A COST-EFFECTIVE MULTISPECTRAL SENSOR SYSTEM FOR LEAF-LEVEL PHYSIOLOGICAL TRAITS

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Electrical and Computer Engineering
University of Saskatchewan
Saskatoon, SK, Canada

By

MOHAMMAD HABIBULLAH

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering

57 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan S7N 5A9

Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9

Canada

Disclaimer

Reference in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan and shall not be used for advertising or product endorsement purposes.

Abstract

With the concern of the global population to reach 9 billion by 2050, ensuring global food security is a prime challenge for the research community. One potential way to tackle this challenge is sustainable intensification; making plant phenotyping a high throughput may go a long way in this respect. Among several other plant phenotyping schemes, leaf-level plant phenotyping needs to be implemented on a large scale using existing technologies.

Leaf-level chemical traits, especially macronutrients and water content are important indicators to determine crop's health. Leaf nitrogen (N) level, is one of the critical macronutrients that carries a lot of worthwhile nutrient information for classifying the plant's health. Hence, the non-invasive leaf's N measurement is an innovative technique for monitoring the plant's health. Several techniques have tried to establish a correlation between the leaf's chlorophyll content and the N level. However, a recent study showed that the correlation between chlorophyll content and leaf's N level is profoundly affected by environmental factors. Moreover, it is also mentioned that when the N fertilization is high, chlorophyll becomes saturated. As a result, determining the high levels of N in plants becomes difficult. Moreover, plants need an optimum level of phosphorus (P) for their healthy growth. However, the existing leaf-level P status monitoring methods are expensive, limiting their deployment for the farmers of low resourceful countries.

The aim of this thesis is to develop a low-cost, portable, lightweight, multifunctional, and quick-read multispectral sensor system to sense N, P, and water in leaves non-invasively. The proposed system has been developed based on two reflectance-based multispectral sensors (visible and near-infrared (NIR)). In addition, the proposed device can capture the reflectance data at 12 different wavelengths (six for each sensor). By deploying state of the art machine learning algorithms, the spectroscopic information is modeled and validated to predict that nutrient status.

A total of five experiments were conducted including four on the greenhouse-controlled environment and one in the field. Within these five, three experiments were dedicated for N sensing, one for water estimation, and one for P status determination. In the first experiment, spectral data were collected from 87 leaves of canola plants, subjected to varying levels of N fertilization. The second experiment was performed on 1008 leaves from 42 canola cultivars, which were subjected to low and high N levels, used in the field experiment. The K-Nearest Neighbors (KNN) algorithm was employed to model the reflectance data. The trained model shows

an average accuracy of 88.4% on the test set for the first experiment and 79.2% for the second experiment. In the third and fourth experiments, spectral data were collected from 121 leaves for N and 186 for water experiments respectively; and Rational Quadratic Gaussian Process Regression (GPR) algorithm is applied to correlate the reflectance data with actual N and water content. By performing 5-fold cross-validation, the N estimation shows a coefficient of determination (R^2) of 63.91% for canola, 80.05% for corn, 82.29% for soybean, and 63.21% for wheat. For water content estimation, canola shows an R^2 of 18.02%, corn of 68.41%, soybean of 46.38%, and wheat of 64.58%. Finally, the fifth experiment was conducted on 267 leaf samples subjected to four levels of P treatments, and KNN exhibits the best accuracy, on the test set, of about 71.2%, 73.5%, and 67.7% for corn, soybean, and wheat, respectively.

Overall, the result concludes that the proposed cost-effective sensing system can be viable in determining leaf N and P status/content. However, further investigation is needed to improve the water estimation results using the proposed device. Moreover, the utility of the device to estimate other nutrients as well as other crops has great potential for future research.

Acknowledgments

At first, all praise goes to the Almighty ALLAH for HIS blessings upon me.

I would like to extend my sincere gratitude towards my supervisors Prof. Khan A. Wahid and Prof. Anh Dinh. Without their continuous support, guidance, and supervision, my program, and this thesis work at the University of Saskatchewan, will not be possible.

I am grateful for the funding support, which made this work possible, from Canada First Research Excellence Fund (CFREF) managed and distributed by the Global Institute for Food Security (GIFS) at the University of Saskatchewan.

I would like to extend my sincere gratitude and appreciation to our collaborator Dr. Raju Soolanayakanahally (Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, Saskatoon) and his associate Shankar Pahari, for helping me designing experimental setup, and providing plant resources as well as the greenhouse-control environment. Moreover, Dr. Raju's valuable feedback in this thesis has been of great help in the assessment of the developed system.

I would like to thank Mohammad Reza Mohebian for his guidance in some technical difficulties. Also, I would like to thank Mohammad Abdul Moin, Rakibul Chowdhury, Rinku Basak, Ali Newaz Bahar, and Shayeb Shahariar for the assistance during the experiments. I thank all my friends in Canada and Bangladesh for believing and inspiring me all the time.

I want to dedicate all my achievements and hard work to my parents (Mohammad Abdul Baset and Laila Sultana), my wife (Marzana Afrin Tuli), my sister (Mahabuba Sultana), and my niece for their love and care. I cannot express in words how much I owe them for tolerating me when I was passing stressful times.

Dedication

Dedicated to my parents and my wife for their incredible love and support.

Table of Contents

Page Number

List of Tables

Page Number

List of Figures

Page Number

List of Abbreviations

°C	Degree Celsius
µW/cm2	microwatt per square centimeters
3D	Three-dimensional
A	Ampere
AAFC	Agriculture and Agri-Food Canada
ADC	Analog to Digital Converter
APPF	Australian Plant Phenomics Facility
cm	centimeter
CRI	Color Rendering Index
DT	Decision Tree
FWHM	Full-Width Half-Max
g	grams
g/L	gram per Liter
GB	Gigabytes
GHG	Global greenhouse gas
GHz	Giga Hertz
GPR	Gaussian Process for Regression
HSI	Hyperspectral Imaging
I2C	Inter-Integrated Circuit
IoT	Internet of Things
IPPN	International Plant Phenotyping Network
IR	Infrared
K	Potassium
KNN	The K-Nearest Neighbor
LED	Light Emitting Diode
MAE	Mean Absolute error
mAh	milli-ampere-hour
MCC	Matthews Correlation Coefficient
ml	milli-liter

MSE	Mean Squared Error
N	Nitrogen
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
nm	nanometer
OLED	Organic Light Emitting Diode
P	Phosphorus
P2IRC	Phenotyping and Imaging Research Center
PCB	Printed Circuit Board
PSO	Particle Swarm Optimization
R^2	Coefficient of Determination
RBF	Radial Basis Function
RGB	Red, Green and Blue - Three channels of an image
RMSE	Root Mean Squared Error
SPAD	Soil Plant Analysis Development
SVM	Support Vector Machine
SWIR	Short-Wave-Infrared
UAV	Unmanned Arial Vehicle
USB	Universal Serial Bus
USD $	U.S. Dollar
V	Volts
VIS	Visible

Chapter 1: Introduction

1.1. Global Food Security Bottleneck and High Throughput Plant Phenotyping

The continuing global demand for food is on the rise, resulting from population increase. In the next few decades, approximately 2.3 billion person increase is expected to be seen in global perspective [1]. To meet the commensurate food demand, agricultural production needs to be enhanced by 1.5 times [2]. Existing solutions to feed these populations include either intensive use of the existing cropland or clearing the land. However, these current practices for agricultural land expansion may create a potential threat for the environment, as both land clearing and extensive use will result in increasing the global greenhouse gas (GHG) emissions. So, "sustainable intensification", meaning "*producing the same food from same area of land while reducing the environmental impacts"* [2], is a demand of the time.

One potential way of sustainable intensification to solve the global food security bottleneck is high throughput plant phenotyping. Plant phenotyping refers to a "*quantitative description of the plant's anatomical, ontogenetical, physiological and biochemical properties*" [3]. In other words, phenotype is the assembly of the characteristics possessed by a cell or organism resulting from the interaction of the environment and the genotype. A more recent definition of plant phenotyping would be the investigation of the plant traits like physiology, yield, growth, ecology, and other basic quantitative parameters [4]. In simple words, plant phenotyping can be referred to as the collection of the methods and techniques utilized to measure plant structure, chemical traits, and growth development. Plant phenotyping can be categorized into several kinds depending on the multiple organizational levels like canopy, whole plant, cellular level, leaf level, and root level.

The efforts from several research domains have been incorporated into utilizing and optimizing the available technologies to address the need for plant phenotyping [5]. However, these technologies for the plant phenotyping are still under progress, and not fully explored, making it a gridlock for plant science research [5]. Several research organizations and institutes, such as Phenotyping and Imaging Research Center (P2IRC) [6], International Plant Phenotyping

Network (IPPN) [7], and Australian Plant Phenomics Facility (APPF) [8], have been formed to address this global food security bottleneck in this respect.

1.2. Leaf-level Plant Phenotyping

Leaf level plant phenotyping for high throughput plant breeding has been in the key focus of the plant science researchers recently. Leaf level traits can be two kinds – morphological traits, and physiological traits. The examples of leaf-level morphological phenotypic traits are leaf number, leaf shape, leaf area index, leaf expansion rate, and leaf thickness; whereas physiological traits include quantifying chemical properties of plants [9]. These properties are water content, macronutrients: nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), sulfur (S), and magnesium (Mg); and micronutrients: sodium (Na), iron (Fe), boron (B), zinc (Zn), manganese (Mn), and copper (Cu) [10]. The ongoing technological advancement has facilitated several non-invasive ways for quantifying leaf-level traits including imaging and non-imaging techniques. Several researchers have applied numerous imaging techniques: hyperspectral, multispectral, 2-dimensional visible and near-infrared, thermal imaging, and 3-dimensional (3D) cameras in plant phenotyping research [11] such as estimating leaf N estimation [12], P content determination [13], K [13], micronutrients [10], water content [14], leaf segmentation and counting [15] and others. In addition to imaging methods, optical spectroscopy, especially in visible and NIR spectroscopy, is also extensively used to determine several physiological traits [16].

