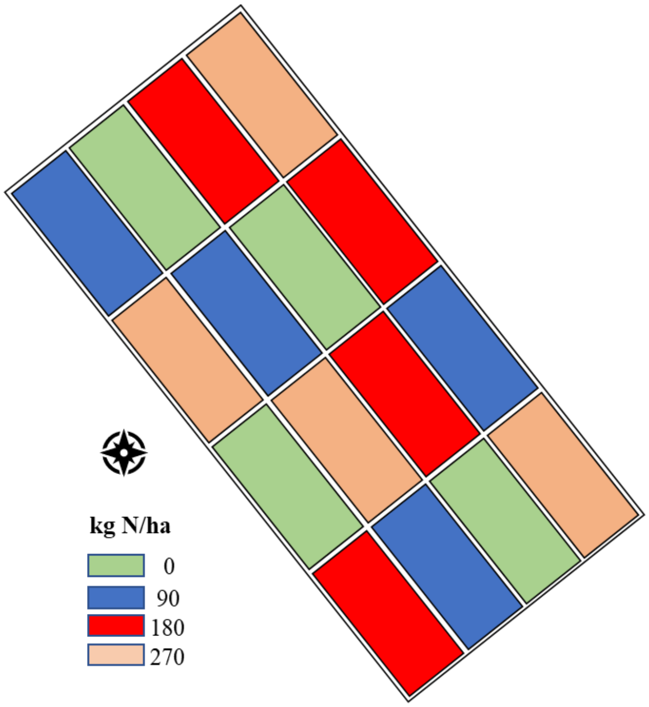


Figure 3.1 Experimental design for a maize field study with four replications of four nitrogen fertilizer treatments, from 2017 to 2019 at Agriculture Systems Research Farm, Mississippi State, US.

Experimental design

Fertilizer N treatments were applied as a side-dressed liquid urea ammonium nitrate (UAN) solution (32-0-0) 2 times during the growing season, using an applicator equipped with coulters and attached liquid knives set at 23 cm from one side of each row and 7.62 cm deep. The maximum N level in corn occurs early in the growing season when plants proliferate (Hogrefe et al., 2017). Therefore, half of the fertilizer N rate for each treatment was applied just after emergence in each year when corn had 1-2 leaves with visible leaf collars (V1-V2 stage) and the remainder was applied when corn had 6-7 leaves with visible leaf collars (V6-V7 stage). Treatments were randomly allocated to the plots and have been repeated each year on the same experimental units. Since 2012, the field study had been on corn and the same N rates of fertilizer had been applied to the plots.

Materials and methods

The multispectral images were captured for three years at 3 different growth stages (V4-V5, V6-V7, and VT) by the MicaSense® RedEdge™ multispectral camera mounted on a UAV. Planting/harvesting, fertilizer N applications, and flight dates are shown in Table 3.2. The spectral reflectance measurements were acquired at 475, 560, 668, 717, and 842 nm (blue, green, red, red-edge, and NIR wavelengths, respectively). All five wavelengths were obtained simultaneously. Each flight took 10 minutes for 0.8 ha. The flight altitude was 60-m in 2017 and 2018, and it was decreased to 30-m in 2019 to increase resolution. The camera was mounted on the bottom of the UAV, and the viewing angle was less than 10^{o} from nadir. The UAV images were taken with at least a 75% horizontal overlap. Then, Pix4Dmapper software was used to stitch and mosaic the images with a spatial resolution of 4-cm. A white calibration reflectance panel was used to compensate for incident light before and after each flight. The CRP has a calibration curve associated with the acquired images across the visible and near-infrared light spectrum. The Geosystems Research Institute (GRI) at Mississippi State University preprocessed all the images and mosaic the image tiles and converted the raw data to the reflectance data.

Table 3.2 Cultural practices and sampling date each year of the study

Cultural Dates	2017	2018	2019
Planting	13-Apr	19-Apr	23-Apr
1st N application	21-Apr	8-May	8-May
2nd N application	16-May	23-May	27-May
Harvest	24-Aug	13/14-Sep	6-Sep
UAV Flight and sampling leaf N	2017	2018	2019
V4	2-May	14-May	23-May
V6	15-May	23-May	30-May
VT	9-Jun	18-Jun	21-Jun

The spectral data were processed to remove soil pixels. The empirical formula of 2*Green – Red – Blue > 0.06 (Barzin et al., 2020) was obtained to distinguish all the vegetative pixels. This formula also removes the corn's shadowed pixels. After removing the soil pixels, five spectral bands (blue, green, red, red-edge, and NIR) were extracted for each pixel and based on these spectral bands, 26 VIs were computed. All these processes were performed by using QGIS v.3.12.0 ("QGIS Development Team," 2020) and R software 'raster' package (R Core Team: Vienna, 2019). The indices and associated spectral bands were documented in Barzin et al. (2020).

Nitrogen sampling

Above ground biomass samples were taken at V4-V5 stages (referred to as V4 sampling), and leaf samples were collected at the V6-V7 stage (referred to as V6 sampling), and before tassel emergence (VT) in order to measure the crop tissue N concentration. Six samples were collected in at three phenological stages from rows 2 & 3, and 10 and 11 (3 samples from each row). The most recent, matured, and fully-collard leaves were selected for analysisand then samples were dried in a forced-air oven at 65°C. The oven-dried leaves were analyzed on a Carlo Erba N/C 1500 automated dry combustion analyzer (Carlo Erba, Milan Italy) to measure the total N concentration.

Multispectral data are visually and numerically similar and highly correlated (Barzin et al., 2017b). Therefore, the analysis of all the original spectral bands is not efficient (Schowengerdt, 2012). Summation, differencing, rationing, and using combinations of different bands make different VIs to compare them in a different time and places (Jackson and Huete, 1991). In this research, all the vegetation indices individually, the 5 original bands individually, and the 5 original bands collectively have been compared with different ML methods. Moreover, three different groups of VIs were used in 3 different data categories to compare the different ML methods for leaf N estimation:

1) Thirty-one variables (five spectral bands + twenty-six VIs).

2) Three different groups of VIs that previous researchers recommended using (2a and 2b) or have a high correlation with N.

2a) The VIs that are important to estimate leaf chlorophyll concentration,

2b) The VIs recommended for N requirement prediction for maize,

2c) The VIs that have a high contribution in N estimation and were statistically significant.

3) Three variables that were selected in the Random Forest feature importance and were statistically significant.

For instance, Green Atmospherically Resistant Index (GARI) is one of the VIs, which has greater sensitivity to a wide range of chlorophyll concentrations (Gitelson et al., 1996). Green Chlorophyll Index (GCI) can be used for chlorophyll content estimation for a variety of plant species (Gitelson et al., 2003). Green Normalized Difference Vegetation Index (GNDVI) is the same as NDVI, except it uses the green wavelength instead of the red wavelength, and is more sensitive to chlorophyll concentration (Gitelson and Merzlyak, 1998a).

Machine learning algorithms

Eight ML techniques were used to find an estimation algorithm to explain the relationship between one dependent variable (%N) and 31 independent variables (5 wavelengths and 26 VIs) to estimate leaf %N. These algorithms include Random Forest (Breiman, 2001), Gradient Boosting (Friedman, 2002), Multi-Layer Perceptron (MLP) (Gardner and Dorling, 1998), Support Vector

Machines (SVM) (Smola and Scholkopf, 2004), Ordinary least squares method (or the standard linear model) (Shiu and Chuang, 2019; Yahya and Olaniran, 2014), and Penalized regression models (Ridge regression (Hoerl and Kennard, 1970), Lasso (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1997) regression, and elastic net regression (Zou and Hastie, 2005). All the ML models were performed in Python 'scikit-learn' (Pedregosa et al., 2011).

Algorithm setup

Several ML regressors were applied to estimate leaf N%. Machine learning models usually have parameters tuned by users known as hyperparameters. Parameter tuning refers tochoosing the optimal value for the settings and optimizing the model for the optimal performance of the model. All related hyperparameters for ML models used in this research, along with their range of possible values to test final optimal values applied for the current dataset in this study are represented in Table 3.3.

Table 3.3 Machine learning models with all related parameters used in this research.

Algorithm	hyperparameter	Range of value for test	Optimal value

Random Forest	Number of estimators (trees)
	Min_sample_split
	Min_sample_leaf
	Max depth of tree
		{20, 40, 60, 80, 100, 120, 140, 160, 180, 200}
		{2, 5, 8, 10, 12, 14}
		{1, 2, 4, 6, 8, 10}
		{2, 3, 5, 6, 8, 10, None}
			40
			8
			6
			5

Gradient Boosting	Max_leaf_node
	Min_sample_leaf
	L_2-regularization
		{2, 4, 6, 8, 10, 12, 14, 16, 18, 20}
		{1, 2, 4, 6, 8, 10}
		$\{1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$
			10
			4
			10^{-4}

SVM	Kernel C (regularization term)	Linear {1, 10, 100, 1000}	Linear 10

Table 3.3 (continued)

MLP	Hidden layer number
	Range of hidden layer neurons
	Activation function
	Optimization technique
	Alpha
	Max iteration
		{1, 2, 3}
		{1-100}
		{logistic, tanh, ReLU}
		{sgd, adam}
		{0, 0.001, 0.0005, 0.00005}
		500
			1
			50
			ReLU
			Adam
			0.001
			500

Ridge regression	L_2 penalty	$\{1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$	10^{-2}
Lasso regression	L_1 Penalty	$\{1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$	10^{-4}
Elastic Net	L_2 penalty L_1 Penalty	$\{1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$ $\{1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$	10^{-2} 10^{-4}

The standard way to tune the hyperparameter and select the optimal value for them in more ML-like approaches is to perform cross-validation and select the value that minimizes the cross-validated sum of squared residuals (error). As can be seen in Table 3.3, a set of values for each hyperparameter were selected through cross-validation. Then the dataset was split into k folds, and each time the k-1 fold was used for training the model and the rest of the data was used for testing it. Later on, the strategy was repeated for k times of k folds and the optimal value was selected for each hyperparameter based on minimizing the sum of squared residuals. In this work, cross-validation with five folds has been applied to detect the optimal value for each of the hyperparameters and the input dataset was scaled to the standard deviation of 1 and mean of zero, as for all ML models standardizing the input data is important to avoid bias in the training process. In addition, between 20 to 200 trees have been tested for Random Forest and gradient boosting to find the optimal number of trees as a predictor. There are several ways to control the overfitting problem (Breiman, 2001; Vezhnevets and Barinova, 2007). Different values for each of these hyperparameters for Random Forest and gradient boosting models were set by using the cross-validation technique and the resulting values were shown in Table 3.3. In the SVM model, a regularization term (C) was added to the loss function to compensate the overfitting issues. The C value serves to identify the greatest value for the? margin with the lowest mis-predicting. In this work, different values between 1-1000 have been tested based on the cross-validation technique, and the optimal value for linear SVM was 10 for input data. For the MLP model, the different combinations of hidden layer numbers (1, 2, and 3), range of hidden layer neurons (between 1-100), different optimization and activation techniques have been tested. The best combination and fit model for MLP based on input data were having only one layer with 50 neurons and the activation function was ReLU with Adaptive Moment Estimation (adam) optimization technique. To avoid the overfitting problem, different values for alpha were tested, and the best value was 0.001. Ridge and Lasso regression were used to quantify the overfitting of the data by measuring the magnitude of coefficients. For all three regularized regression models, different values have been tested for L1 and L2 penalty terms which ended up to10^{-2} for L_1 and 10^{-4} for L_2.

Results and Discussion

All the VI individually, the 5 wavebands individually, and the 5 wavebands collectively were compared in different ML methods for estimating corn leaf N leaf %. Figure 3.2 illustrates the resulting R-square for each method. Gradient boosting and the Random Forest method had the greatest R-squares when using the VIs and band variables individually. Although some variables such as VARIGREEN, SCCCI, Red-Edge, NIR, etc. resulted in small R-squares in the estimatation of leaf N%, it does not mean they are not valuable when they are used in a combination with other variables. The correlation of all the variables was considered in the Random Forest feature selection method and this method can be used for selecting the features. As can be seen in the figure 3.2, using VI as an input substantially improved the models in comparison to using wavebands individually or even all the 5 bands collectively. The R-squeres wereincreased using VIs and spectral bands together.

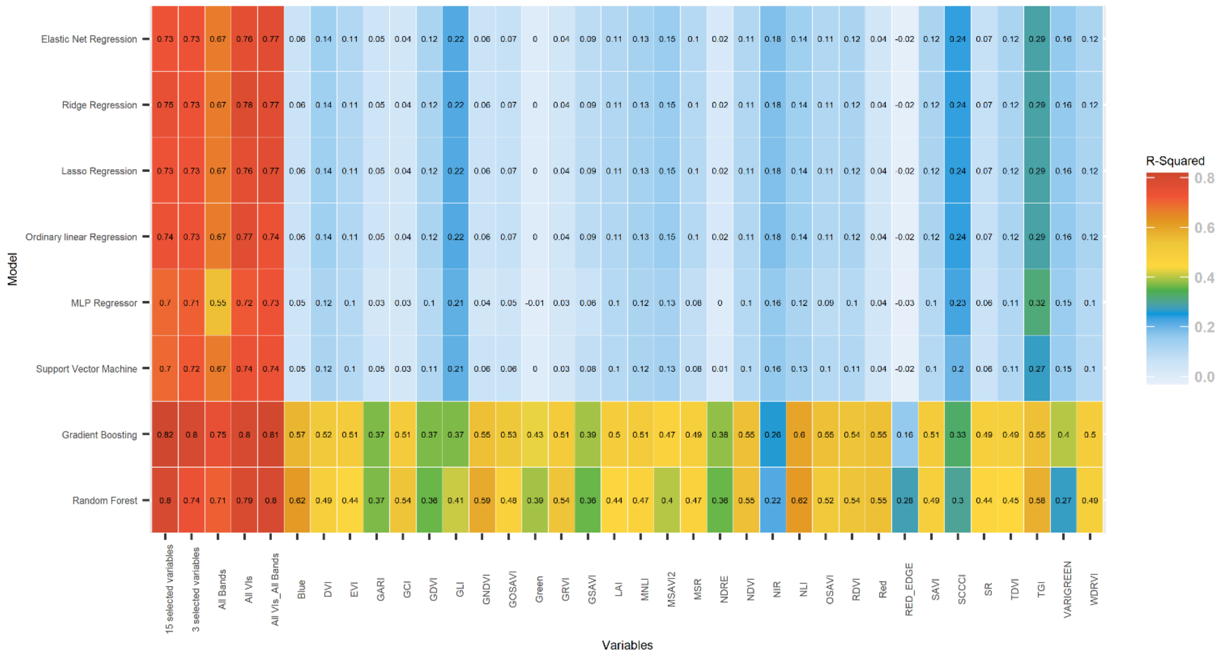


Figure 3.2 R-squared derived from comparison of all the 26 VIs individually, the 5 wavebands individually, and the 5 wavebands collectively to estimate leaf N content of corn using different ML models.