Although these noninvasive technologies provide important information about plant traits, they need to be integrated into a larger context. Most of them are highly expensive and bulky that limits their application on a large scale. As a result, more robust, inexpensive and accessible technologies are needed to be explored to alleviate global food security bottleneck and make plant phenotyping a high throughput.

1.3. Research Objectives

The primary objective of this thesis is to develop a low-cost, portable, and light-weight system to monitor leaf-level physiological traits noninvasively. The following research objectives were set to meet the goal of developing an affordable system:

- To design and develop a reflectance-based multispectral sensing system using low-cost light detectors.

- To develop a software program to operate the hardware components and data collection procedures using existing libraries.
- To develop a machine learning pipeline including data cleaning, feature engineering, normalizing, optimizing, and modeling using state of the art algorithms.
- To determine the N levels in leaves, and correlate with ground truth N content in multiple crops.
- To compare the N estimation performance with commercially used SPAD meter.
- To estimate water content in leaves by correlating with reflectance data.
- To predict the P status of leaves.
- To compare the feasibility of the proposed system in terms of accuracy and cost over the existing devices.

1.4. Thesis Organization

The chapters are organized as follows:

Chapter 2 reviews the literature relating to existing technologies and recent advances in leaf-level plant phenotyping research. Moreover, different devices for sensing leaf N, P sensing, and water content are explained.

Chapter 3 provides the details of the methodology and experimental setup. This chapter elaborately explains the design requirements, a hardware description of the proposed low-cost sensing system, utilized machine learning algorithms, feature engineering methods, validation technique, experimental setup, and data modeling of five different experiments.

Chapter 4 describes the results of each experiment separately. Also, this chapter provides discussions regarding the results and comparisons with existing techniques.

Finally, *Chapter 5* concludes the findings of this research work and provides direction for further investigations and improvements to the designed system.

Chapter 2: Literature Review

This chapter includes the literature review of the existing technologies for sensing leaf-level physiological traits, specifically N, P, and water. Section 2.1 describes the current leaf N monitoring methods. Section 2.2 provides an overview of all available devices related to P monitoring in leaves, and finally, section 2.3 shows related kinds of literature for water content estimation. Moreover, the limitations of the current methodologies are included in each section.

2.1. Technologies for Sensing Leaf-level N Content

Researchers have been trying to discover several methods to monitor plants' N status over time. The current techniques include destructive as well as nondestructive approaches. Invasive determinations basically are chemical methods [17] namely Kjeldahl-digestion and Dumas-combustion. There exist two approaches to determine plant N noninvasively- light spectroscopy and hyperspectral imaging. The popularly used spectroscopic devices for sensing N are Soil Plant Analysis Development (SPAD) [18], FieldSpec [19], GreenSeeker [20], imagery from QuickBird satellite [21]. Although these are widely used for N correlation, they have some limitations. For example, the basis of SPAD meter is determining chlorophyll and it saturates at high N fertilization [22]. According to Xiong *et al.* [22], the fraction of N assigned to chlorophyll is very small and most of them are allocated to photosynthetic proteins. A good correlation (R^2=.86) with N by using FieldSpect 3 spectrometer was shown by Wang et al. [19]. But this device is expensive and less flexible to operate in the field. In addition, GreenSeeker is also expensive equipment and the determination saturates with increasing biomass/leaf area [23]. Moreover, the use of satellite imagery has some drawbacks such as- satellite's constant movement, cloudy weather, subscription cost, etc. Recently, hyperspectral imaging (HSI) is being used in several plant phenotyping applications including N sensing [24]. It facilitates not only spectral information but also spatial information. Yu et al. (2014) in [12] showed how HSI can be used to investigate the mapping of N distribution in leaves. However, HSI is normally used in research purposes as it is very expensive. So, developing a low cost, quick read, portable, light weight device to determine leaf

N content is very challenging. Figure 2-1 summarizes the overall techniques related to N estimation in crops.

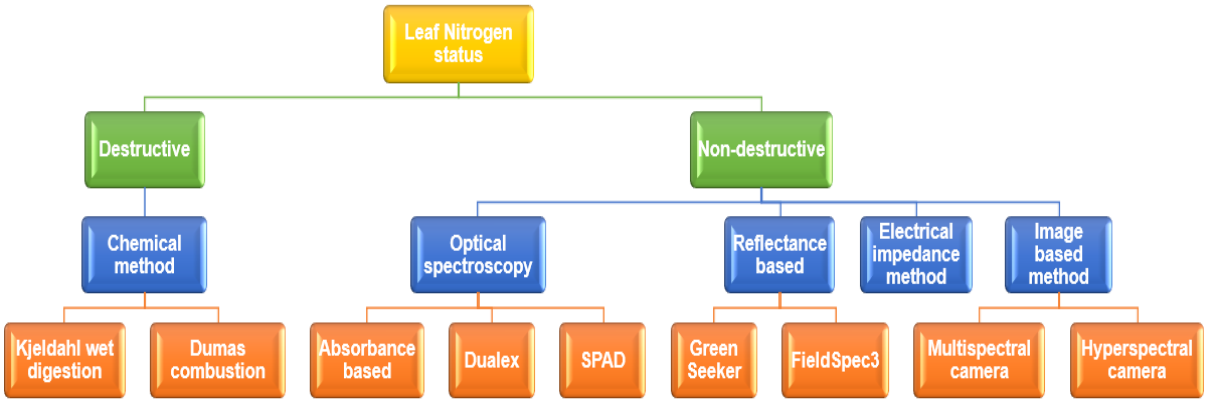


Figure 2-1: Methods used for estimating leaf N status.

Majority of the plant's uptake N in forms of nitrate although it depends on species. The chemical formation of N in plants is basically protein having peptide bonds (-CO-NH-)[25] and these are sensitive to visible and NIR regions. Blackmer et al. (1994) [26] investigated that the reflectance at 550 nm wavelength is effective for different N treatment separation. Many researchers recently have developed NIR spectroscopy based models to determine N in several plants including spring wheat [27], corn [28], winter oilseed rape [29]. In the article by Zhang et al. (2013)[13], authors published important wavelengths significant for detecting N by using HSI in oilseed rape leaves and showed $R^2 = 0.882$. According to the authors, twelve optical bands around 440, 473, 513, 542, 659, 718, 744, 865, 928, 965, 986 and 1015 nm are effective for sensing N contents.

2.2. Methods for Sensing P in Leaves

Phosphorus (P) is a vital element of some important macromolecules namely nucleic acid, phospholipids and phosphates [30]. Besides, some organic P molecules take part in energy transfer reactions and in respiration [13]. Also, P plays a crucial part in harvesting energy from the sun to generate sugar molecules in plants [31]. However, ensuring optimal levels of P content in the plant is difficult, as insufficient or excessive use of P fertilizer affects the overall growth and

development of the plant's health [32]. In addition, the cost of P fertilizer (phosphate) is getting increased, as a result, the optimal application of it has become an urgent need [33].

There exist several techniques including destructive and non-destructive to sense leaf's P levels. The traditional technique to measure P is laboratory-based stoichiometry. The common processes are P-vanadium-molybdenum and molybdenum-antimony anti-absorption spectrometry. Although the accuracy of this stoichiometry method is very high, they are invasive, complex, labor-intensive, time-consuming and expensive [34]. The development of machine learning algorithms has facilitated several non-destructive methods to model the plant's nutrition characteristics, especially P, with spectral imaging and electrical signals [4][35][36]. The existing imaging methods are - monocular vision, multispectral, hyperspectral and fluorescence imaging. In the monocular vision method, the plant images are collected by an RGB camera, on the aerial or ground level, to sense plant nutrition. In a comparative study between ground vs. aerial RGB to assess the plant's nutrition profile under different P fertilization, Adrian et al. suggested that RGB indices can be correlated with leaf fits application in the field setting.

Multispectral imaging is another imaging technique where several spectral bands, commonly in the visible and NIR range, are incorporated in the image to model plant nutrition [37]. Guoxiang et al. showed how multispectral 3D imaging techniques with appropriate modeling algorithms can be utilized for highly accurate determinations of p contents in tomato plants [34]. However, these cameras are expensive and show unstable performance in different light environments [34]. Finally, the utility of hyperspectral imaging is getting popular as it can provide both spatial and spectral information at the same time [38]. With that information, researchers have been able to find a correlation between these spectral pixel values and the plant's chemical traits. For example, Piyush Pandey et al. have created a model to determine macronutrients and micronutrients in leaves using lab-based hyperspectral imaging [10]; Liu Yanli et al. predicted the P content in citrus leaves using hyperspectral camera [32]. Hyperspectral imaging can also be utilized for investigating characteristic wavelengths of nutrients [37][14]. Xiaolei et al. conducted an experiment on oilseed rape leaves using hyperspectral imaging and found out that the optimal wavelengths for sensing P content are 468, 522, 698, 721, 817,967, 979 and 1025 nm. However, hyperspectral cameras are very expensive, and analyzing their data is very complex because of high dimensionality [39]. In addition to those cameras, VIS-NIR-SWIR spectroscopy is another

method to sense P in leaves. Basically, plant's P compounds (sugar-phosphate, phospholipids) are bonded by covalent bonding of carbon that absorbs VIS-NIR-SWIR light. That's why light reflectance/absorbance in these spectra can be correlated to P contents. In a case study with a maize diversity panel, Yufeng et al. investigated that leaf chemical traits can be modeled by VIS-NIR-SWIR spectroscopy [40]. In their experiments, they utilized FieldSpec 4 (Analytical Spectral Devices) spectroradiometer to correlate reflectance data to P content. Although these devices are portable and accurate, they are very expensive. Also, these instruments are bulky and less flexible for a quick read.