Three different groups of VIs which were derived from UAV imagery were used in 8 ML models. All 31 variables (26 VIs and five wavebands) were used for the first comparison between these models to find which of the ML models can predict leaf N %most accurately. Figure 3.3 shows leaf N% estimation using different ML methods in category 1. As can be seen in figure 3.3, the coefficient of determination for Random Forest, Gradient Boosting, SVM, MLP, Ridge

Regression, Lasso Regression, and Elastic Net models were 0.80, 0.81, 0.74, 0.73, 0.74, 0.77, 0.77, and 0.77, respectively. Therefore, it can be concluded that gradient boosting and Random Forest are the best methods to estimate plant %N with the highest coefficients of determination among all other models. Random Forest and gradient boosting are regression trees methods (Liaw and Wiener, 2002) and both of them utilize a similar strategy to combine a set of weak learners into strong learners by a different method of training data (Barzin et al., 2020; Friedman, 2002). Random Forest regression applied on hyperspectral data has shown potential to accurately predict the leaf N concentration of sugarcane (Abdel-rahman et al., 2013). Li et al., (2016), indicated that the Random Forest model provides the most accurate prediction in comparison to regression models (Z. Li et al., 2016). Zha et al., (2020), illustrated that the Random Forest method performed better than other ML algorithms in the estimation of a N nutrition index (Zha et al., 2020).

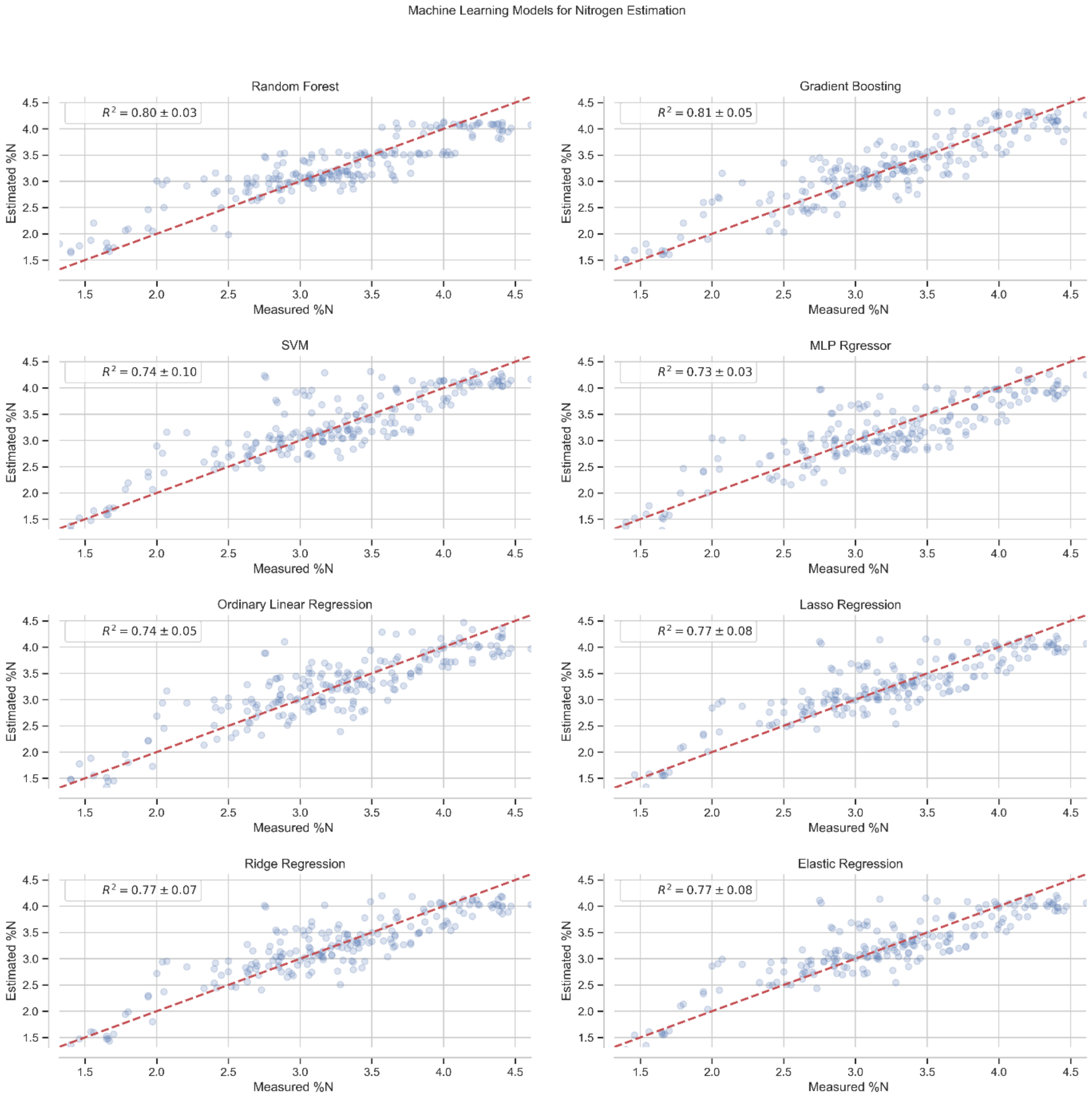


Figure 3.3 Comparing different models for corn N estimation based on 31 variables.

The relative importance of each variable by the Random Forest feature selection method is illustrated in Figure 3.4. As can be seen in this figure, the height of each bar is indicative of the variables contributionto the model derived by this method. The results showed that the most effective variables to estimate leaf N % were NDRE, SCCCI, and Red-edge, as gathered from the feature importance plot. These variables are statistically significant based on the Analysis of Variance (ANOVA). While TGI was the most important variable in the Random Forest feature importance with the tallest bar, it was not significant at a 95% confidence level. Therefore, for category 2, 15 out of 31 variables were selected as acceptable variables to be used to predict N requirements for maize (GDVI, GSAVI, and GOSAVI), and leaf chlorophyll content (GARI, GCI, GNDVI, NDVI, and TGI), or have a greater contribution in estimating leaf N concentration based on Figure 3.4 (NDRE, NIR, SCCCI, Red-edge, GRVI, LAI, EVI, and Green). Previous studies have also shown similar VIs selection for N requirement prediction for corn (Louhaichi et al., 2001; Sripada et al., 2006) and leaf chlorophyll content estimation (Gitelson et al., 1996, 2003; Gitelson and Merzlyak, 1998b; Hunt et al., 2011; Rouse et al., 1974).

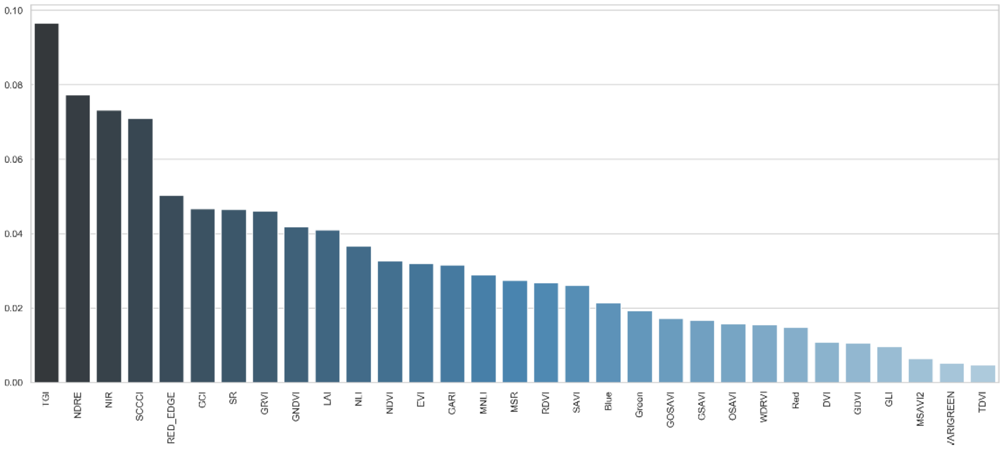


Figure 3.4 Feature importance based on Random Forest

As a result, eight different ML models were applied to these selected variables to estimate leaf N%. Figure 3.5 illustrates the comparison of ML methods in using 15 variables of category 2. As can be seen in Figure 3.5, the coefficients of determination for Random Forest, Gradient Boosting, SVM, MLP, Ridge Regression, Lasso Regression, and Elastic Net models were 0.81, 0.82, 0.70, 0.71, 0.74, 0.73, 0.75, and 0.73, respectively. Therefore, it can be concluded that the gradient boosting and Random Forest methods resulted in the best prediction of leaf N%.

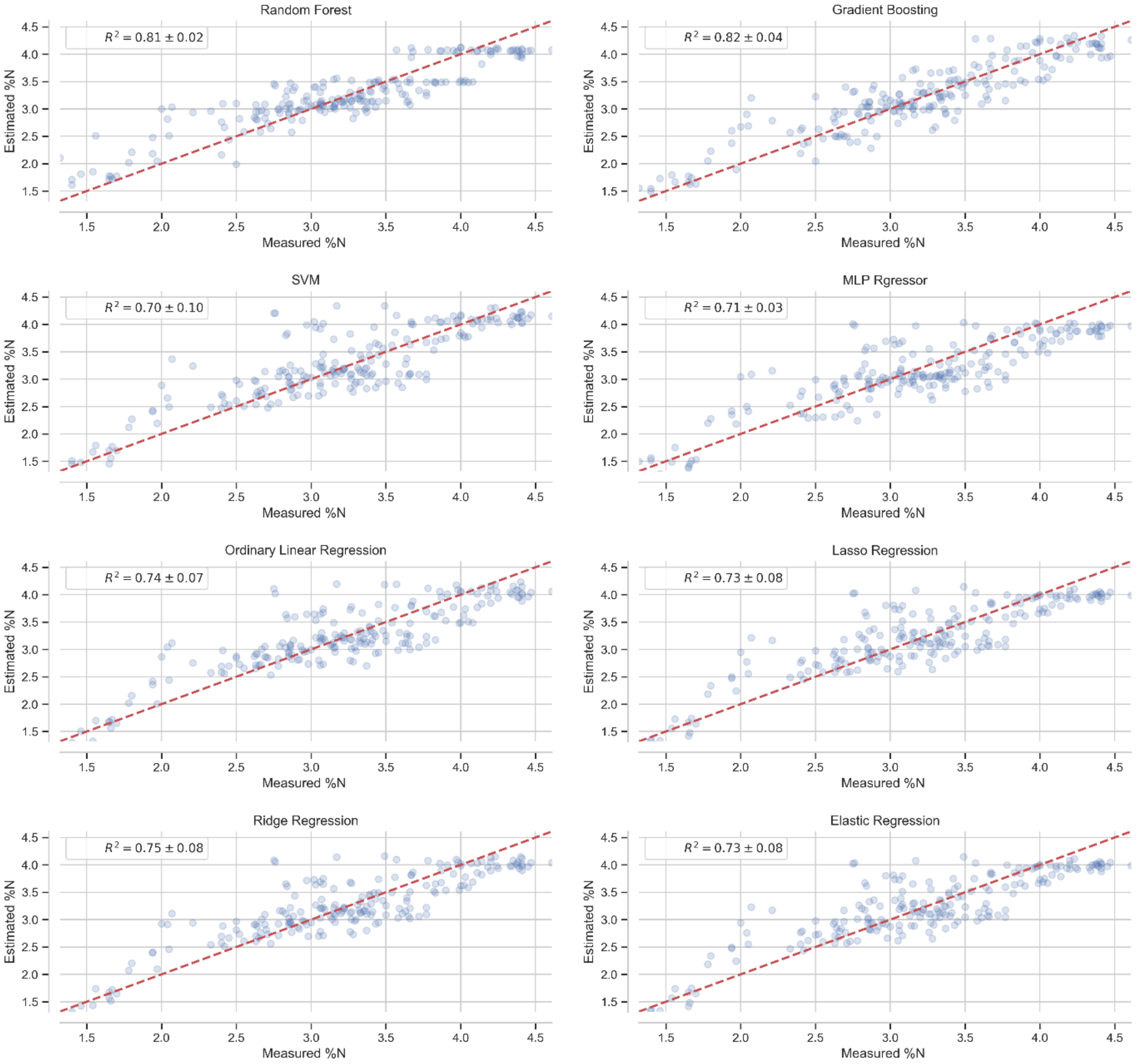


Figure 3.5 Comparison of different models for leaf N% estimation based on 15 selected variables (GDVI, GSAVI, GOSAVI, GARI, GCI, GNDVI, NDVI, TGI, NDRE, SCCCI, CGI, GRVI, EVI, Red-edge, LAI).

For the third category, the three variables that had the most influence in the Random Forest feature selection method and were statistically significant (i.e., Red-edge, SCCCI, and NDRE) were used to import in the ML models. The correlation of all the variables was considered in the Random Forest feature selection method and this method can be used for feature selection.