2.3. Devices for Water Content Monitoring

Leaf water content is another major factor for the overall health of the plants. One of the reasons is that water stress limits transpiration affecting crops' photosynthesis mechanism [41]. In addition, at the growing stage, the fertilizer application and irrigation rely on leaf water content [42]. So, the determination of leaf water content is of great importance for monitoring the health status of the plants. One of the common methods of determining leaf water content is calculating the difference between fresh leaf weight and dried leaf weight. This method is destructive and time-consuming. However, the applications of remote sensing such as spectrometry, HSI have been seen in several studies featuring non-destructive approaches. The determination of leaf water content in Miscanthus by using VIS/NIR was investigated by [43]. In that article, least-square support machine regression was used to model leaf water content with reflectance spectra. It also identified 75 significant wavelengths between 450-2500 nm range showing R^2 about 0.9899. R. Gente *et al.* utilized the terahertz time-domain spectroscopy technique that calculated the relative volumetric fraction of water present in the tissue that correlates very well with the direct determination of water content [44]. The recent development HSI has been proven to be very effective for in vivo analysis of plant chemical properties including water content [10]. In that study, Piyush Pandey et al. showed how HSI can be used to correlate hyperspectral images with leaf water content ($R^2 = 0.93$). In another study, where UV-VIS spectrometer (Evolution 300) was used, it was shown that 8 efficient wavelength intervals were effective for water content determination in leaves [45]. These are 553-556 nm, 689-720 nm, 755-842 nm, 950-970 nm, 1013-1034 nm, and 1055-1075 nm. The common drawbacks of these previous techniques are their high cost.

On-going technological advancements in plant phenotyping research have brought about significant opportunities to improve current farming practice. However, the prime bottleneck is the implementation of these techniques on a large scale. The available technologies and platforms are still in the research phase and are not ready for commercial use, and those available commercially are highly expensive, inaccessible, and bulky. Hence, there is a need to develop cost-effective solutions for this purpose. This thesis attempts to develop a low-cost solution for sensing leaf-level phenotyping like N, P, and water content. The proposed device/system will have the ability to utilize existing off the shelf sensors, hardware components, and algorithms. The primary target of the application is to make it a cost-effective, portable, and lightweight for monitoring nutrient status so that it can be implemented on large scale in the lab, greenhouse environment, and in the field.

Chapter 3: Methodology

This chapter describes the design and development of the proposed device, and overall methodologies. Section 3.1 sets the design requirement of the low-cost sensing system, section 3.2 describes the hardware components, the setup used to build the system, and a summary of the cost for individual parts of the system. Section 3.3 discusses the machine learning modeling algorithms used to create models. Section 3.4 overviews the feature engineering part and section 3.5 includes the validation techniques. Finally, section 3.6 separately describes the experimental setup and data modeling techniques of five conducted experiments.

3.1. Design Requirements

For developing a low-cost system for sensing leaf-level chemical traits of the plant, the following design requirements have been set:

- The system should be low-cost. It is one of the prime goals of this thesis.
- The developed system should have multifunctionality, that is, should be capable of estimating N, P, and water content at the same time.
- The system should sense spectral information at multiple bands in the visible and NIR regions.
- The system should be portable and capable of remote sensing operations wirelessly.
- The device should be lightweight to be carried around the field.

3.2. Hardware Design for the System

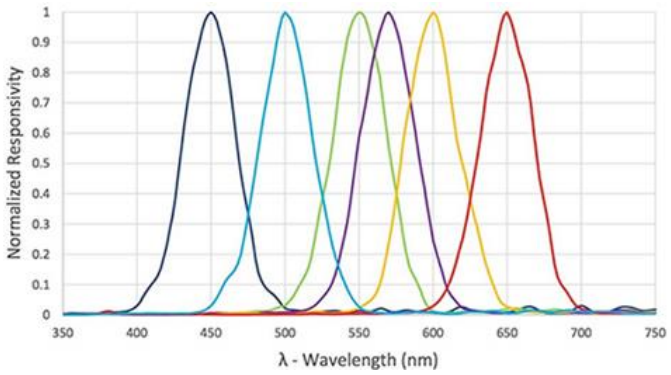
Any low cost sensing device that can detect light reflectance in visible and NIR regions will be applicable. However, most of the available spectral sensors and imaging systems working in those regions are expensive. One potential way might be Raspberry Pi NOIR camera, which is cost-effective ($30-$40) and can capture images in those regions. However, this camera does not provide light excitation on-board, so separate arrangements are needed to illuminate samples. That is why, this device was not used in this thesis.

Considering the design requirements, the proposed system prototype is designed based on two optical sensors- Sensor1 and Sensor2. Sensor1 is a visible multispectral sensor (AS7262, AMS), and Sensor2 is a NIR multispectral sensor (AS7263, AMS). Also, a Qwiic mux breakout board (TCA9548A, SparkFun Electronics) as a multiplexer and a Raspberry Pi version 3 as control circuitry are used. In addition, a power bank (BWA18WI035C, Blackweb), and an OLED display (DS-OLED-MOD, Cytron Technologies) are utilized in this prototype system. The descriptions of these components are discussed in the following sections.

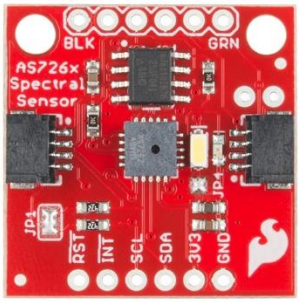
3.2.1. Sensor1

Sensor1 (Figure 3-1 a-d) is a 6-channel multispectral sensor in the visible range around 430 nm to 670 nm with full-width half-max (FWHM) of 40 nm. In this study, the visible AS7262 spectral breakout (SparkFun number SEN-14347) was used, where sensor1 is integrated. Here, it has built-in aperture controls of the light entering the process of the sensor array. Also, it has an I2C register set by which spectral data can be accessed. Here, the six visible channels are 450 nm (channel V), 500 nm (channel B), 550 nm (channel G), 570 nm (channel Y), 600 nm (channel O) and 650 nm (channel R). It has a 16-bit ADC (Analog to Digital Converter). Moreover, it's operating voltage ranges from 2.7V to 3.6V with I2C interface. The package field of view of the sensor is ±20 degrees. Calibration and measurements are made using diffused light. Each channel is tested with GAIN = 16× at ambient temperature (25°C) under a 5700K white LED test condition with an irradiance of ~600 μW/cm^2 (300-1000 nm). The measurement unit of the channel is μW/cm^2 with an accuracy of 12%. The energy at each channel is calculated with a ±40 nm bandwidth around the center wavelengths. The built-in excitation light source is used in this study. It is a 5700K white LED (L130-5780HE1400001, Lumileds) having a Color Rendering Index (CRI) of 80.

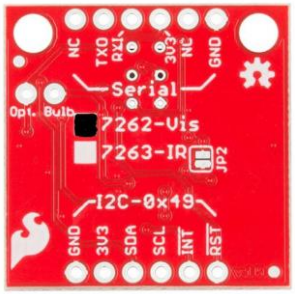
(a)



(b)



(c)



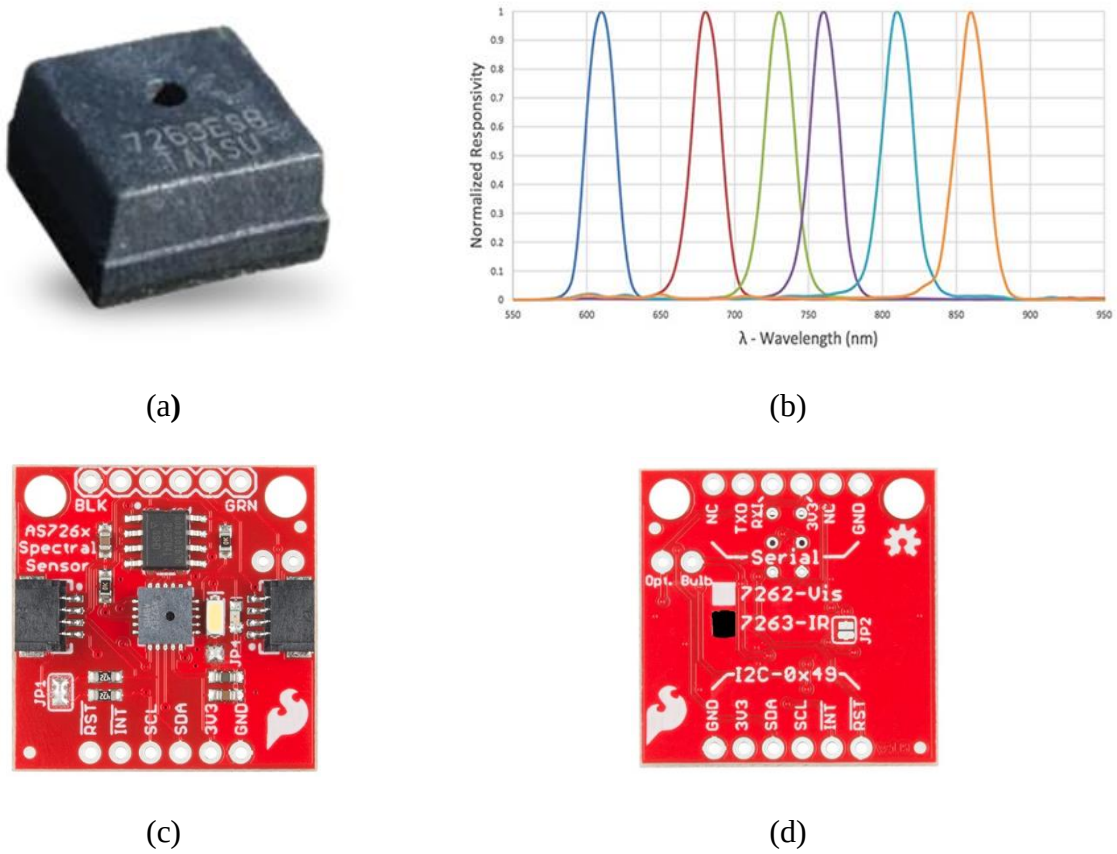
(d)

Figure 3-1: Sensor1. (a) Optical sensor AS7262 [46]. (b) Normalized spectral responsivity versus wavelength of Sensor1 which detects at 6 visible channels — 450 nm (channel V), 500 nm (channel B), 550 nm (channel G), 570 nm (channel Y), 600 nm (channel O) and 650 nm (channel R) each with 40 nm FWHM [47]. (c) front side and (d) backside of AS7262 spectral breakout (SparkFun number SEN-14347) [48]. Here, sensor1 is integrated on the board.

3.2.2. Sensor2

Sensor 2 (Figure 3-2 a-d) is a digital six-channel spectrometer in the NIR light region. The NIR AS7262 spectral breakout (SparkFun number SEN-14351**)** was used, where sensor2 is integrated. It has six independent optical filters whose spectral response is defined in the NIR wavelengths from approximately 600nm to 870 nm with full-width half-max (FWHM) of 20 nm. The channels are 610 nm (channel R), 680 nm (channel S), 730 nm (channel T), 760 nm (channel U), 810 nm (channel V) and 860 nm (channel W). The light source used in the test condition is an incandescent light with an irradiance of ~1500 μW/cm^2 (300-1000 nm). Also, the energy at each channel is calculated with a ±33 nm bandwidth around the center wavelengths. As a NIR source

light, the onboard 2700K warm LED (L130-2790001400001) is utilized having a CRI of 90. The other configurations that are like Sensor1 are not mentioned.



(a) (b)

(c) (d)

Figure 3-2: Sensor2. (a) Optical sensor AS7263 [49]. (b) Normalized spectral responsivity versus wavelength of Sensor2 which detects at 6 channels — 610 nm (channel R), 680 nm (channel S), 730 nm (channel T), 760 nm (channel U), 810 nm (channel V) and 860 nm (channel W) each with 20 nm FWHM [47]. (c) front side and (d) backside of AS7263 spectral breakout (SparkFun number SEN-14351) [50]. Here, sensor2 is integrated on the board.

3.2.3. Control circuitry

Raspberry Pi 3 Model B (RP3) was used for controlling the sensors (Figure 3-3). The RP3 has a quad-core processor having 1.2 GHz and 1 GB LPDDR2 RAM. Also, it has BCM43438 wireless LAN and Bluetooth low energy on board. It has 40 pins extended GPIO pins and 4 USB ports. It also provides full-size HDMI and micro SD port for loading your operating system and storing data. It can be powered through Micro USB power source up to 2.5 A. This widely popular board is used in several applications such as image processing [51], IoT systems [52], etc.

Figure 3-3: Raspberry Pi [53]

3.2.4. Multiplexer

Sensor1 and Sensor2 have the same I2C address. So, a multiplexer (Figure 3-4) was used that has eight configurable addresses of its own providing 8 I2C buses (TCA9548A, SparkFun Electronics). The operating voltage of the component is 1.65V - 5.5V and the operating voltage is -40°C to 85°C.

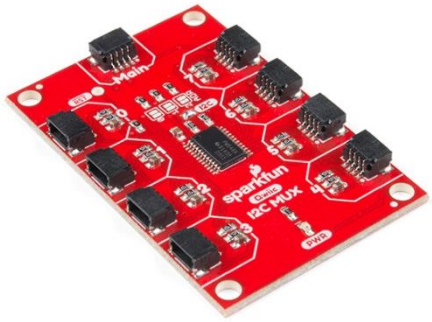


Figure 3-4: Multiplexer having 8 configurable pins [54].

3.2.5. Power Module

As a portable power supply for the device, a 5V/2A rated power bank (Figure 3-5) was used to power up the control circuit (BWA18WI035C, Blackweb). It has a built-in 5000 mAh high capacity rechargeable Li-polymer battery. This has overcharging/discharge, short circuit, and

current protection. Moreover, it provides dual USB output, where a Micro USB charge cable can be connected.

Figure 3-5: Power Module. A 5V/2A rated power bank was used to power up the control circuit having 5200 mAh battery [55].

3.2.6. Display

The prototype also includes a 1.3-inch I2C OLED display ((DS-OLED-MOD, Cytron Technologies) for visualization. The module (Figure 3-6) has a resolution of 128×64 resolution working without backlight. It shows the text color in white and operating input voltage is 3.3V/5V. Moreover, it has four pins: VCC (power), GND (ground), SDA (data cable), and SCL (clock). This display is controlled by RP3 for displaying.

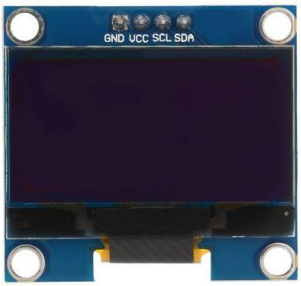
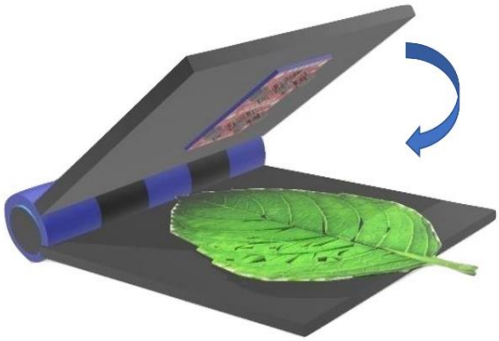
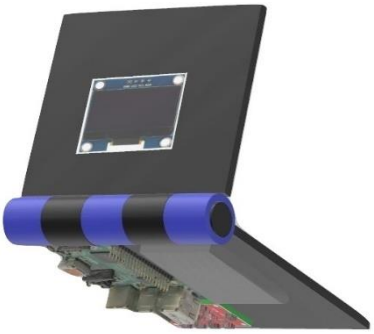


Figure 3-6: 1.3-inch I2C OLED display ((DS-OLED-MOD, Cytron Technologies) [56].

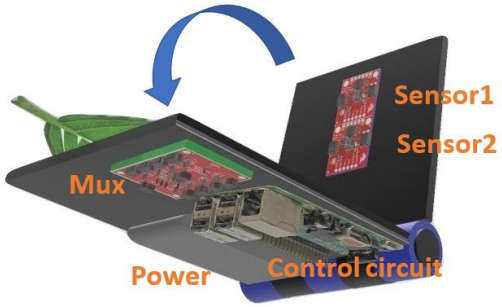
Figure 3-7 a-d shows the graphical setup of the prototype, whereas Figure 3-7 e-f shows the sample of the prototype. The leaf is placed as shown in Figure; both the sensors get in contact with the leaf while scanning.



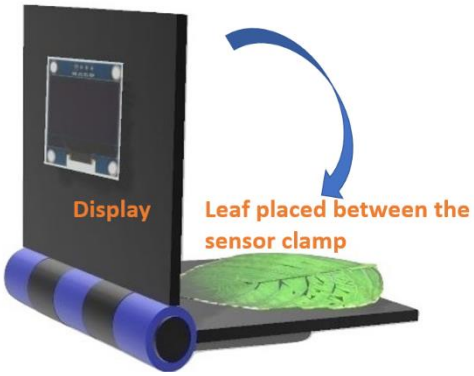
(a)



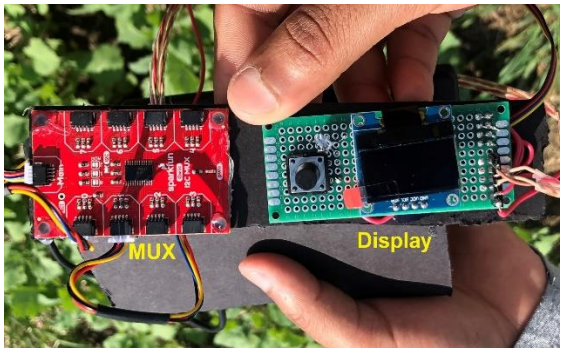
(b)



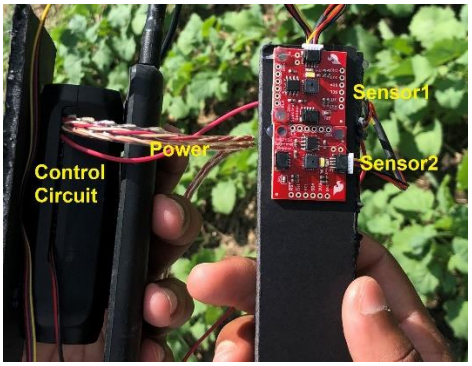
(c)



(d)



(e)



(f)

Figure 3-7: Graphical setup of the device from (a) front side, (b) back side, (c) bottom, and (d) right side. Sample of the device showing (e) MUX and display, and (f) control circuit, power module, sensor1, and sensor2. Here, sensor1 and sensor2 are connected to the multiplexer, and the multiplexer is connected to the control circuit. The whole is getting powered from power supply, and display is connected to the control circuit. The leaf is placed accordingly and sensor1 and sensor2 scan the leaf surface by contacting.

3.2.7. Full Design of the System

Figure 3-8 shows the connection diagram of the prototype. Here, the Qwiic ports of sensor1 and sensor2 are connected to the port1 and port2 of the mux through Qwiic connectors. From the main port of the mux, four wires namely: SDA, SCL, VCC, and GND are connected to pin numbers 3, 5, 1, and 9 of the control circuit respectively. Moreover, the power bank is connected to the micro-USB port of the control circuit; and the OLED display to the SDA, and SCL. Data collected from the sample leaves are modeled and processed using MATLAB 2018b.