Moreover, using VIs as the inputs substantially improved the models in comparison to wavebands individually or even all the 5 bands collectively. Previous studies illustrated that SCCCI has a strong relationship with fertilizer N rates and can estimate leaf N concentration accurately (Fox, 2015; T. B. Raper and Varco, 2015; Sumner, 2019). Besides, it is sensitive to N status early-season when N fertilization decisions have a huge impact on yield results (T. B. Raper, 2011). Moreover, NDRE and red-edge have a strong relationship with the N status indicators (Cao et al., 2018; T. B. Raper and Varco, 2015; Sumner, 2019). Figure 3.6 illustrated the comparison of different ML methods in category 3 with three variables. Based on the trends shown in **Error! Reference source not found.**this figure, gradient boosting as the non-parametric models have provided the best results with a 0.80 coefficient of determination. In general, gradient boosting can better prevent the possibility of overfitting which normally occurs with decision tree algorithms. More accurately, gradient boosting has a greater robustness, in that it is less likely to be affected by the scale of training datasets, and that outliers and less-related variables cannot simply change its performance (Ye et al., 2009). In this analysis, MLP compared to Random Forest and gradient boosting presented the least correlation with 0.71 for R2. Ordinary linear regression has shown a good correlation with other regression models. Moreover, ridge, Lasso, and elastic net regression provided very close results to each other all with a 0.73 coefficient of determination. Although some variables individually such as VARIGREEN, SCCCI, Red-Edge, NIR, etc. have a small R-squared with leaf N%, it does not mean they are not valuable when they used in a combination with other variables.

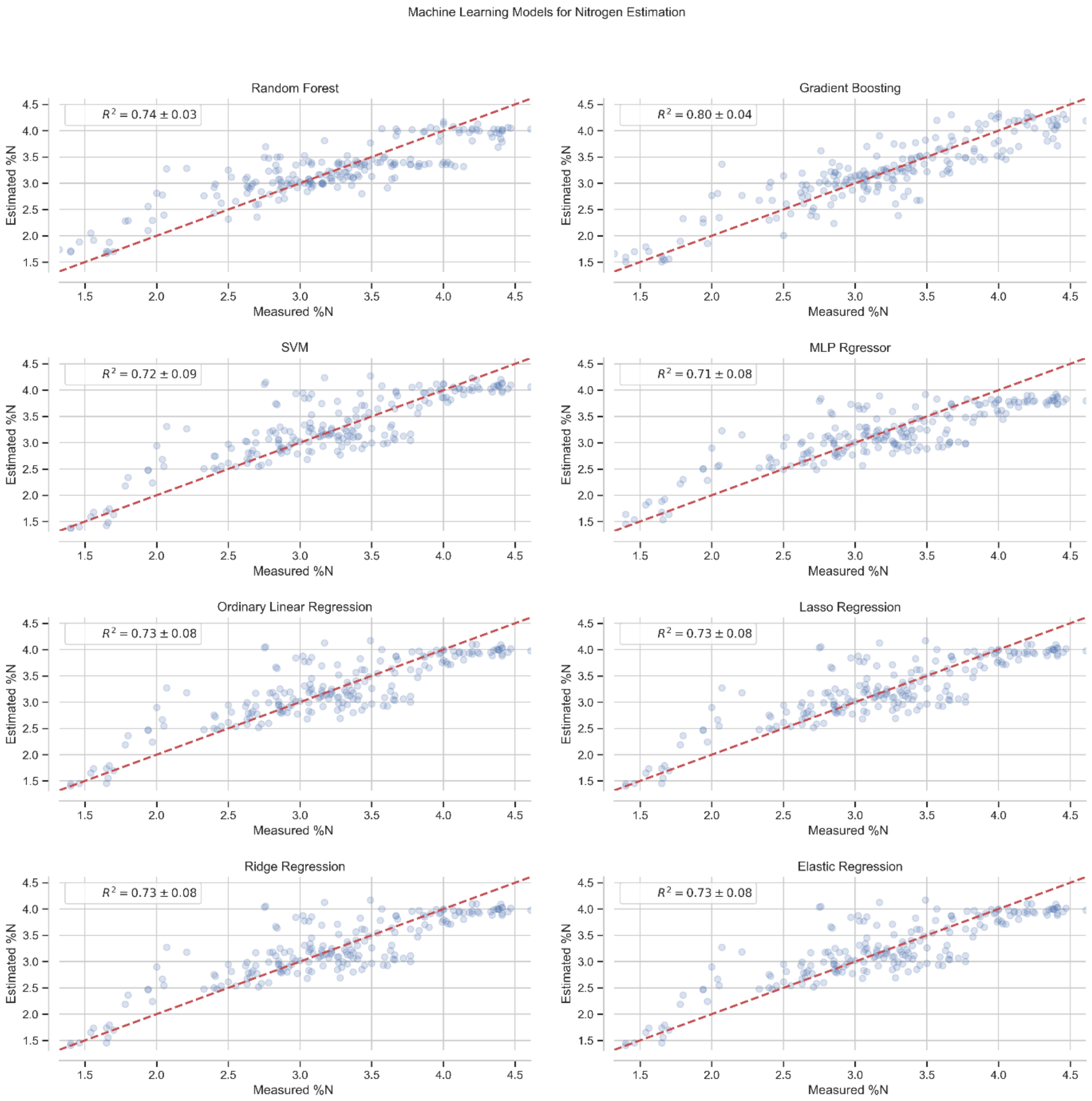


Figure 3.6 Comparing different models for N estimation based on three selected variables (Red-edge, SCCCI, and NDRE).

As a result, between 8 ML models used in this research, gradient boosting was the best-fitted model to estimate leaf N % with the greatest coefficient of determination by using a different number of variables in the models (the 31 variables, 15 selected variables, and even three variables used in category 1, 2, and 3, respectively. Moreover, using VIs as substantially improved the models in comparison to wavebands individually or even all the 5 bands collectively.

Conclusion

This study was conducted to develop ML models for estimating leaf N % of maize by using several VIs and spectral wavelengths. Three approaches were used to select the inputs for the ML models: 1) twenty six VIs and five wavebands (31 inputs) in the ML models, 2) 15 out of 31 variables which previous research projects recommended for leaf N concetration estimation, leaf chlorophyll content and the VIs that have a high correlation with N, and 3) three VIs that Random Forest selected as the important variables in estimation leaf N and those VIs were statistically significant at 95% confidence interval. The Random Forest feature selection method illustrated that red-edge, SCCCI, and NDRE were the most effective variables in estimating leaf N% in maize. Furthermore, this research attempted to evaluate the performance of a ML algorithm and find the most accurate method for leaf N concentration estimation. Based on the 8 ML algorithms, the gradient boosting had the greatest coefficient of determination in the estimatation of leaf N%.

CHAPTER IV

EVALUATING CROP CIRCLE MULTISPRCTRAL ACTIVE CANOPY SENSOR FOR PREDICTION OF CORN LEAF NITROGEN CONCENTRATION AND YIELD USING MACHINE LEARNING

Introduction

Agricultural products play a major role in feeding the world's population and have a remarkable impact on people's life and work by providing food and fuels. Agricultural lands expanded around 10 million ha every year during 1980 and 2007 (West et al., 2010) to aid meet the needs of a growing population, changing diets, and boosted biofuel requirement. However, in 2010–2012, around 870 million people consumed an insufficient quantity of food to cover their minimum dietary energy needs (McGuire, 2013). The United States Department of Agriculture (USDA) reported that approximately 36 million hectares of corn were planted in the US and 0.27 million hectares in Mississippi and 11.3 ton/ha and 11.7 ton/ha corn grain were harvested in the US and Mississippi, respectively in 2019 (USDA/NASS, 2019). Therefore, the importance of increasing production efficiency to meet global food demands is critical to maintaining or increasing economic viability while minimizing environmental hazards.

Nitrogen (N) is a primary nutrient required for plant growth, development, and reproduction of healthy plants, and it plays a key role in the developing phenological stages of corn since N is directly related to photosynthesis (Andrews et al., 2013). An increase fertilizer N application rates has contributed substantially to an increase in yield of grain crops throughtout the world (Cassman et al., 2003). However applying fertilizer N in excess of crop demand can result in decreased economic returns, low Nitrogen Use Efficiency (NUE), poor nutritional quality, and a negative environmental impact (Gautam and Panigrahi, 2007; Kim and Dale, 2008). On the other hand, in response to N deficiency, N transfers from lower (older) tissues to the meristematic region (younger ones) (Raper et al., 2013). Nitrogen deficiency always leads to a decrease in leaf chlorophyll concentration, which is coupled with changing leaf color from dark green to light green or yellow (Bronson et al., 2005). This distinction was associated with physiological and structural changes in cotton leaves (Fridgen and Varco, 2004), which results in an increase leaf spectral reflectance in visible wavelength range (400-700 nm) (Zhao et al., 2003). Moreover, NIR reflectance changes due to N deficiencies, in which these wavelengths are increasingly used for estimating crop N, especially at early growth stages (F. Li et al., 2010). Therefore, N deficiency strongly influences the phenotypic characteristics of crops (Zhu et al., 2014).

Applying the optimum amount of fertilizer N can have a considerable impact on grain yield. Improving grain yield production and quality of crops using optimal fertilizer N application rates , as well as proper application of pesticides, herbicides, and other inputs are the primary goals of precision agriculture. Critical parameters such as fertilizer and irrigation management, weather conditions, topography, and soil properties affect potential growth and yield. Multispectral sensors can be used to provide to assist in the accurate and timely application of these inputs for large agricultural fields providing spectral, spatial, and temporal information related to crop growth. Tracking corn growth rates during the growing season to estimate yield is important to efficiently managing N fertilization. Several studies used remotely sensed sensor's data to predict yield (Chang et al., 2003; Reyniers et al., 2006; Sakamoto et al., 2014; Solari et al., 2008; Tadesse et al., 2015; Uno et al., 2005). In the past, it was common to use a chlorophyll meter (SPAD) and associated leaf color charts to obtain point measurements of crop N status (Dobermann et al., 2002; Gitelson and Merzlyak, 1998a); however, this method is time-consuming and labor-intensive when used to define spatial structure in large production fields. In recent years, there is an increasing interest in using active field sensors and passive multispectral sensors for estimating yield and plant N due to a greater efficiency in mapping large areas (Cao et al., 2013; Yao et al., 2012).

Multispectral sensors can quickly collect spectral information on actively growing crops aross fields. One such commercially available handheld multispectral sensor is the Crop Circle ACS-430 (Holland Scientific Inc., Lincoln, Nebraska, USA). It is an active canopy sensor with its own source of illumination. It simultaneously measures reflectance at 3 wavelengths continuously, which can be used for computing Vegetation Indices (VIs). A vegetation index is a single value calculated index using several mathematical combinations of different spectral wavelengths. Vegetation indices can be used for estimating various physiological characteristics such as biomass, leaf area, vegetation cover, leaf chlorophyll content at different growth stages for different crops (Hatfield and Prueger, 2010). For instance, Cao et al. (2013) evaluated 43 VIs gained from reflectance at the green, red-edge and NIR wavelengths acquired with a Crop Circle ACS-470 to estimate N status in rice, Trotter et al. (2008) used three VIs derived from a Crop Circle^{TM} for biomass assessment under differing farming environments. Raper et al. (2013) used the normalized difference vegetation index (NDVI) to predict N status of cotton leaves, while Cao et al. (2017) used a Crop Circle ACS-470 to develop a precision N management strategy for winter wheat in the north China plain and compare it with GreenSeeker, evaluated the performance of a Crop Circle ACS-470 (CC-470) and Crop Circle ACS-430 (CC-430) for N status estimation of winter wheat at different height and growth stages, and Shi et al. (2013) evaluated the Crop Circle ACS-470 sensor observations to estimate N status and yield for rice in China There are other handheld devices such as GreenSeeker handheld crop sensor (Trimble Navigation Limited, Sunnyvale, California, USA), Crop Circle ACS-211 (Holland Scientific Inc., Lincoln, Nebraska, USA), and CropSpec (Topcon Positioning Systems, Inc., Livermore, California, USA), which measure just two wavelengths (red and NIR, NIR and green, NIR and red-edge respectively). One of the limitations of using these sensors is a lower number of spectral bands and VIs limiting the capacity for using differing indices for different plant biophysical parameters and phenological stages (Hatfield and Prueger, 2010; F. Li et al., 2010). So, having more than 2 wavelengths can improve a canopy sensors' utility (Cao et al., 2013).

Machine learning (ML) techniques are applied to multispectral images which can be utilized to illustrate physiological and structural attributes of plants and their response to environmental stress (Wahabzada et al., 2016); Moreover, ML imports several variables such as meteorological data, soil moisture, irrigation, spectral bands as inputs to predict fertilizer N requirements or automated recommendations for irrigation (Goldstein et al., 2018). Barzin et al. (2020) applied 8 ML methods on 5 spectral bands and various VIs to find the best method for estimating leaf N in corn. Gutiérrez et al. (2018) applied an ML algorithm on thermal imagery in order to develop a new technique for fast and reliable water status estimation in a vineyard. Weng et al. (2018) used one of the ML methods (least squares-support vector machine classifier) on hyperspectral images in order to detect Huanglongbing disease and nutrient deficiency in the citrus orchard during hot and cold seasons.

The primary objective of this study was to estimate corn leaf N concentration and grain yield using a multispectral handheld sensor to acquire canopy reflectance. This study employed four ML algorithms to find the best prediction model by using Ccrop Circle ACS-430 field sensor measurements at different growth stages. The results of this project provided detailed information regarding three spectral bands (red, red-edge, and NIR) and 25 VIs to estimate the leaf N status and predict the yield of corn potentially. Additionally, the accuracy of wavelength-based and VI-based models were compared to examine the best model inputs.

Data and Methods

The study was undertaken on an experimental plot at Mississippi State University.

Study site description and Experimental design

The data was collected during the 2019 corn growing season using a Crop Circle ACS-430 at the W.B. Andrews Agriculture Systems Research Farm at Mississippi State, MS, USA (33°28’13.5” N, 88°45’48.0” W). The field area was 0.8 ha with Marietta fine sandy loam (Fine-loamy, siliceous, active, thermic Fluvaquentic Eutrudepts). The field was rainfed and the average precipitation and temperature during the 2019 growing season was 76 cm and 22 degrees Celsius, respectively (Barzin et al., 2020).

The experimental field study was divided into 16 plots. Twelve rows of corn were planted in each plot. There was a 97 cm space between each row that had a 38 m length and a 3 m alley in between each plot. Four fertilizer N levels (0, 90, 180, and 270 kg/ha) with four replicates were applied (Figure 3.1). All the treatments were randomly assigned to each plot. The experimental design of the field was a randomized complete block. Corn (DeKalb Brand-DKC67-72 variety) was planted on April 23, 2019, at the Mississippi State research farm. Soil samples were taken before planting and analyzed utilizing Mississippi soil test extraction. The field received uniform applications of P-K-Mg-S before planting based on soil test results: one part concentrated super phosphate (0-46-0), two parts muriate of potash (0-0-60), and one part sulfate of potash-magnesia (0-0-22-11Mg-22S) and was applied at a material rate of 224 kg ha-1. Moreover, weeds and pests were managed based on Mississippi State University Extension recommendations. The fertilizer N source was liquid urea ammonium nitrate (32-0-0), which was applied as a side dress. The N fertilizer was applied as two splits: 50% after emergence, when corn had 1-2 leaves with visible leaf collars (V1-V2 stage) on May 8, 2019, and 50% of N fertilizer was applied when corn had 6-7 leaves with visible leaf collars (V6 stage) on May 27, 2019. The experimental field has been devoted to corn production since 2012 with the same fertilizer N rates assigned to individual plots.

Data Collection

The Crop Circle ACS-430 (Holland Scientific Inc., Lincoln, Nebraska, USA) was used in 2019 to collect canopyreflectance data at 670, 730, and 780 nm (red, red-edge, and NIR spectral bands) from rows 2 and 3 and rows 10 and 11 of each plot (Figure 4.1). Spectral reflectance data was simply and instantly recorded as a CSV file on an SD flash card using the Holland Scientific GeoSCOUT X datalogger (Figure 4.2). It also measured the NDVI and NDRE values directly with geolocation of each. The sensor's field of view was an oval of ~30 degrees by ~14 degrees. The sensor to canopy distance can be typically between 25 to 180 cm based on the device's operation manual; however, in this research, the sensor was held approximately 60 cm above the canopy with a speed of 10 Hz, while walking thru each plot with a constant speed. The Crop Circle ACS-430 refers to reflectance measurements as Pseudo Solar Reflectance (PSR), which means the spectral reflectance wavelengths are scaled as percentages and will not differ with sensor height above a target (Cao et al., 2018). The Crop Circle ACS-430 has an internal GPS to record the latitude, longitude, and elevation of each point, but it was not considered accurate enough for our purposes. Therefore, it was connected to a Piksi Multi Evaluation Kit (Swift Inc., Canada) as a Real-time Kinematic (RTK) GPS. Data was extracted with a 3-m reduction from the beginning and end of each plot length. This reduction was applied to skip the first and last crop canopy of each plot to eliminate border effects. Data was collected at three phenological stages (V4, V6, and VT) around 10:30 am for each stage. The average reflectance values were computed to represent rows 2, 3 and 10, 11 of each plot (Figure 4.1). The calculated spectral VIs using red, red-edge, and NIR are listed in Table 4.2. R software was employed for all the mathematical and statistical analysis used in this research.

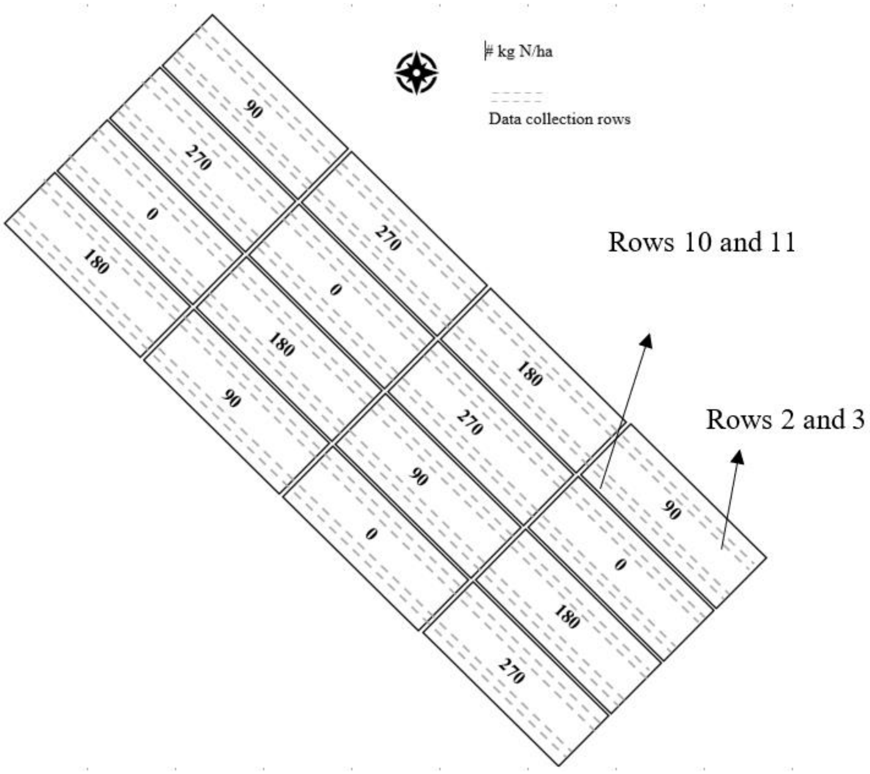


Figure 4.1 Fertilizer N treatment and 4 replicates for corn in 2019 at Agriculture Systems Research Farm, Mississippi State, US.

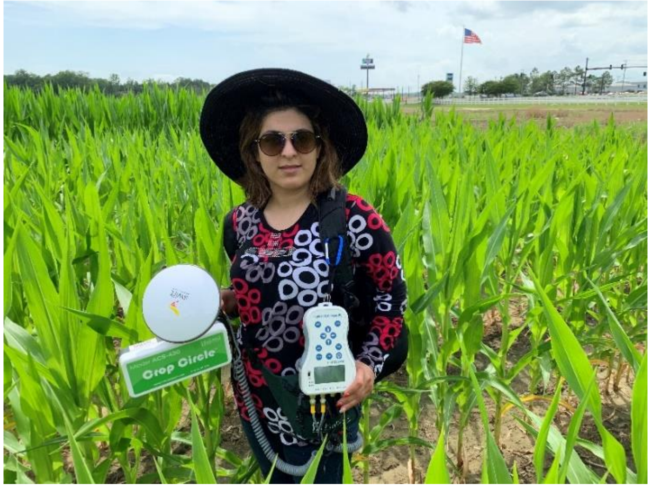


Figure 4.2 Crop Circle ACS-430 and GeoSCOUT X datalogger connected to RTK GPS.

Table 4.2 calculated spectral vegetation indices using red, red-edge, and NIR spectral bands.

	Vegetation Indices		Formula	Reference
1	Normalized Difference Vegetation Index	NDVI	(NIR-Red)/(NIR+Red)	(Rouse et al., 1973)
2	Renormalized Difference Vegetation Index	RDVI	(NIR-Red)/$\sqrt{NIR + Red}$	(Roujean and Breon, 1995)
3	Transformed Difference Vegetation Index	TDVI	1.5(NIR-Red)/$\sqrt{\mathrm{NIR}^2 + \mathrm{Red} + 0.5}$	(Bannari et al., 2002)
4	Difference Vegetation Index	DVI	NIR-Red	(Tucker, 1979)
5	Red edge difference vegetation index	REDVI	NIR-Red-edge	(Tucker, 1979)
6	Red edge re-normalized different vegetation index	RERDVI	(NIR-Red-edge)/$\sqrt{NIR + Red - edge}$	(Cao et al., 2013)

Table 4.2 (continued)

7	Normalized Difference Red-Edge	NDRE	(NIR-Red-edge) / (NIR+Red-edge)	(Gitelson and Merzlyak, 1994) (Raper & Varco, 2014)
8	Simplified Canopy Chlorophyll Content Index	SCCCI	NDRE / NDVI	(T. B. Raper and Varco, 2015)
9	Non-Linear Index	NLI	$(NIR^2-Red)/(NIR^2+Red)$	(Vescovo and Gianelle, 2008)
10	Modified Non-Linear Index	MNLI	$(NIR^2-Red) *(1+0.5)/ (NIR^2+Red+0.5)$	(Gong et al., 2003) (Feng et al., 2019)
11	Soil Adjusted Vegetation Index	SAVI	1.5*(NIR-Red)/(NIR+Red+0.5)	(Rondeaux et al., 1996)
12	Optimized Soil Adjusted Vegetation Index	OSAVI	(NIR-Red)/(NIR+Red+0.16)	(Rondeaux et al., 1996)
13	Modified Soil Adjusted Vegetation Index 2	MSAVI2	(2NIR+1-$\sqrt{(2\mathrm{NIR}+1)^2-8(\mathrm{NIR}-\mathrm{Red})}$)/2	(Qi et al., 1994)
14	Simple Ratio	SR	NIR/Red	(Fraser and Latifovic, 2005)
15	Modified Simple Ratio	MSR	(NIR/Red)-1/$\sqrt{(\mathrm{NIR}\,/\mathrm{Red})+1}$	(J. M. Chen, 1996)
16	Wide Dynamic Range Vegetation Index	WDRVI	(0.1 NIR-Red) / (0.1 NIR + Red)	(Gitelson, 2004)
17	Red-edge wide dynamic range vegetation index	REWDRVI	= (0.12* NIR − Red-edge)/(0.12 * NIR + Red-edge)	(Cao et al., 2013)
18	Red-edge ratio vegetation index	RERVI	NIR/ Red-edge	(Tucker, 1979)
19	Red-edge difference vegetation index	REDVI	NIR − Red-edge	(Tucker, 1979)
20	Red-edge chlorophyll index	CIRE	(NIR/Red-edge) – 1	(Gitelson et al., 2005)

Table 4.2 (continued)

21	Modified red-edge simple ratio	MSR_RE	((NIR/Red-edge) – 1) /$\sqrt{(\mathrm{NIR/Red-edge}) + 1}$	(Cao et al., 2013)
22	Red-edge soil adjusted vegetation index	RESAVI	1.5 * [(NIR − Red-edge)/(NIR + Red-edge + 0.5)]	(Cao et al., 2013)
23	Modified RESAVI	MRESAVI	0.5 * [2 * NIR + 1 − $\sqrt{(2 * \mathrm{NIR} + 1)^2 - 8 * (\mathrm{NIR} - \mathrm{Red-edge})}$]	(Cao et al., 2013)
24	Red-edge optimal soil adjusted vegetation index	REOSAVI	1.16 * (NIR − Red-edge)/(NIR + Red-edge + 0.16)	(Cao et al., 2013)
25	Red-edge re-normalized different vegetation index	RERDVI	(NIR − Red-edge)/$\sqrt{\mathrm{NIR} + \mathrm{Red-edge}}$	(Cao et al., 2013)

Leaf Nitrogen sampling

Whole plant or leaf sampleswere collected at 3 stages: Whole plant samples were collected at V4 stage (May 23, 2019), leaf samples were taken at V6 (May 30, 2019), and just before tassel emergence (VT) (Jun 21, 2019). Six samples were collected from rows 2 and 3 and six samples from rows 10 and 11 (three samples from each row). The most recently matured and fully-collared leaf on individual corn plants were selected for sampling. Samples were placed in a forced-air oven and dried at 65 ^{o}C and weighed before they were ground through a 40-mesh sieve in a Willy Mill and placed in airtight plastic vials. They were again dried and stored in sealed polypropylene vials until analysis and were processed for total N concentration on a Carlo Erba N/C 1500 automated dry combustion analyzer (Carlo Erba, Milan Italy).

Grain yield

Corn grain was harvested with a two-row plot combine for the entire plot length and grain yield was calculated on ton/ha for rows 2 and 3 and rows 10 and 11 of each plot (Figure 4.1). Grain yield was adjusted to a moisture content of 15.5 %.

Statistical Analysis

Feature selection

Due to the availability of a large number of VI's, there is a need to to select ones that optimally can predict crop yield and and tissue N concentration. In this research, the Recursive Feature Elimination (RFE) method (Granitto et al., 2006), was used which is a popular feature selection method used to select predictors from the training data set that are more effective in predicting the independent variable and maximizing model accuracy. Most feature selection methods are able to determine important features. However, before inputing variables into the ML algorithm, it should be considered that only the predictors that significantly improve the prediction should be imported to the models. Thus, in this research the Random Forest RFE (RF-RFE) method was used to select the most appropriate VIs, derived from a Crop-Circle spectral sensor, to estimate corn leaf N concentration and predict grain yield. This method works mainly in three steps including:

1- RFE builds a model and estimates the feature importance by using a training data set.

2- RFE sets the priority of the important features. It takes a subgroup of the selected variables in step 1 and builds models of a given subset size. In each iteration, the ranking of each feature is recalculated. In this step, the repeated cross-validations were implemented within the RFE method.

3- the model performance is evaluated across different subset sizes to derivean optimal list of predictors.

Most importantly, the flexibility of this method in terms of hyperparameters and ability to control what algorithms are utilized makes it an appropriate feature selection model for most of the ML applications. Since we were interested in fitting an appropriate model with a limited number of predictors, the RFE method chooses the optimum number of features, without affecting the model accuracy.

Machine Learning methods

Since the observed grain yield and leaf N concntration data did not follow a normal distribution and VIs and multispectral observation are highly correlated, this type of study is best for performed through nonparametric models; therefore, this project utilized four nonparametric ML models to develop corn leaf N concentration and grain yield prediction models. Four ML methods (Random Forest (RF), Gradient Boosting Model (GBM), XGBoost, and Support Vector Regression (SVR)) were used in this research to find the best model to predict grain yield and leaf N concentration of corn.

With its build-in ensembling capacity, RF (Breiman, 2001) is one of the most versatile ML algorithms, which is used for either regression or classification problems. This technique is robust to correlated predictors, such as spectral bands or different VIs, and is used to solve both supervised and unsupervised ML problems. Random Forest models can provide variable interaction detection, nonlinear relationship detection, handling of missing values, and modeling of local effects.

However, one of the disadvantages of the Random Forest ML is that it tends to return erratic prediction in the case where observations are out of range of training data. Therefore, to come up with a robust prediction model, it is required to not utilize out of range values in the validation data set.

Similarly, the GBM (Friedman, 2002; Natekin and Knoll, 2013) is an RF model in that it runs numerous decision trees and uses these trees to compute an average. The GBM is a sequential modeling approach, though, the value added by this model is that each step learns from the previous step, whereas, with a Random Forest model all trees are run separately and they do not learn from each other. In the GBM model, on the other hand, high residuals from one step are upweighted when they get fed into the next step and, as a result, each tree can learn from previous trees.