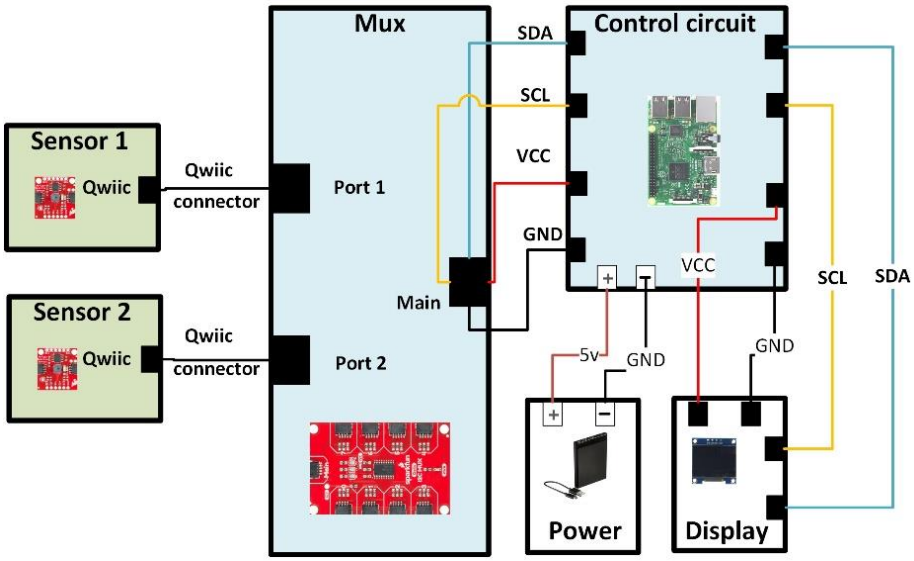


Figure 3-8: Connection overview of the proposed device. Sensor1 and sensor2 are connected to the port1 and port2 of the mux through Qwiic connectors. The four lines from the main port of the mux are connected to the SDA, SCL, VCC, and GND of the control circuit. The power bank is connected to the micro-USB port of the control circuit; and the OLED display to the SDA, and SCL.

3.2.8. Cost of the Parts of the Proposed Sensor

The total cost of the device is $200, as of 10 February 11, 2020. The cost breakdown of the components of the system is shown in Table 3-1.

Table 3-1: Cost of the components

Device components	**Approximate cost (USD)**
Sensor1	$25
Sensor1	$25
Multiplexer	$20
Raspberry Pi 3	$50
Power bank	$10
Display	$5
Manufacturing	$50
Others	$15
Total	**$200**

3.3. State-of-the-art Algorithms

In this work, several machine learning techniques are utilized for data modeling purposes. These are K-means clustering, KNN, SVM, Decision Tree, Ensemble, and GPR. They are discussed below.

3.3.1. K-means Clustering

K-means clustering is an unsupervised data analysis method that is commonly used to segment data set in groups. It is an iterative algorithm which attempts to partition the dataset into K-pre-defined different non-overlapping clusters where each data point falls into only one group. Also, the inter-cluster data points are maintained as similar as possible while also keeping the clusters as distinct as possible. In this method, data sub-groups (clusters) can be identified based on the similarity of the data points. Here, similarity can be measured based on Euclidean distance between the data points. This technique is widely used for plant phenotyping [57], pattern recognition, segmentation of medical image [58]. In the cluster, the center of the cluster is called centroid which is nothing but the mean of the data points belonging to each cluster. In this research,

K-means clustering is used in the preprocessing stage using the value of K as three to cluster three different observations.

3.3.2. KNN

The K-Nearest Neighbor (KNN) is an instance-based supervised classification method that works based on the closest training examples in feature space [59]. It is the simplest technique when there is no prior knowledge about the distribution of the dataset. Moreover, when the number of samples is larger than the number of features, KNN works better than classifiers that have a learning step. KNN has been applied in many fields; such as plant phenotyping [60], plant disease classification [61], and detection of N status on plants [62]. In this thesis, the value of K is chosen based on trial and error method. Different values of K are implemented, and the value one shows minimum error.

3.3.3. SVM

Support Vector Machine (SVM) is a supervised learning algorithm applicable to both classification and regression problems, but mostly classification. It has been widely used in several research domains like plant disease detection [63], blood glucose classification [64], and speech emotion recognition [65]. SVM relies on a set of hyperplanes in a high dimensional space that separates the features according to the number of class labels. For separation of the data by the hyperplanes, data points are mapped from input space to high dimensional feature space, where data are sparse and more separable. Radial basis function (RBF) is commonly used as its kernel, which simplifies the computation of the inner product value of the transformed data in the feature space. The performance of the classifier depends highly on choosing the parameters of the kernel. The soft-margin and RBF kernel parameters are optimized using the Wang [66] method. In this method, inter-cluster distance is calculated in the feature space to determine parameters. The distance index with the kernel parameter combination that leads to best separation index are calculated iteratively and thus optimized.

3.3.4. Decision Tree

Decision Tree (DT) divides the dataset to small subsets and forms a tree structure; thus develop a decision tree with decision and leaf nodes [67]. Here, each brunch represents a decision,

each node represents a feature, and each leaf represents an outcome. There are several DT algorithms in which ID3 (Iterative Dichotomiser 3) is mostly used. This algorithm uses entropy function and information gain as metrics. The entropy of features is used to determine the homogeneity of samples to construct the tree. It is essential that the nodes are aligned as such that the entropy decreases with splitting downwards. In this research, A C4.5 decision tree is used [68]. Data at every node of the tree is sorted by C4.5 for the best separation attributes. Also, it inherently employs single pass pruning process to mitigate overfitting. DT has been used previously in various applications such as theft detection in smart grid [69], crop disease detection [70], and N content estimation [71].

3.3.5. Ensemble

Ensemble Bagged Decision Tree is a bag of the decision trees that uses ensemble technique for aggregating results [72]. This bagging method was introduced by Breimann [73]. The details of bagging can be found in [74]. In this method, the training dataset is divided into several bags, and a decision tree model is built for each bag. Bagging is used when our goal is to reduce the variance of a decision tree. The aggregation of these models together forms the final classifier. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner. This is a popular method applied in several fields like cancer detection [75], plant segmentation [76], and leaf N estimation [77]. In this thesis, the ensemble bagged decision tree is applied in classification tasks.

3.3.6. GPR

For, correlating the spectral data with the actual N content, Gaussian Process for Regression (GPR) was found to be the most effective [78]. This technique has been used for classification [79] and regression [80] in different domains. This method achieves a significant interest in statistical modeling for its good performance in prediction [81]. Gaussian process regression is a nonparametric, probabilistic, Bayesian approach based on kernels. GPR explains the response by introducing latent variable from gaussian process, and the covariance of the latent function is calculated by joint distribution of those latent variables. The latent variables are introduced for each of the observations which make GPR non-parametric. The covariance of the

variable is defined using different kernel functions. In this study, 'Rational Quadratic' as kernel is used as it shows best result. Also, constant basis functions were specified in the GPR method

3.4. Feature Engineering

Normalization changes the underlying probability distribution of features. For this study, a modified standard score (z-score) is used for normalization [82]. Moreover, the median was also applied to make the normalization robust against outliers [83]. Then, the feature selection is performed. Feature selection approaches can be categorized into three categories including wrapper, filter, and embedded methods [84]. The filter methods work unaided from the classifier. On the other hand, the wrapper feature selection methods formulate an objective function and search all the problem space with the combination of features for the best selection. An independent-sample t-test, which is a filter method, is used to identify statistically discriminative normally distributed features [85].

Finally, the embedded method evaluates the performance of the classifier for predicting the best feature set with searching that is conducted by a learning classification process. We use the embedded method, in which features are weighted based on the Particle Swarm Optimization (PSO) algorithm during learning. PSO is an evolutionary computational method inspired by flocking birds [86], applied in many different areas, including manufacturing [87], plant phenotyping [88], optimum design [89], etc.. In this research, the swarm size of the PSO is set to 200, and the maximum number of iterations is 500. Also, the range of the weights of the features is -5 to 5, and the tolerance limit is set to 10^{-12}. Moreover, the Matthews Correlation Coefficient (MCC) is used as a cost function to optimize the feature weights [82]. The parameters of PSO are chosen based on Mohebian *et el.* [84]. The overview feature engineering using PSO is shown in Figure 3-9.

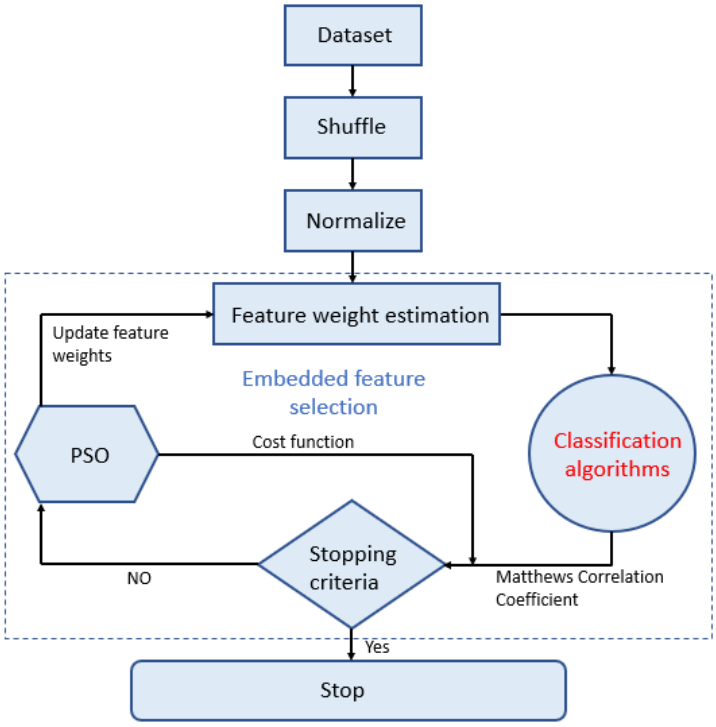


Figure 3-9: Overview feature engineering using PSO that optimizes the weights of the features in every cycle.

3.5. Model Validation

Validation technique is used to assess how a model behaves on new data. There are several ways to perform validation; in which cross-validation is the most popular one. It is basically used in predictive analytics, and determining how accurately a predictive model will behave in practice. First, a model is usually trained on one portion of the dataset (training data), later the model is tested on the remaining dataset (test set). Cross-validation tackles problems like overfitting or bias and helps model generalize. The effectiveness of a model can be evaluated by using several evaluation metrics.

3.5.1. Hold-out Method

Hold-out method is a validation technique where data is divided into a train and test sets. After training the system, the trained model is applied to the test set. In this thesis, hold out is used in which the train-test ratio was 75%-25%.

3.5.2. K-fold Cross-Validation

For validating the model on new data, k-fold cross-validation (CV) [90] was also performed in this work using the value of k as five that divides the dataset into five subsets. Each model was trained on 4 subsets and tested on the remaining set, and this was run five times to compute the average performance. So, every data point gets the chance to be tested once and trained 4 times.

3.5.3. Validation Metrics

Validation metrics are used to effectively measure the performance of a model. These metrics are selected based on the problem- classification or regression.

3.5.3.1. Classification Metrics

The final task of the modeling process is classification, specifically multiclass classification. In a systematic analysis regarding classification [91], researchers have defined several validation metrics. The validation metrics used in our study are accuracy, sensitivity/recall, specificity, precision, and F1-score. The classification performance measures used in our study are listed below in Table 3-2:

3.5.3.2. Regression Metrics

The performance of the cross-validation model was evaluated in terms of primary metrics stated in [92]. They are the root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). The definition of the metrics is given in Table 3-3.

Table 3-2: Classification metrics

Parameter	**Evaluation focus**	**Definition**
$\mathrm{Accuracy_i}$	Effectiveness of a classifier for i-th class	$\frac{\mathrm{TP_i}+\mathrm{TN_i}}{\mathrm{TP_i}+\mathrm{TN_i}+\mathrm{FN_i}+\mathrm{FP_i}}$
$\mathrm{Sensitivity_i}$ $/\mathrm{Recall_i}$	Effectiveness of a classifier to identify positive labels for i-th class	$\frac{\mathrm{TP_i}}{\mathrm{TP_i}+\mathrm{FN_i}}$
$\mathrm{Specificity_i}$	Effectiveness of a classifier to identify negative labels for i-th class	$\frac{\mathrm{TN_i}}{\mathrm{TN_i}+\mathrm{FP_i}}$
$\mathrm{Precision_i}$	Class agreement of the data labels with the positive labels for i-th class	$\frac{\mathrm{TP_i}}{\mathrm{TP_i}+\mathrm{FP_i}}$
$\mathrm{F1-Score_i}$	Relations between positive labels and those given by a classifier for i-th class	$\frac{2\times\mathrm{Precision_i}\times\mathrm{Sensitivity_i}}{\mathrm{Precision_i}+\mathrm{Sensitivity_i}}$
$\mathrm{Accuracy_m}$	The average per-class effectiveness of a classifier	$\frac{\sum_{i=1}^{l}(\frac{\mathrm{TP_i}+\mathrm{TN_i}}{\mathrm{TP_i}+\mathrm{FN_i}+\mathrm{FP_i}})}{l}$
$\mathrm{Sensitivity_m}$ $/\mathrm{Recall_m}$	Effectiveness of a classifier to identify class labels if calculated from sums of per-category decisions	$\frac{\sum_{i=1}^{l}\mathrm{TP_i}}{\sum_{i=1}^{l}(\mathrm{TP_i}+\mathrm{FN_i})}$
$\mathrm{Specificity_m}$	The average per class effectiveness of a classifier to identify negative labels	$\frac{\sum_{i=1}^{l}\mathrm{TN_i}}{\sum_{i=1}^{l}(\mathrm{TN_i}+\mathrm{FP_i})}$
$\mathrm{Precision_m}$	Agreement of the data class labels with those of a classifiers if calculated from sums of per-category decisions	$\frac{\sum_{i=1}^{l}\mathrm{TP_i}}{\sum_{i=1}^{l}(\mathrm{TP_i}+\mathrm{FP_i})}$
$\mathrm{F1-Score_m}$	Relations between data's positive labels and those given by a classifier based on a per-class average	$\frac{2\times\mathrm{Precision_m}\times\mathrm{Sensitivity_m}}{\mathrm{Precision_m}+\mathrm{Sensitivity_m}}$

l is the number of categories. True positive ($\mathrm{TP_i}$): Spectral data belong to the i-th category which is correctly identified; false positive ($\mathrm{FP_i}$): Spectral data which do not belong to the i-th category,

incorrectly identified; true negative (TN_i): Spectral data which do not belong to the i-th category, correctly identified; false negative (FN_i): Spectral data belong to the i-th category, incorrectly identified.

Table 3-3: Regression Metrics

Validation parameter	**Definition**
Root mean square error (RMSE)	$\sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$
Mean squared error (MSE)	$\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}$
Mean absolute error (MAE)	$\frac{\sum_{i=1}^{N}\lvert(\hat{y}_i - y_i)\rvert}{N}$
Co-efficient of determination (R^2)	$1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\overline{y} - y_i)^2}$

y_i stands for actual content (N/water), $\hat{y}_i$ for predicted content (N/water) and N for the number of test data points in each fold.

3.6. Experimental Setup and Data Modeling

In this thesis, a total of five experiments were performed including four on the greenhouse-controlled environment and one in the field. Within this five, three experiments were for N sensing, one for water estimation, and one for P status determination. The experimental setup and data modeling techniques in each experiment are discussed in the following sections.

3.6.1. Experiment 1: N Level Classification in Canola

Canola seeds were sowed on the 2nd November of 2018 in a controlled greenhouse environment situated in Agriculture and Agri-Food Canada (AAFC), Saskatoon. During the first 3 weeks, all 24 pots were fertilized with slow-release 15-30-15 (15% N, 30% P, and 15% K) fertilizer at a rate of 4 g/L to ensure uniform establishment. At the end of 3 weeks, the plant-pots were brought to an indoor room and organized on a shelf as Figure 3-11. Later, the twenty-four pots were divided into four N concentrations each containing six replicate plants (4 N levels × 6

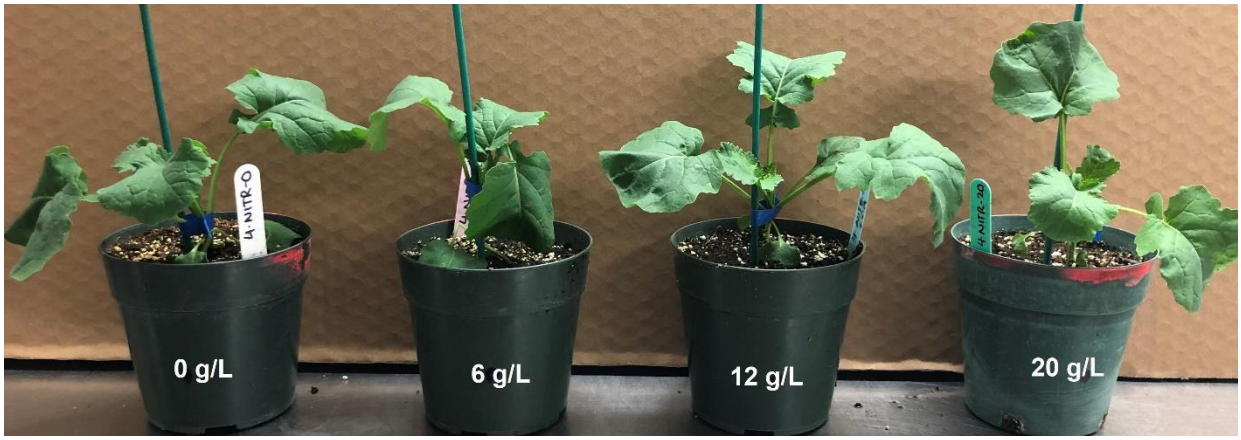


Figure 3-10: One random canola plant from each of the four-N fertilization. The photograph was taken during week seven.

pots = 24 plants, Figure 3-10). Henceforward, only N fertilizer 30-0-0 (30% N, 0% P, and 0% K) was applied three times a week at four concentration levels: 0 g/L, 6 g/L, 12 g/L, and 20 g/L.

(a) (b)

Figure 3-11: (a) Canola plants at AAFC (Agriculture and Agri-Food Canada) control environment on the last day of the 3rd week, (b) Indoor plants under LED lights.

Two panels of 45 W LED grow light is used which has 117pcs of red LEDs (630 nm) and 52 pcs of blue LEDs (470 nm) in a ratio of 9:4, which illuminated 16 hours a day. The day temperature was kept at 25˚C and night at 20˚C with a relative humidity of ~45%.

The proposed prototype measures the reflectance at different wavelengths. At each time of collecting data, the device is calibrated by taking reflectance data from a white surface. It is worth to mention that a regular white mirror paper is used as a reference for calibration. The reflectance data at 12 different wavelengths (450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm) are collected for each of the leaves. reflectance data collection started at week seven after the sowing. Here, 87 fresh leaves are selected from 24 plants. All the data is collected from three different positions around the midrib of a leaf by scanning 15 times. K-means clustering is performed on the 15 scans by using the value of K as 3. From the 3 most different clusters, 3 centroids are selected corresponding to the reflectance from the three different positions. The reflectance data collected at 12 wavelengths from the leaves are used as features to classify N levels. The purpose of the classification is to investigate if a model can be built using the reflectance by the sensor to determine the four categories of responses from four N treatments. Next, the data set is randomly shuffled and normalized using z-score. After that embedded feature selection with particle swamp optimization is applied. Finally, the KNN classifier is utilized to train a model on 75% of the data and tested on 25%. The process flow diagram is shown in Figure 3-12.

3.6.2. Experiment 2: High-low N Identification in the Field

This experiment was conducted at Lewellyn Farm, Saskatoon, Saskatchewan, Canada during the summer of 2019 (Figure 3-13 a). In this experiment, a total of 42 canola cultivars were subjected to low-N and high-N levels. Here, a total of 336 plots (42 cultivars × 4 replicates per N levels) were subjected to two levels of N treatments. From each plot, three leaves (one from each plant) from three central and south-facing plants, were used in this work. So, the total number of leaf samples was 1008 (3 × 336). Reflectance data are collected from these 1008 leaves and preprocessed similarly as shown in Figure 3-12. Here, the reflectance data collected at 12 wavelengths from the leaves are used as features to classify N levels. The purpose of the classification is to investigate if a model can be built using the reflectance by the sensor to determine two categories in the field experiment.