The other model used in this research was the XGBoost (T. Chen and Guestrin, 2016), which stands for Extreme Gradient Boosting. This model is an optimized distributed algorithm that has recently been dominating applied ML competitions for structured or tabular data. It has a fast, and accurate performance on regression and classification. It also can prevent overfitting by adding a regularization term (Mo et al., 2019).

The regression model of Support Vector Machine (SVM) (Vapnik, 1998; Vapnik et al., 1997), called Support Vector Regression (SVR), is a supervised-learning approach and useful tool in real-value function estimation. This model is used to model linear and nonlinear relationships and the essential data points are chosen to solve the regression function. Support vector regression is a classifier used for predicting discrete categorical data whereas SVR is a regressor that is applied for predicting continuous variables. One of the advantages of SVR is that it is robust to the outliers and generalizing capability with high prediction accuracy (Awad and Khanna, 2015).

Two groups of features were used to train the ML models including spectral bands and VIs. Therefore, two strategies were used to train the N and yield prediction models. In the first method, ML models were trained based on spectral bands. In this step, four features including Red, NIR, RedEdge, and growth stages were used as inputs in the training models. In the second strategy, the VIs selected by the RF-RFE method were used as input features to the training models. Each data set was randomly divided into training and test set, such that the training set contained 75% of the samples.

Results and Discussion

The Randon Forest recursive feature elimination method selected SCCCI and RERVI, which are two commonly red-edge-based VIs (Barzin et al., 2020; Fox, 2015; T. B. Raper and Varco, 2015; Sumner, 2019) as predictors for estimating leaf N % and SCCCI, CIRE, RERVI, SAVI, and NDVI were chosen for predicting corn grain yield. Cao et al. (2013) and Erdle et al. (2011) found that RERVI was the most influential and temporally stable index for estimating N concentration. Besides, growth stages have a considerable effect on performing VIs for estimating plants biophysical parameters (Hatfield and Prueger, 2010; F. Li et al., 2010, 2012; Miao et al., 2009; Yu et al., 2013) and in this study, growth stage was imported to the models as an input for both leaf N estimation and yield prediction.

Regression analysis

The relationship between the SCCCI and RERVI with leaf N concentration and grain yield was compared in Figure 4.3. In this figure, the data was organized concerning three phenological stages including V4 (purple), V6 (orange), and VT (pink). The density plots illustrated the distribution of leaf N% as a response and SCCCI and RERVI as predictors that have a huge impact on estimating leaf N%. The density plots for leaf N% (Figure 4.3, upper left) demonstrated that tissue N does not follow the normal distribution at any of stages. It has a similar probability distribution pattern for VIs at different phenological stages. For example, in Figure 4.3, the SCCCI almost followed the multimodal distribution. Scatter plots (lower panel) and associated correlations (upper panel) illustrated the relationship between the response variables (yield and N) and two independent variables (RERVI and SCCCI) for the three growth stages. Regarding the SCCCI index, the correlation coefficients between this index and N were -0.35, 0.90, and 0.92 at V4, V6, and VT stages, respectively. Correspondingly, the correlation coefficients between the RERVI and N were 0.48, 0.74, and 0.95 at the V4, V6, and VT stages, respectively. The correlation coefficients between grain yield and SCCCI were -0.44, 0.72, and 0.93 and between yield and RERVI were 0.67, 079, and 0.97 for V4, V6, and VT, respectively. The correlation coefficients between SCCCI and others was negative at V4 stages and it maybe happen because of the corn leaves were small at V4 stage. Four histograms were associated with each phenological stage and each variable was shown at the bottom of this figure. In the right panel, the boxplots showed the variation of each variable at different stages. For example, as corn development progressed the average of leaf N% increased (around %0.25) from V4 to V6, then at the VT stage, it decreased to almost the same level as the V4 stage. There is a variation in RERVI as phenological stages change from V6 to VT , however, unlike the SCCCI, there is a trend according to each stage. As illustrated in the scatterplot between RERVI and leaf N%, the variation in RERVI for each growth stage is independent of the other stages; therefore, separate regression models were fitted for each stage individually (Figure 4.4). Shen et al. (2014) reported that RERVI had a consistent better correlation with plant N uptake across different growth stages The results of the regression analysis are illustrated in Table 4.3 and the fitted lines are shown in Fig 4.4.

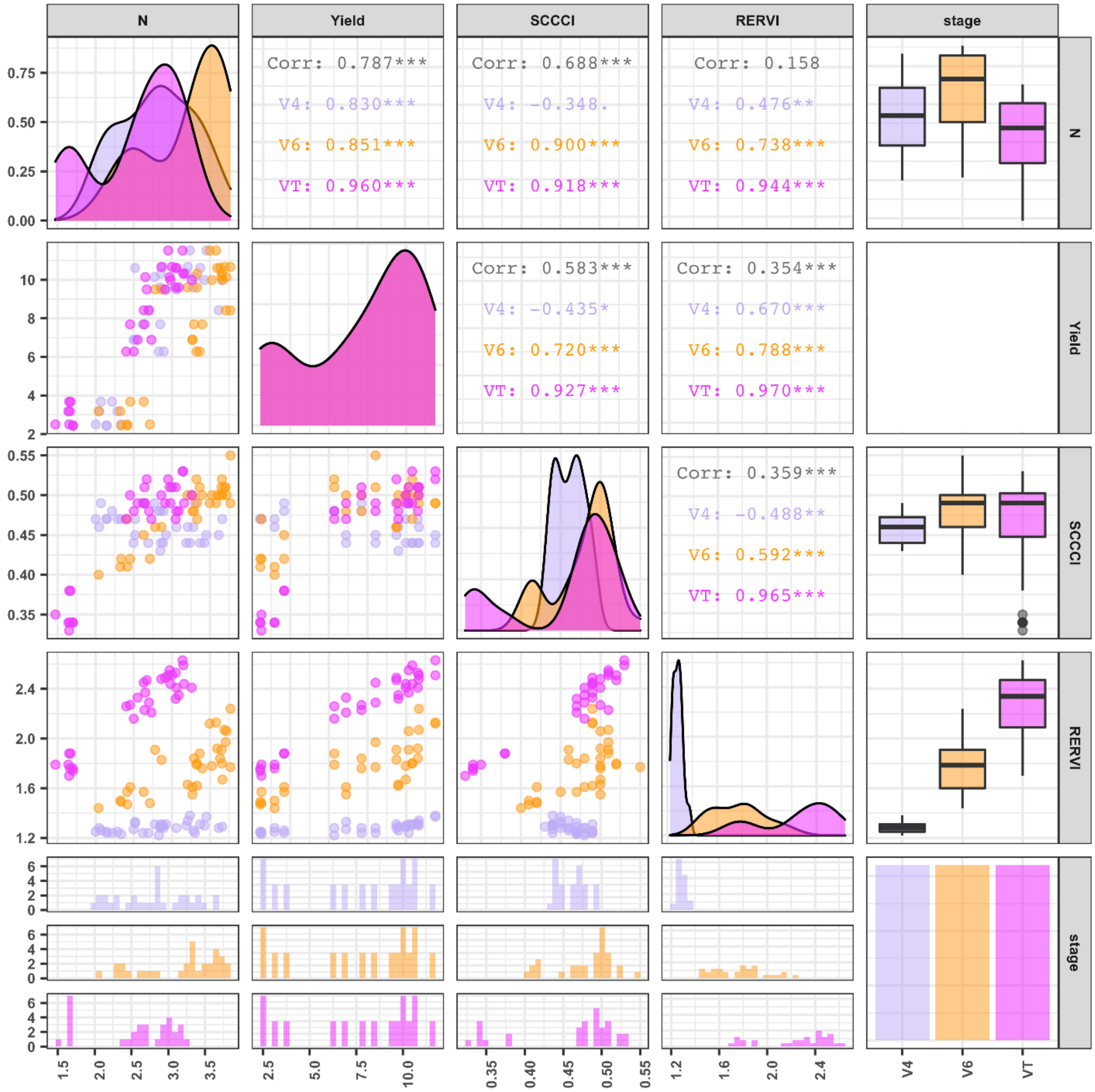


Figure 4.3 Exploratory data analysis for N, yield, RERVI, and SCCCI at different phenological stages.

Table 4.3 Linear regression results of leaf N (%) and RERVI

Stage	Residual Standard Error	R^2	p-value
V4	0.44	0.23	0.006 **
V6	0.36	0.54	1.45 ×10-6 ***
VT	0.19	0.89	5.8 × 10-16 ***

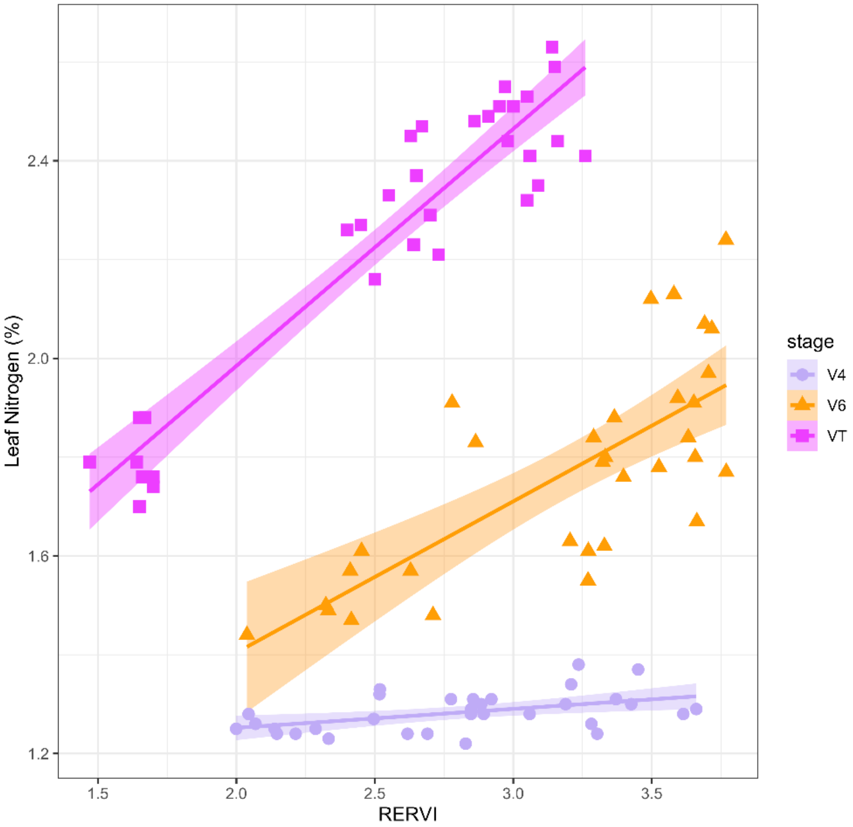


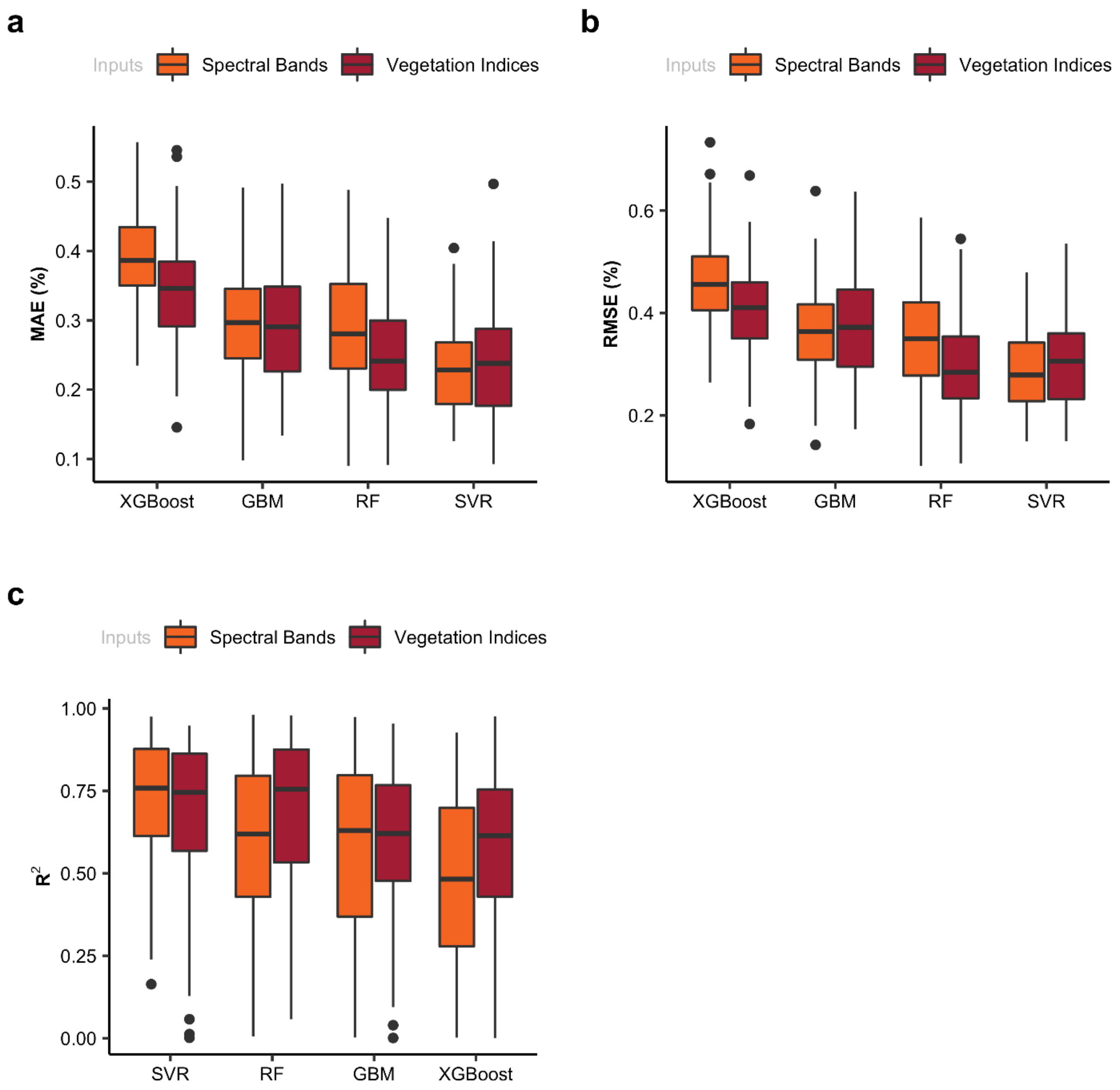
Figure 4.4 Scatter plots of RERVI versus leaf N percent in different corn growth stages

The regression analysis was performed for other VIs to evaluate the relationship between leaf N concentration and VIs. The preliminary results of regression analysis were used to assess the importance of independent variables in leaf N estimation. As illustrated in Table 4.3, RERVI has a statistically significant relationship with leaf N at all three stages (all the p-values were less than 0.05 in 5% confidence interval).