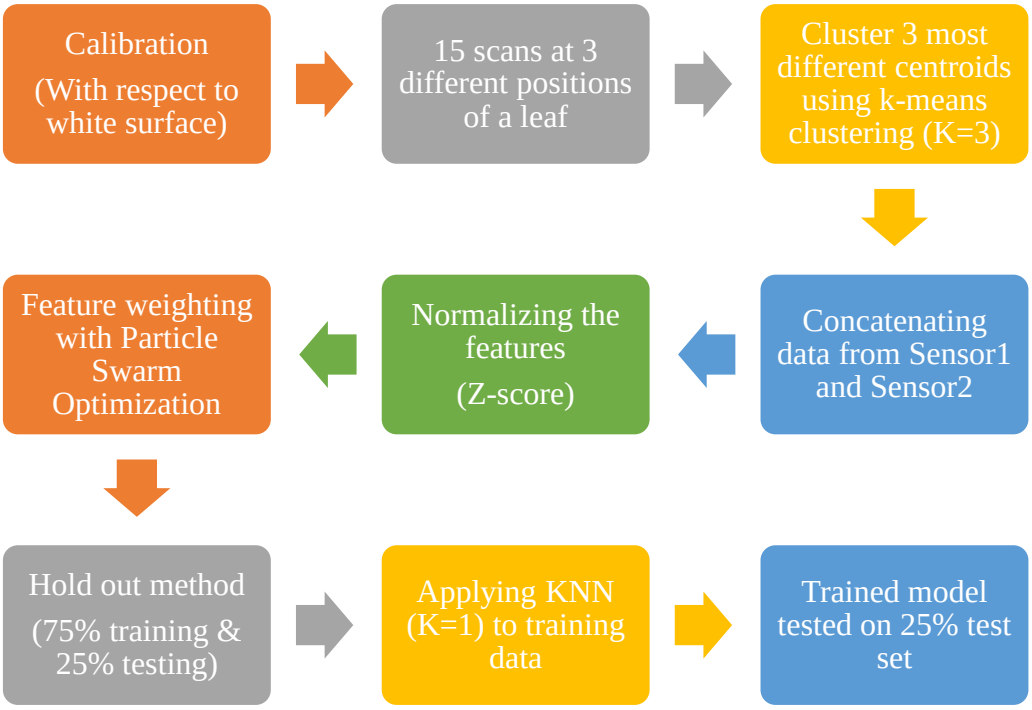


Figure 3-12: Process flow diagram of the methodology starting from calibrating the data with respect to the white surface to hold out testing.

3.6.3. Experiment 3: Actual N Content Determination in Multiple Crops

This N experiment was conducted on a total of 64 plants consisting of canola, corn, soybean, wheat each having 16 pots. All the seeds were sowed on the 2nd of February, 2019 in a controlled greenhouse environment (Figure 3-14 a-b) situated in the Agriculture and Agri-Food Canada (AAFC), Saskatoon. During the first 3 weeks, the plants were fertilized with slow-release 15-30-15 (15% N, 30% P, 15% K) fertilizer at a rate of 4 g/L to ensure uniform establishment. Later, for the N experiment, the 64 pots were divided into four concentration levels ensuring equal distribution of plants from each species (Figure 3-14 c-f). Henceforward, only N fertilizer (30-0-0) was applied three times in a week at four concentrations 0 g/L, 6 g/L, 12 g/L, and 20 g/L.

Data were collected each day starting from March 19 to March 26, 2019. Here, the leaves from 20 g/L fertilizer were not used in the experiment as they were intoxicated due to overfertilization. Also, 12 g/L samples of corn and soybean were not used for the same reason. So,

(a)

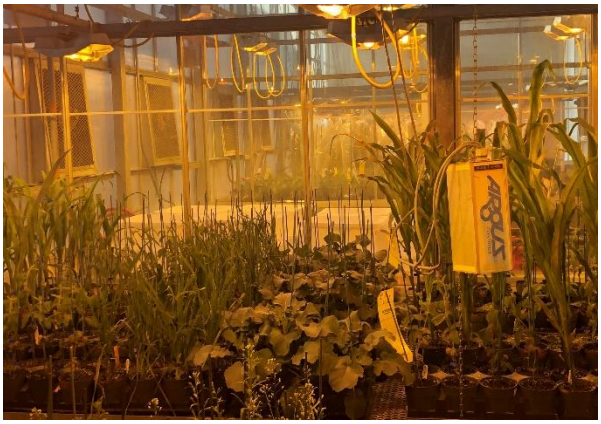
(b)

(c)

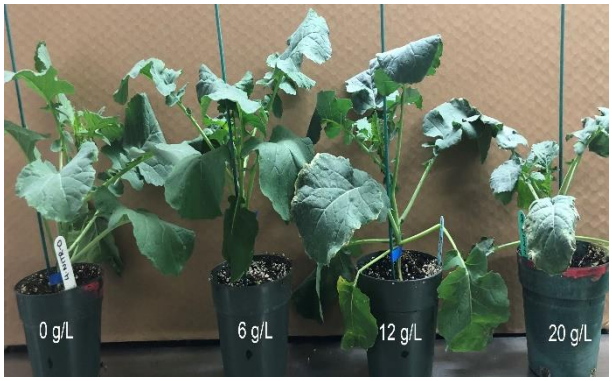
Figure 3-13: Field view of different canola plots (a). The proposed sensor (b), is used for taking measurements in the fields (c).

the number of leaf samples used for this experiment was 121 including 28 corn, 21 soybean, 36 wheat, and 36 canola leaves. For measuring the actual N content, the leaves were cut and then placed in the oven-dried at 50ºC for 72 to 96 hours to make them completely dried. The LECO TruMac N analyzer was used as shown in Figure 3-15 a-c.

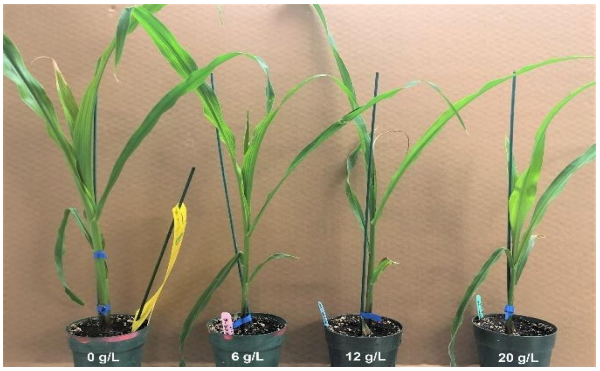
(a)



(b)

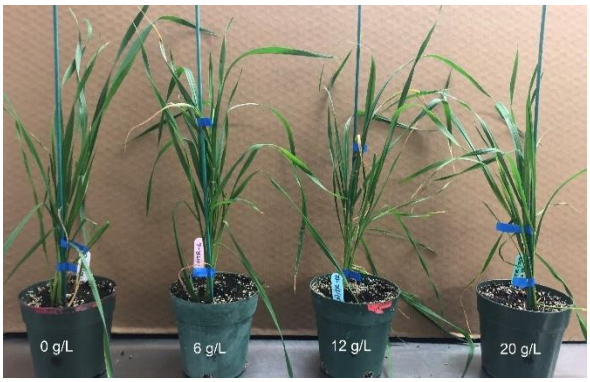


(c)



(d)

(e)

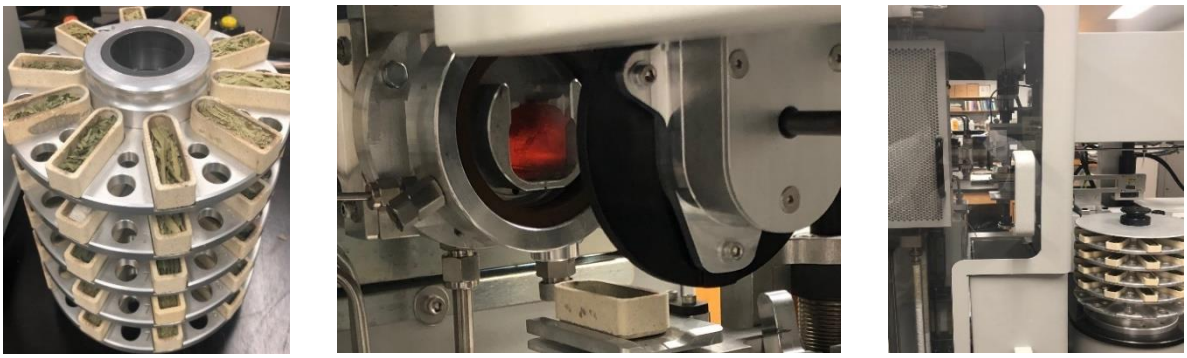


(f)

Figure 3-14: (a) and (b) Two different views of the greenhouse-controlled environment. Sample pots of (c) canola, (d) corn, (e) soybean and (f) wheat pots subjected to 4 levels of N fertilization. applications.

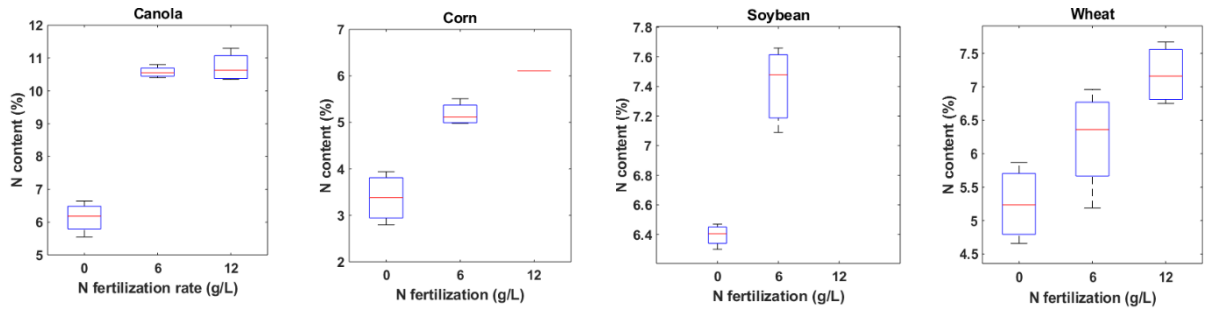
The analyzer is a macro combustion N degerminator that utilizes a pure oxygen environment in a ceramic boat for the macro sample combustion process. The box plots of N contents are shown in Figure 3-16, and Figure 3-17. The range of the N content for 3 levels of

fertilizer application (0, 6, 12 g/L) for wheat, canola and 2 levels of fertilizer application (0, 6 g/L) for corn and soybean can be observed in Figure 3-16 a-d for 4 types of plants. The variations of wheat 36 wheat samples are very significant followed by canola. The variations combining all the N samples were shown in Figure 3-17. The lowest, mean, and highest N contents were found to be 2.8%, 6.8%, and 11.3%.