Machine Learning results

Machine learning results for N estimation

The box plots in Figure 4.5 displayed the variation of mean absolute error (a), root mean square error (b), and R-squares (c) resulting from the cross-validation of training models. Two groups of box plots were shown in Figure 4.5: the orange plots showed the cross-validation results derived from the ML model, which were trained by the VIs, and the brown plots illustrated cross-validation results derived from the models, which were trained by the spectral raw data. In this study, four non-parametric ML models were used to estimate the leaf N concentration. As indicated in Figure 4.5, evaluation metrics were different in all models. For example, in the wavelength-based models, the average R^2 resulting from cross-validations in the training data set were 0.61, 0.48, 0.62, and 0.75 for RF, XGBoost, GBM, and SVR models, respectively. Results indicate that the SVR model outperformed the other models in almost all performance measures. In other words, the SVR model was able to predict leaf N% in 75% of the cases using wavelength-based inputs. Additionally, the same ML models were trained by two VIs including SCCCI and RERVI, which were the best predictors among the 25 VIs derived from Crop Circle spectral bands. Comparing the performance of two groups of ML models indicated that the models trained by the VIs have the greatest accuracy in comparison to models trained by the spectral bands in N prediction. The results of this study ranked the ML leaf N% estimation models from the best to the worst, according to the statistical evaluation metrics in the following order: SVR, RF, GBM, and XGBoost. In addition, regarding the performance of the ML models, the same order can be seen for both VI-based and wavelength-based modeling approaches.



2

Figure 4.5 Statistical evaluation metrics resulted from cross-validation in two sets of machine learning models to estimate leaf N content.

The performance of well-trained models was evaluated on the test data (25% of the samples), which had no contribution in training the models. Similar statistical evaluation metrics were used to validate model performance on the test data set (Figure 4.6). As illustrated in figure 4.6, the SVR model had achieved the best performance measures as compared to the other models. As a result, this study illustrated that VIs derived from the Crop Circle sensor can be used as reliable inputs for the SVR model to predict corn leaf N% accurately.

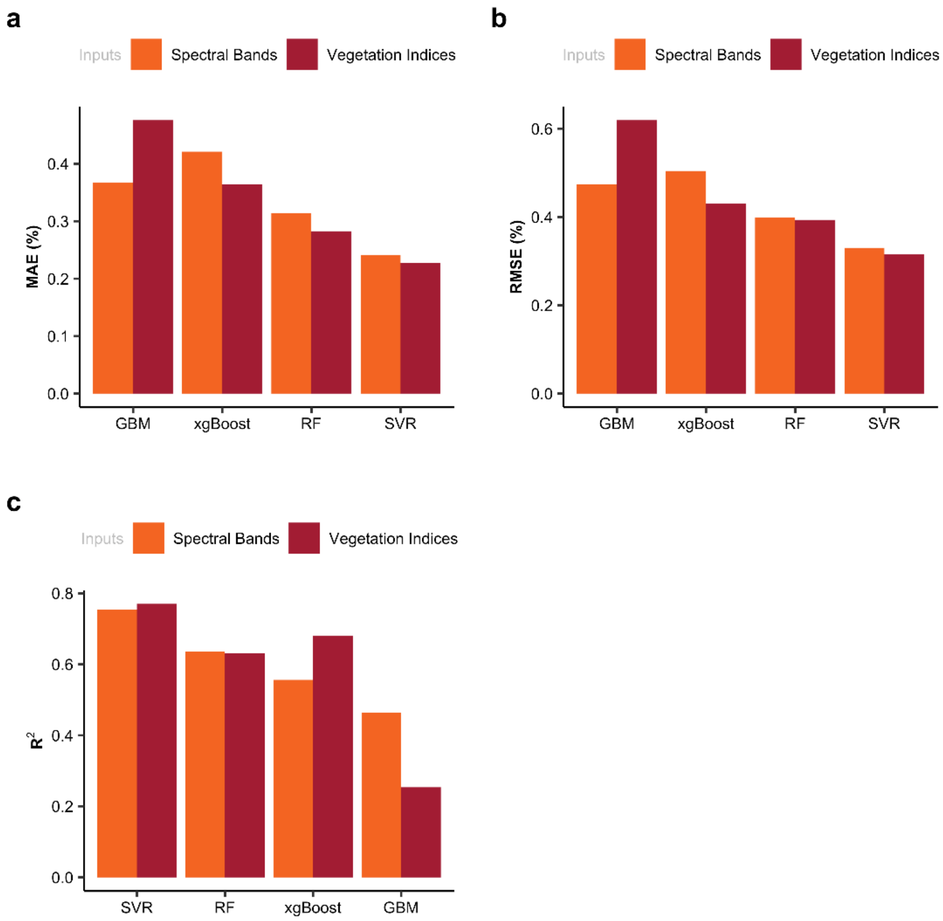


Figure 4.6 Statistical evaluation metrics derived from ML models' performance on the validation (test) data set.

The results also indicate that the VIs-based models achieved slightly better performance measures than reflectance-based models. In both modeling strategies, the SVR model had significantly enhanced performance in comparison to the other models. While it may be more difficult to interpret the non-parametric models such as RF, in addition to the SVR model, it was found in this study to better predict leaf N concentration.

Machine learning results for Yield estimation

The variation in mean absolute error (a), root mean square error (b), and R-squared (c) resulted from the cross-validation of training models demonstrated in Figure 4.7. Regarding the wavelength-based models, the average R^2 values resulted from cross-validations in the training data set were 0.73, 0.76, 0.74, and 0.72 for RF, XGBoost, GBM, and SVR models, respectively. Results indicated that there is no statistically significant difference between models. The results of cross-validation showed that the first quartile, median, and third quartile of the ML models were almost the same, indicating that there was no major difference between these models (Figure 4.7). It is notable to point out that the spectral-based models did not accuraretly predict grain yield.

Conclusions

This study was conducted to evaluate the reliability of a handheld Crop Circle ACS-430 to estimate corn leaf N% and predict grain yield corn using ML algorithms. The Random Forest recursive feature elimination method was applied and determined that SCCCI and RERVI were the most effective VIs to estimate corn leaf N concentration. This method also determined that SCCCI, CI_{RE}, RERVI, SAVI, and NDVI were the most efficient VIs in predicting corn grain yield.

Furthermore, between the four ML models utilized in this research, SVR achieved accurate results for leaf N% estimation using either the spectral bands or VIs as the model inputs. While VIs can predict grain yield well, spectral bands were not reliable for predicting grain yield. More studies are needed to further evaluate this sensor with larger dataset collection across differing environments.

CHAPTER V

CONCLUSIONS

This study used spectral wavelengths and calculated VIs derived from a red-edge multispectral camera mounted on a UAV to develop regression-based and tree-based learning models at different growth stages. These models were developed in order to estimates corn leaf N concentration and predict grain yield. The effect of the input variables was found to vary with phenological stages. OSAVI and SCCCI were the single dominant variables in the yield prediction models at V3 and V4-5 stages, respectively. In general, SCCCI was a VI that contributed to most of the models to predict yield, suggesting the importance of red-edge-based VIs in yield estimation. The applied Gradient Boosting Machines (GBM) for yield prediction resulted in the greatest coefficient of determination (R^2) of 0.97 and 0.95 at V10 and VT stages, respectively. Likewise, the greatest R^2 values were obtained at the same stages using regression-based models. As corn development progressed, the accuracy of both regression-based and tree-based models would be increased. GBM was the most accurate model among the eight ML models used to estimate the leaf N content in this study. The Random Forest feature selection method showed that SCCCI and NDRE were the most efficient VIs to estimate the leaf N content using a red-edge Multispectral camera mounted on the UAV. However, applying the Crop Circle ACS-430 illustrated that SCCCI and RERVI were the most effective VIs to estimate corn leaf N concentration which showed the importance of including the red-edge band.. This method also demonstrated that the SCCCI, RERVI, CIRE, SAVI, and NDVI idices were the most important ones in predicting corn grain yield. Moreover, among the four ML models used for leaf N% estimation by the Crop Circle, Support Vector Regression (SVR) attained the most accurate results. More data collection is required to further evaluate this sensor. The methodology used in this research can be extended to predict yield for other crops or in other regions as well, where yield prediction is mainly reliant on weather and climatic conditions.

REFERENCES

Abdel-rahman, E. M., Ahmed, F. B., Ismail, R. (2013). Random Forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data, *1161*. https://doi.org/10.1080/01431161.2012.713142

Akbarzadeh Baghban, A., Younespour, S., Jambarsang, S., Yousefi, M., Zayeri, F., Azizi Jalilian, F. (2013). How to test normality distribution for a variable: a real example and a simulation study, *4*(1), 2008–4978.

Al-Abbas, A. H., Barr, R., Hall, J. D., Crane, F. L., V, M. F. (1972). Spectra of Normal and Nutrient-Deficient Maize Leaves. *Agronomy Journal*, *66*(1), 16–20.

Andrews, M., Raven, J. A., Lea, P. J. (2013). Do plants need nitrate ? The mechanisms by which nitrogen form affects plants, *163*(3), 174–199. https://doi.org/10.1111/aab.12045

Aquino, A., Millan, B., Diago, M., Tardaguila, J. (2018). Automated early yield prediction in vineyards from on-the-go image acquisition Automated early yield prediction in vineyards from on-the-go image acquisition. *Computers and Electronics in Agriculture*, *144*(January), 26–36. https://doi.org/10.1016/j.compag.2017.11.026

Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., Kolb, F. L. (2015). Genomic Selection for Predicting Fusarium Head Blight Resistance in a Wheat Breeding Program, (November). https://doi.org/10.3835/plantgenome2015.01.0003

Awad, M., Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer Nature.

Báez-González, A. D., Chen, P. Y., Tiscareño-López, M., Srinivasan, R. (2002). Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. *Crop Science*, *42*(6), 1943–1949. https://doi.org/10.2135/cropsci2002.1943

Baez-gonzalez, A. D., Kiniry, J. R., Maas, S. J., L, M. T., C, J. M., Mendoza, J. L., … Manjarrez, J. R. (2005). Large-Area Maize Yield Forecasting Using Leaf Area Index Based Yield Model, 418–425.

Bannari, A., Asalhi, H., Teillet, P. M. (2002). Transformed difference vegetation index (TDVI) for vegetation cover mapping. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, *5*(C), 3053–3055. https://doi.org/10.1109/igarss.2002.1026867

Baret, F., Houlès, V., & Guerif, M. (2007). Quantification of plant stress using remote sensing observations and crop models : the case of nitrogen management. *Ournal of Experimental Botany*, *58*(4), 869–880. https://doi.org/10.1093/jxb/erl231

Barzin, R., Pathak, R., Lotfi, H., Varco, J., Bora, G. C. (2020). Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn, 1–21. https://doi.org/10.3390/rs12152392

Barzin, R., Shirvani, A., Lotfi, H. (2017a). Estimation of daily average downward shortwave radiation from MODIS data using principal components regression method: Fars province case study. *International Agrophysics*, *31*(1), 23–34. https://doi.org/10.1515/intag-2016-0035

Barzin, R., Shirvani, A., Lotfi, H. (2017b). Estimation of daily average downward shortwave radiation from MODIS data using principal components regression method: Fars province case study. *International Agrophysics*, *31*(1), 23–34. https://doi.org/10.1515/intag-2016-0035

Bender, B. R. R., Haegele, J. W., Ruffo, M. L., Below, F. E. (2012). Modern Corn Hybrids ' Nutrient Uptake Patterns, 7–10.

Berger, K., Verrelst, J., Féret, J., Wang, Z., Wocher, M., Strathmann, M., … Hank, T. (2020). Crop nitrogen monitoring : Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, *242*(March), 111758. https://doi.org/10.1016/j.rse.2020.111758

Bojović, B., Marković, A. (2009). Correlation between nitrogen and chlorophyll content in wheat (Triticum aestivum L.). *Kragujevac Journal of Science*, *31*(31), 69–74.

Bondi, E., Salvaggio, C., Montanaro, M., Gerace, A. D. (2016). Calibration of UAS imagery inside and outside of shadows for improved vegetation index computation. *In Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, 9866*, 98660J. https://doi.org/10.1117/12.2227214

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1007/978-3-662-56776-0_10

Brinkhoff, J., Dunn, B. W., Robson, A. J., Dunn, T. S., Dehaan, R. L. (2019). Modeling Mid-Season Rice Nitrogen Uptake Using Multispectral Satellite Data, 1–22. https://doi.org/10.3390/rs11151837

Broge, N. H., Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, *76*(2000), 156–172.

Bronson, K. F., Booker, J. D., Keeling, J. W., Boman, R. K., Wheeler, T. A., Lascano, R. J., Nichols, R. L. (2005). Cotton canopy reflectance at landscape scale as affected by nitrogen fertilization. *Agronomy Journal*, *97*(3), 654–660. https://doi.org/10.2134/agronj2004.0093

Butchee, K. S., May, J., Arnall, B. (2011). Sensor Based Nitrogen Management Reduced Nitrogen and Maintained Yield. *Cm*, *10*(1), 0. https://doi.org/10.1094/cm-2011-0725-01-rs

Cao, Q., Miao, Y., Li, F., Gao, X. (2017). Developing a new Crop Circle active canopy sensor-based precision nitrogen management strategy for winter. *Precision Agriculture*, *18*(1), 2–18. https://doi.org/10.1007/s11119-016-9456-7

Cao, Q., Miao, Y., Shen, J., Yuan, F., Cheng. (2018). Evaluating Two Crop Circle Active Canopy Sensors for In-Season Diagnosis of Winter Wheat. *Agronomy, 8*(10), 201. https://doi.org/10.3390/agronomy8100201

Cao, Q., Miao, Y., Wang, H., Huang, S., Cheng, S., Khosla, R., Jiang, R. (2013). Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor. *Field Crops Research, 154*, 133–144. https://doi.org/10.1016/j.fcr.2013.08.005

Cassman, K. G., Dobermann, A., Walters, D. T., Yang, H. (2003). Meeting Cereal Demand While Protecting Natural Resources and Improving Environmental Quality. *Annual Review of Environment and Resources*, *28*(1), 315–358. https://doi.org/10.1146/annurev.energy.28.040202.122858

Chang, J., Clay, D. E., Dalsted, K., Clay, S., O'Neill, M. (2003). Corn (Zea mays L.) Yield Prediction Using Multispectral and Multidate Reflectance. *Agronomy Journal*, *95*(6), 1447–1453. https://doi.org/10.2134/agronj2003.1447

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, *22*(3), 229–242. https://doi.org/10.1080/07038992.1996.10855178

Chen, T., Guestrin, C. (2016). XGBoost : A Scalable Tree Boosting System. In *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chlingaryan, A., Sukkarieh, S., Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*(November 2017), 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Clevers, J. G., Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation, 23*, 344–351.

Dobermann, A., Witt, C., Dawe, D., Abdulrachman, S., Gines, H. C., Nagarajan, R., … Adviento, M. A. A. (2002). Site-specific nutrient management for intensive rice cropping systems in Asia. *Field Crops Research, 74*(1), 37–66. https://doi.org/10.1016/S0378-4290(01)00197-6

Doraiswamy, P. C., Moulin, S., Cook, P. W., Stern, A. (2003). Crop Yield Assessment from Remote Sensing, *69*(6), 665–674.

Erdle, K., Mistele, B., Schmidhalter, U. (2011). Field Crops Research Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. *Field Crops Research*, *124*(1), 74–84. https://doi.org/10.1016/j.fcr.2011.06.007

Esri, Redlands, Ca, U. (2019). Using ArcMap.

Feng, W., Wu, Y., He, L., Ren, X., Wang, Y., Hou, G., … Guo, T. (2019). An optimized non-linear vegetation index for estimating leaf area index in winter wheat. *Precision Agriculture*, *20*(6), 1157–1176. https://doi.org/10.1007/s11119-019-09648-8

Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., … Székely, B. (2010). Crop yield estimation by satellite remote sensing, *1161*. https://doi.org/10.1080/01431160410001698870

Fox, A. A. A. (2015). *An Integrated Approach for Predicting Nitrogen Status in Early Cotton and Corn*.

Fraser, R. H., Latifovic, R. (2005). Mapping insect-induced tree defoliation and mortality using coarse spatial resolution satellite imagery. *International Journal of Remote Sensing*, *26*(1), 193–200. https://doi.org/10.1080/01431160410001716923

Fridgen, J. L., Varco, J. J. (2004). on Nitrogen and Potassium Availability, 63–69.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Funk, C., Budde, M. E. (2009). Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, *113*(1), 115–125. https://doi.org/10.1016/j.rse.2008.08.015

Gallo, B. C., Demattê, J. A. M., Rizzo, R., Safanelli, J. L., Mendes, W. de S., Lepsch, I. F., … Lacerda, M. P. C. (2018). Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. *Remote Sensing*, *10*(10), 1571. https://doi.org/10.3390/rs10101571

Gardner, M. W., Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Gautam, R. K., Panigrahi, S. (2007). Leaf nitrogen determination of corn plant using aerial images and artificial neural networks. *Canadian Biosystems Engineering / Le Genie Des Biosystems Au Canada*, *49*, 1–9.

Ghasemi, A., Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, *161*(2), 165–173.

Gitelson, A. A., Gritz, Y., Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., Kaufman, Y. J., Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A. A., Kaufman, Y. J., Stark, R., Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87.

Gitelson, A. A., Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology*, *22*(3), 247–252.

Gitelson, A. A., Merzlyak, M. N. (1998a). Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, *18*(12), 2691–2697. https://doi.org/10.1080/014311697217558

Gitelson, A. A., Merzlyak, M. N. (1998b). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, *22*(5), 689–692.

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, *32*(8), 1–4. https://doi.org/10.1029/2005GL022688

Gnädinger, F., Schmidhalter, U. (2017). Digital counts of maize plants by Unmanned Aerial Vehicles (UAVs). *Remote Sensing*, *9*(6), 544. https://doi.org/10.3390/rs9060544

Goldstein, A., Fink, L., Meitin, A. (2018). Applying machine learning on sensor data for irrigation recommendations : revealing the agronomist ' s tacit knowledge. *Precision Agriculture*, *19*(3), 421–444. https://doi.org/10.1007/s11119-017-9527-4

Gong, P., Pu, R., Biging, G. S., Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6 PART I), 1355–1362. https://doi.org/10.1109/TGRS.2003.812910

González-audícana, M., Saleta, J. L., Catalán, R. G., García, R. (2004). Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA Mergers Based on Wavelet Decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(6), 1291–1299.

GopalaPillai, S., Tian, L. (1999). In-field variability detection and yield prediction in corn using digital aerial imaging, *42*(6), 1911–1920.

Granitto, P. M., Furlanello, C., Biasioli, F., Gasperi, F. (2006). Recursive feature elimination with Random Forest for PTR-MS analysis of agroindustrial products, *83*, 83–90. https://doi.org/10.1016/j.chemolab.2006.01.007

Gutiérrez, S., Marı´a, Diago, P., Ferna´ndez-Novales, J., Tardaguila, J. (2018). Vineyard water status assessment using on- the-go thermal imaging and machine learning, 1–18. https://doi.org/10.6084/m9.figshare.5808720.Funding

Hastie, T., Tibshirani, R., Jerome, F. (2004). *The elements of statistical learning: data mining, inference and prediction*. *Springer Science & Business Media*.

Hatfield, J. L., Prueger, J. H. (2010). Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sensing*, *2*(2), 562–578. https://doi.org/10.3390/rs2020562

He, M., Kimball, J. S., Maneta, M. P., Maxwell, B. D. (2018). Regional Crop Gross Primary Productivity and Yield Estimation Using Fused Landsat-MODIS Data. *Remote Sensing*, *10*(3), 372. https://doi.org/10.3390/rs10030372

Hoerl, A. E., Kennard, R. W. (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, *12*(1), 69–82. https://doi.org/10.1080/00401706.1970.10488635

Hogrefe, K. R., Patil, V. P., Ruthrauff, D. R., Meixell, B. W., Budde, M. E., Hupp, J. W., Ward, D. H. (2017). Normalized difference vegetation index as an estimator for abundance and quality of Avian Herbivore Forage in Arctic Alaska. *Remote Sensing*, *9*(12), 1234. https://doi.org/10.3390/rs9121234

Hunt, Daughtry, C. S. T., Eitel, J. U. H., Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, *103*(4), 1090–1099. https://doi.org/10.2134/agronj2010.0395

Iizuka, K., Itoh, M., Shiodera, S., Matsubara, T., Dohar, M., Watanabe, K. (2018). Advantages of unmanned aerial vehicle (UAV) photogrammetry for landscape analysis compared with satellite data: A case study of postmining sites in Indonesia. *Cogent Geoscience*, *4*(1), 1498180. https://doi.org/10.1080/23312041.2018.1498180

Jackson, R. D., Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, *11*(3–4), 185–200. https://doi.org/10.1016/S0167-5877(05)80004-2

Janitza, S., Tutz, G., Boulesteix, A. L. (2016). Random Forest for ordinal responses: Prediction and variable selection. *Computational Statistics and Data Analysis*, *96*, 57–73. https://doi.org/10.1016/j.csda.2015.10.005

Johnson, D. M. (2014). Remote Sensing of Environment An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, *141*, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Kakkar, A. (2017). DigitalCommons @ USU Nitrogen Availability and Use Efficiency in Corn Treated with Contrasting Nitrogen Sources.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, *153*(April), 213–225. https://doi.org/10.1016/j.compag.2018.07.016

Kim, S., Dale, B. E. (2008). Effects of nitrogen fertilizer application on greenhouse gas emissions and economics of corn production. *Environmental Science and Technology*, *42*(16), 6028–6033. https://doi.org/10.1021/es800630d

Kizil, Ü., Genç, L., İnalpulat, M., Şapolyo, D., Mİrİk, M. (2012). Lettuce (Lactuca sativa L .) yield prediction under water stress using artificial neural network (ANN) model and vegetation indices. *Žemdirbystė= Agriculture*, *99*(4), 409–418.

Kox, M. A. R., Lüke, C., Fritz, C., van den Elzen, E., van Alen, T., Op den Camp, H. J. M., … Ettwig, K. F. (2016). Effects of nitrogen fertilization on diazotrophic activity of microorganisms associated with Sphagnum magellanicum. *Plant and Soil*, *406*(1–2), 83–100. https://doi.org/10.1007/s11104-016-2851-z

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). New York: McGraw-Hill Irwin.

Li, F., , Mistele, B., Hu, Y., Yue, X., Yue, S., Miao, Y., Chen, X., Cui, Z., Meng, Q. and Schmidhalter, U. (2012). Remotely estimating aerial N status of phenologically differing winter wheat cultivars grown in contrasting climatic and geographic zones in China and Germany. *Field Crops Research*, *138*, 21–32. https://doi.org/10.1016/j.fcr.2012.09.002

Li, F., Miao, Y., Hennig, S. D., Gnyp, M. L., Chen, X., Jia, L., Bareth, G. (2010). Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. *Precision Agriculture*, *11*(4), 335–357. https://doi.org/10.1007/s11119-010-9165-6

Li, Z., Wang, J., Tang, H., Huang, C., Yang, F. (2016). Predicting Grassland Leaf Area Index in the Meadow Steppes of Northern China : A Comparative Study of Regression Approaches and Hybrid Geostatistical Methods. *Remote Sensing*, *8*(8), 632. https://doi.org/10.3390/rs8080632

Liaw, A., Wiener, M. (2002). Classification and Regression by randomForest. R News 2, *3*(December 2002), 18–22.

Lillesand, T., Kiefer, R., Chipman, J. (2015). Remote sensing and image interpretation. 6th edition. *John Wiley & Sons*.

Lobell, D. B., Ortiz-monasterio, J. I., Asner, G. P., Naylor, R. L., Falcon, W. P. (2005). Combining Field Surveys , Remote Sensing , and Regression Trees to Understand Yield Variations in an Irrigated Wheat Landscape, 241–249. https://doi.org/10.2134/agronj2005.0241a

Louhaichi, M., Borman, M. M., Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, *16*(1), 65–70. https://doi.org/10.1080/10106040108542184

Magri, A., Es, H. M. V. A. N., Glos, M. A., Cox, W. J. (2005). Soil Test , Aerial Image and Yield Data as Inputs for Site-specific Fertility and Hybrid Management Under, 87–110.

Matsushita, B., Yang, W., Chen, J., Onda, Y., Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: A case study in high-density cypress forest. *Sensors*, *7*(11), 2636–2651. https://doi.org/10.3390/s7112636

McGuire, S. (2013). WHO , World Food Programme , and International Fund for Agricultural Development . 2012 . The State of Food Insecurity in the World 2012 . Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition. Rome, FAO, 126–127.

Miao, Y., Mulla, D. J., Randall, G. W., Vetsch, J. A., Vintila, R. (2009). Combining chlorophyll meter readings and high spatial resolution remote sensing images for in-season site-specific nitrogen management of corn, 45–62. https://doi.org/10.1007/s11119-008-9091-z

Miyoshi, G. T., Arruda, S., Osco, L. P., Junior, M., Gonçalves, D. N., Imai, N. N. (2020). A Novel Deep Learning Method to Identify Single Tree Species in UAV-Based Hyperspectral Images. *Remote Sensing*, *12*(8), 1294.

Mo, H., Sun, H., Liu, J., Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. *Energy & Buildings*, *205*, 109564. https://doi.org/10.1016/j.enbuild.2019.109564

Morales, A., Nielsen, R., Camberato, J. (2019). Effects of Removing Background Soil and Shadow Reflectance Pixels from RGB and NIR-based Vegetative Index Maps, (Vi), 91.

Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., … Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, *152*, 104–116. https://doi.org/10.1016/j.biosystemseng.2016.04.018

Natekin, A., Knoll, A. (2013). Gradient boosting machines , a tutorial, *7*(December). https://doi.org/10.3389/fnbot.2013.00021

Nurunnabi, A., West, G., Belton, D. (2015). Outlier detection and robust normal-curvature estimation in mobile laser scanning 3D point cloud data. *Pattern Recognition*, *48*(4), 1404–1419.

Oppelt, N., Mauser, W. (2010). Hyperspectral monitoring of physiological parameters of wheat during a vegetation period using AVIS data, *1161*. https://doi.org/10.1080/0143116031000115300

Osco, L. P., Junior, M., Paula, A., Ramos, M., Elis, D., Furuya, G., … Teodoro, P. E. (2020). Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques.

Pathak, R., Barzin, R., Bora, G. C. (2018). Data-driven precision agricultural applications using field sensors and Unmanned Aerial Vehicle. *International Journal of Precision Agricultural Aviation*, *1*(1), 19–23. https://doi.org/10.33440/j.ijpaa.20180101.0004

Pedregosa, F., Weiss, R., Brucher, M. (2011). Scikit-learn: Machine Learning in Python.

QGIS Development Team. (2020). QGIS Geographic Information System. Retrieved from http://qgis.osgeo.org/

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., Sorooshian, S. (1994). A modify soil adjust vegetation index. *Remote Sensing of Environment*, *126*, 119–126.

Qin, Z., Myers, D. B., Ransom, C. J., Kitchen, N. R., Liang, S. Z., Camberato, J. J., … Shanahan, J. F. (2018). Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. *Agronomy Journal*, *110*(6), 2596–2607. https://doi.org/10.2134/agronj2018.03.0222

R Core Team: Vienna, A. A language and environment for statistical computing. R Foundation for Statistical Computing (2019).

Raper, T. B. (2011). *Template Created By : James Nail 2010 EFFECTIVENESS OF CROP REFLECTANCE SENSORS ON DETECTION OF COTTON (Gossypium hirsutum L .) GROWTH AND NITROGEN STATUS By Submitted to the Faculty of Mississippi State University in Partial Fulfillment of the Requireme*.

Raper, T. B., Varco, J. J. (2015). Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. *Precision Agriculture*, *16*(1), 62–76. https://doi.org/10.1007/s11119-014-9383-4

Raper, Tyson B., Varco, J. J., Hubbard, K. J. (2013). Canopy-Based Normalized Difference Vegetation Index Sensors for Monitoring Cotton Nitrogen Status. https://doi.org/10.2134/agronj2013.0080

Rattanakaew, T. (2015). *Utilization of canopy reflectance to predict yield response of corn and cotton to varying nitrogen rates*. Starkville, MS, Mississippi State University.

Razali, N. M., Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33. https://doi.org/doi:10.1515/bile-2015-0008

Reyniers, M., Vrindts, E., De Baerdemaeker, J. (2006). Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat. *European Journal of Agronomy*, *24*(2), 87–94. https://doi.org/10.1016/j.eja.2005.05.002

Rondeaux, G., Steven, M., Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Roujean, J. L., Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, *51*(3), 375–384. https://doi.org/10.1016/0034-4257(94)00114-3

Rouse, J. W., Hass, R. H., Schell, J. A., Deering, D. W., Harlan, J. C. (1974). Monitoring the Vernal Advancement and Retrogradation (GreenWave Effect) of Natural Vegetation [Great Plains Corridor].

Ruwaimana, M., Satyanarayana, B., Otero, V., Muslim, A. M., Muhammad Syafiq, A., Ibrahim, S., … Dahdouh-Guebas, F. (2018). The advantages of using drones over space-borne imagery in the mapping of mangrove forests. *PLoS ONE*, *13*(7), 1–22. https://doi.org/10.1371/journal.pone.0200288

Sader, S. A., Winne, J. C. (1992). RGB-NDVI colour composites for visualizing forest change dynamics. *International Journal of Remote Sensing*, *13*(16), 3055–3067. https://doi.org/10.1080/01431169208904102

Sakamoto, T., Gitelson, A. A., Arkebauer, T. J. (2014). Near real-time prediction of U.S. corn yields based on time-series MODIS data. *Remote Sensing of Environment*, *147*, 219–231. https://doi.org/10.1016/j.rse.2014.03.008

Santamar, L. (2017). Modeling Biomass Production in Seasonal Wetlands Using MODIS NDVI Land Surface Phenology, 1–18. https://doi.org/10.3390/rs9040392

Sawyer, J. Nutrient Deficiencies and Application Injuries in Field Crops (2004).

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

Schubert, E., Kriegel, H. (2014). Generalized Outlier Detection with Flexible Kernel Density Estimates. In *Proceedings of the 2014 SIAM International Conference on Data Mining. pp. 542-550*. Society for Industrial and Applied Mathematics.

Senay, G. B., Ward, A. D., Lyon, J. G. (1998). Manipulation of High Spatial Resolution Aircraft Remote Sensing Data for Use in Site-Specific Farming.

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., … Major, D. J. (2001). Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. *Agronomy Journal*, *93*(3), 583–589.

Sharma, L. K., Bu, H., Denton, A., Franzen, D. W. (2015). Active-Optical Sensors Using Red NDVI Compared to Red Edge NDVI for Prediction of Corn Grain Yield in North Dakota, U.S.A., 27832–27853. https://doi.org/10.3390/s151127832

Shen, J., Miao, Y., Cao, Q., Wang, H., Yu, W., Hu, S., … Lu, J. (2014). Estimating Rice Nitrogen Status Using Active Canopy Sensor Crop Circle 430 in Northeast China. In *The Third International Conference on Agro-Geoinformatics,* (pp. 1–7). IEEE.

Shi, W., Lu, J., Miao, Y., Cao, Q., Shen, J., Wang, H., … Hu, S. (2013). Evaluating a Crop Circle Active Canopy Sensor-based Precision Nitrogen Management Strategy for Rice in Northeast China.

Shiu, Y. S., Chuang, Y. C. (2019). Yield estimation of paddy rice based on satellite imagery: Comparison of global and local regression models. *Remote Sensing*, *11*(2), 111. https://doi.org/10.3390/rs11020111

Silvestro, P. C., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., … Casa, R. (2017). Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sensing*, *9*(5), 509. https://doi.org/10.3390/rs9050509

Siqueira, R. (2016). *CHARACTERIZING NITROGEN DEFICIENCY OF MAIZE AT EARLY GROWTH STAGES USING FLUORESCENCE MEASUREMENTS Submitted*. Colorado State University.

Smola, A. J., Scholkopf, B. (2004). A tutorial on support vector regression - art%3A10.1023%2FB%3ASTCO.0000035301.49549.88.pdf. *Statistics and Computing*, *14*, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Solari, F., Shanahan, J., Ferguson, R., Schepers, J., Gitelson, A. (2008). Active sensor reflectance measurements of corn nitrogen status and yield potential. *Agronomy Journal*, *100*(3), 571–579. https://doi.org/10.2134/agronj2007.0244

Sripada, R. P., Heiniger, R. W., White, J. G., Meijer, A. D. (2006). Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn, *977*, 968–977. https://doi.org/10.2134/agronj2005.0200

Sripada, R. P., Schmidt, J. P., Dellinger, A. E., Beegle, D. B. (2008). Evaluating multiple indices from a canopy reflectance sensor to estimate corn N requirements. *Agronomy Journal*, *100*(6), 1553–1561. https://doi.org/10.2134/agronj2008.0017

Stefanini, M., Larson, J. A., Lambert, D. M., Yin, X., Boyer, C. N., Scharf, P., … Buschermohle, M. J. (2019). Effects of optical sensing based variable rate nitrogen management on yields, nitrogen use and profitability for cotton. *Precision Agriculture*, *20*(3), 591–610. https://doi.org/10.1007/s11119-018-9599-9

Sumner, Z. T. S. (2019). *Multi-platform comparison of canopy reflectance on corn whole plant and leaf tissue nitrogen status, PhD diss.* Starkville, MS.

Sun, J. (2000). Dynamic Monitoring and Yield Estimation of Crops by Mainly Using the Remote Sensing Technique in China, (May).

Tadesse, A., Kim, H. K., Debela, A. (2015). Calibration of Nitrogen Fertilizer for Quality Protein Maize (Z Ea Mays L .) Based on In-Season Estimated Yield Using a Handheld NDVI Sensor in the Central, *2*(1), 25–32.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine*, *16*(4), 385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

Torres, J. M., Nieto, P. J. G., Alejano, L., Reyes, A. N. (2011). Detection of outliers in gas emissions from urban areas using functional data analysis. *Journal of Hazardous Materials*, *186*(1), 144–149. https://doi.org/10.1016/j.jhazmat.2010.10.091

Trotter, T. F., Frazier, P. S., Trotter, M. G., Lamb, D. W. (2008). Objective biomass assessment using an active plant sensor (Crop Circle), preliminary experiences on a variety of agricultural landscapes. In *Ninth International Conference on Precision Agriculture'*. Denver, Colorado.(Ed. R. Khosla.)(Colorado State University: Fort Collins, CO.).

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150.

Tuia, D., Volpi, M., Verrelst, J., Camps-valls, G., Tuia, D. (2018). *Advances in Kernel Machines for Image Classification and Biophysical Parameter Retrieval*.

Uno, Y., Prasher, S. O., Lacroix, R., Goel, P. K., Karimi, Y., Viau, A., Patel, R. M. (2005). Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Computers and Electronics in Agriculture*, *47*(2), 149–161. https://doi.org/10.1016/j.compag.2004.11.014

USDA. (2019). NASS_USAD.pdf. Retrieved from http://www.nass.usda.gov/Quick_Stats

Van Wittenberghe, S., Verrelst, J., Rivera, J. P., Alonso, L., Moreno, J., Samson, R. (2014). Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. *Journal of Photochemistry and Photobiology B: Biology*, *134*(2014), 37–48. https://doi.org/10.1016/j.jphotobiol.2014.03.010

Vapnik, V. (1998). The support vector method of function estimation. In Nonlinear Modeling. *Springer, Boston, MA.*, 55–85.

Vapnik, V., Golowich, S. E., Ave, M., Hill, M. (1997). Support Vector Method for Function Approximation , Regression Estimation , and Signal Processing ·. *In Advances in Neural Information Processing Systems*, 281–287.

Vellidis, G., Tucker, M., Perry, C., Reckford, D., Butts, C., Henry, H., … Edwards, W. (2013). *A soil moisture sensor-based variable rate irrigation scheduling system*. *Precision Agriculture 2013 - Papers Presented at the 9th European Conference on Precision Agriculture, ECPA 2013*.

Vescovo, L., Gianelle, D. (2008). Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). *Advances in Space Research*, *41*(11), 1764–1772. https://doi.org/10.1016/j.asr.2007.07.043

Vezhnevets, A., Barinova, O. (2007). Avoiding Boosting Overfitting by Removing Confusing Samples, 430–441.

Wahabzada, M., Mahlein, A., Bauckhage, C., Steiner, U. (2016). Plant Phenotyping using Probabilistic Topic Models : Uncovering the Hyperspectral Language of Plants. *Nature Publishing Group*, (September 2015), 1–11. https://doi.org/10.1038/srep22482

Wang, J., Shen, C., Liu, N., Jin, X., Fan, X., Dong, C., Xu, Y. (2017). Non-Destructive Evaluation of the Leaf Nitrogen Concentration by In-Field Visible/Near-Infrared Spectroscopy in Pear Orchards. *Sensors*, *17*(3), 538. https://doi.org/10.3390/s17030538

Wang, T., Shi, J., Letu, H., Ma, Y., Li, X., Zheng, Y. (2019). Detection and Removal of Clouds and Associated Shadows in Satellite Imagery Based on Simulated Radiance Fields. *Journal of Geophysical Research: Atmospheres*, *124*(13), 7207–7225. https://doi.org/10.1029/2018JD029960

Weng, H., Lv, J., Cen, H., He, M., Zeng, Y., Hua, S., … He, Y. (2018). Hyperspectral reflectance imaging combined with carbohydrate metabolism analysis for diagnosis of citrus Huanglongbing in different seasons and cultivars. *Sensors & Actuators: B. Chemical*, *275*(August), 50–60. https://doi.org/10.1016/j.snb.2018.08.020

West, P. C., Gibbs, H. K., Monfreda, C., Wagner, J., Barford, C. C., Carpenter, S. R. (2010). Trading carbon for food : Global comparison of carbon stocks vs . crop yields on agricultural land, *107*(46), 19645–19648. https://doi.org/10.1073/pnas.1011078107

Wu, W. (2014). The Generalized Difference Vegetation Index (GDVI) for dryland characterization. *Remote Sensing*, *6*(2), 1211–1233. https://doi.org/10.3390/rs6021211

Xing, L., Pittman, J. J., Inostroza, L., Butler, T. J., Munoz, P. (2018). Improving Predictability of Multisensor Data with Nonlinear Statistical Methodologies, (april). https://doi.org/10.2135/cropsci2017.09.0537

Yahya, W. B., Olaniran, O. (2014). On Bayesian Conjugate Normal Linear Regression and Ordinary Least Square Regression Methods : A Monte Carlo Study On Bayesian Conjugate Normal Linear Regression and Ordinary Least Square Regression Methods : A Monte Carlo Study, (December 2016).

Yang, C., Anderson, G. L. (2000). Mapping Grain Sorghum Yield Variability Using Airborne Digital Videography.

Yang, P., Tan, G. X., Zha, Y., Shibasaki, R. (2004). Integrating remotely sensed data with an ecosystem model to estimate crop yield in north China. In *Proceedings of XXth ISPRS Congress Proceedings Commission VII, WG VII/2* (pp. 150–156). Istanbul, Turkey.

Yao, Y., Miao, Y., Huang, S., Gao, L., Ma, X., Zhao, G., … Zhu, H. (2012). Active canopy sensor-based precision N management strategy for rice. *Agronomy for Sustainable Development*, *32*(4), 925–933. https://doi.org/10.1007/s13593-012-0094-9

Ye, J., Chow, J., Chen, J., Zheng, Z. (2009). Stochastic Gradient Boosted Distributed Decision Trees, 2061–2064.

Yoder, B. J., Pettigrew-Crosby, R. E. (1995). Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, *53*(3), 199–211. https://doi.org/10.1016/0034-4257(95)00135-N

Yu, K., Li, F., Gnyp, M. L., Miao, Y., Bareth, G., Chen, X. (2013). Remotely detecting canopy nitrogen concentration and uptake of paddy rice in the Northeast China Plain. *ISPRS Journal of Photogrammetry and Remote Sensing*, *78*, 102–115. https://doi.org/10.1016/j.isprsjprs.2013.01.008

Zermas, D., Teng, D., Stanitsas, P., Bazakos, M., Kaiser, D., Morellas, V., … Papanikolopoulos, N. (2015). Automation solutions for the evaluation of plant health in corn fields. *IEEE International Conference on Intelligent Robots and Systems*, (September), 6521–6527. https://doi.org/10.1109/IROS.2015.7354309

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W. (2020). Improving Unmanned Aerial Vehicle Remote Sensing-Based Rice Nitrogen Nutrition Index Prediction with Machine Learning. *Remote Sensing*, *12*(2), 215.

Zhang, C., Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, *13*(6), 693–712. https://doi.org/10.1007/s11119-012-9274-5

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., Carter, G. A. (2003). Corn (Zea mays L.) growth, leaf pigment concentration, photosynthesis and leaf hyperspectral reflectance properties as affected by nitrogen supply. *Plant and Soil*, *257*(1), 205–218. https://doi.org/10.1023/A:1026233732507

Zhu, Y., Fan, X., Hou, X., Wu, J., Wang, T. (2014). ScienceDirect Effect of different levels of nitrogen deficiency on switchgrass seedling growth. *CJ*, *2*(4), 223–234. https://doi.org/10.1016/j.cj.2014.04.005

Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net, 301–320.