(a) (b) (c)

Figure 3-15: Actual N content measurement using a LECO TruMac N analyzer system. Samples placed in the (a) stacked tray, are collected automatically by a (b) combustion zone. (c) shows the overall setup.



(a) (b) (c) (d)

Figure 3-16: Boxplot of leaf N content of (a) canola, (b) corn, (c) soybean, and (d) wheat. Here, the horizontal axis represents the rate of N fertilization in g/L, and the vertical axis represents the N content.

The collected reflectance data are shuffled and normalized. Then t-test is applied. After that, GPR is deployed using 5-fold cross-validation. The overall methodology is shown in as a process flow diagram in Figure 3-18.

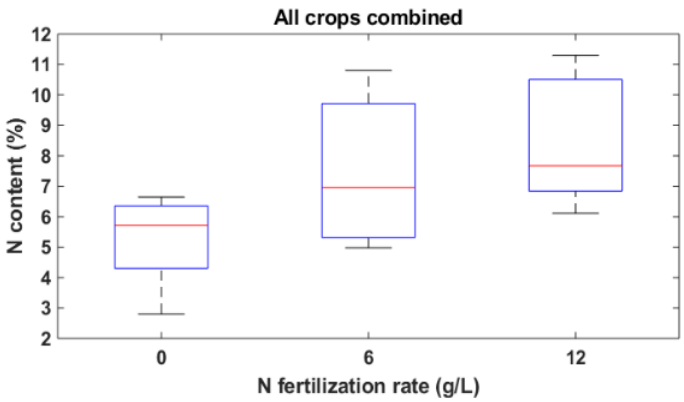


Figure 3-17: Boxplot of leaf N content of combined samples of four species

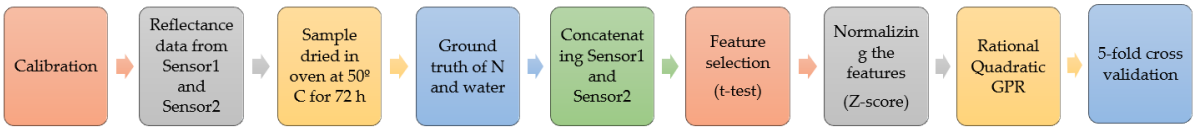


Figure 3-18: Process flow of the methodology.

3.6.4. Experiment 4: Water Content Estimation

Water experiment was conducted on 64 plants including canola, corn, soybean, and wheat each having 16 pots overall. All the seeds were sowed on the 2nd of February, 2019 in a controlled greenhouse environment situated in the Agriculture and Agri-Food Canada (AAFC), Saskatoon. During the first 3 weeks, the plants were fertilized with slow-release 15-30-15 (15% N, 30% P, 15% K) fertilizer at a rate of 4 g/L to ensure uniform establishment. In the case of water experiment, 64 pots were divided into 4 sections making the same number of plants for each species for each section. After that, water was varied by applying it daily at a rate of 50 ml, 100 ml, 150 ml, and 200 ml respectively in these four sections. Figure 3-19 a-d shows 4 different species of plants.

Data were collected each day starting from March 19 to March 26, 2019. After calibration, reflectance data are collected from the leaf surface. In the water experiment, one plant from corn and one plant from soybean were not used as they were dry at the time of data collection. So, a total of 186 leaves including 45 corn, 45 soybeans, 48 wheat, and 48 canola leaves were used in the water experiment. For determining the actual water contents of the samples, the leaves were

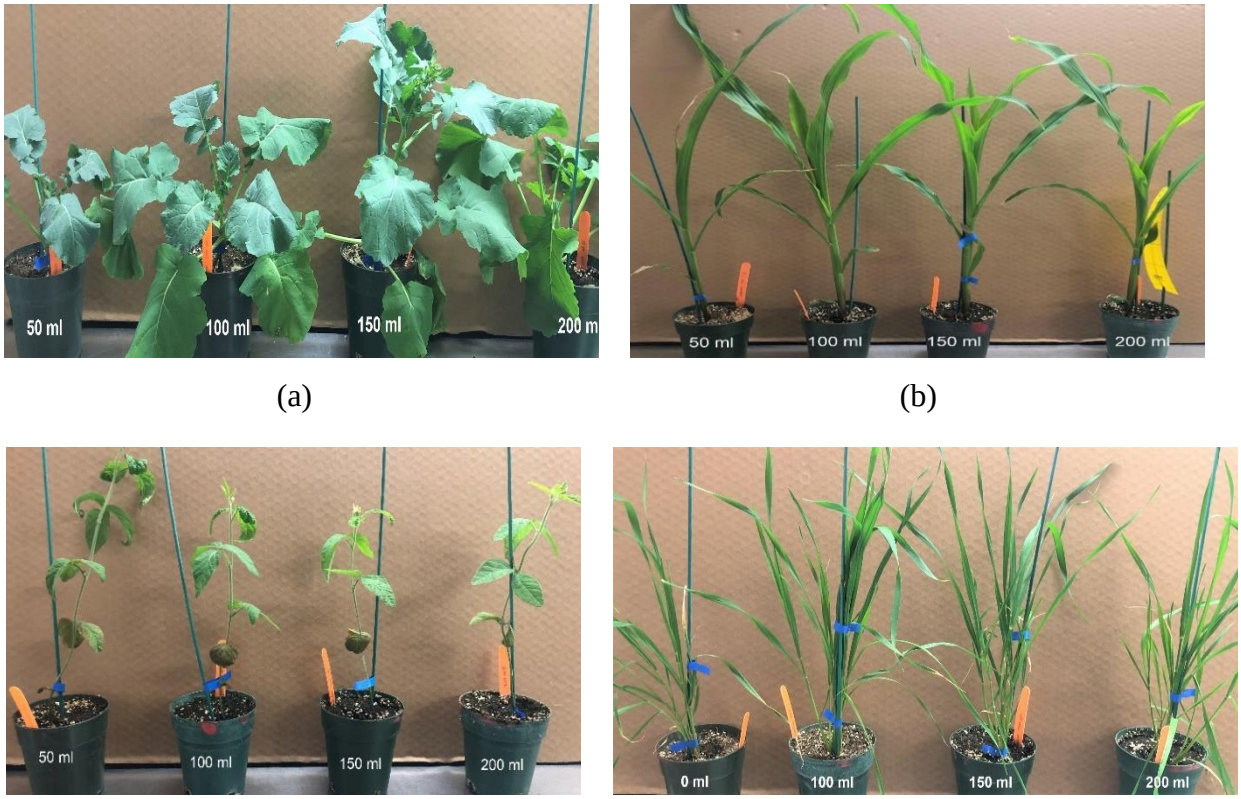


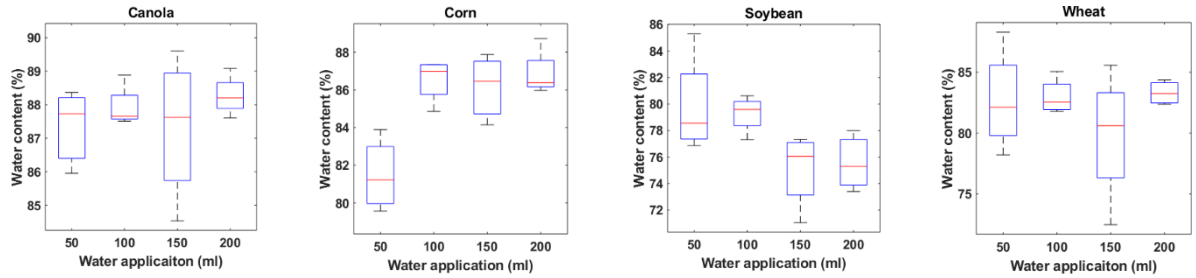
Figure 3-19: Sample pots of (a) canola, (b) corn, (c) soybean and (d) wheat pots subjected to 4 levels water application.

cut and then the fresh weighs were measured (W_{fresh}). All the samples were then placed in the oven-dried at 50º C for 72 to 96 hours to make them completely dry. The dry weight of the samples was then obtained (W_{dry}). Leaf water content was calculated $\mathrm{WC} = (W_{fresh} - W_{dry})/W_{fresh} \times 100$ %. The water content variations of four species as a result of the different application rates were depicted in Figure 3-20 a-d. Also, Figure 3-21 shows the boxplots of the combined water contents. In these cases, the lowest, mean and highest water contents are 71.04%, 83.16%, and 89.6%. After getting the ground truth data, the reflectance data are modeled using GPR as shown in Figure 3-18.

3.6.5. Experiment 5: Leaf P Level Classification in Crops

The experiment was performed on 96 plants consisting of corn, soybean, and wheat each having 32 pots. All the seeds were sowed in the first week of February 2019. The plants were grown in a controlled greenhouse environment in the Agriculture and Agri-Food Canada (AAFC),

Saskatoon, Canada. During the first three weeks, the plants were fertilized with slow-release 15-30-15 (15% N, 30% P, 15% K) fertilizer at a rate of 4 g/L to make sure uniform growth. Later, the



(a) (b) (c) (d)

Figure 3-20: Boxplot of the leaf water content of (a) canola, (b) corn, (c) soybean, and (d) wheat. Here, the horizontal axis represents the rate of water application in mL, and the vertical axis represents the water content.

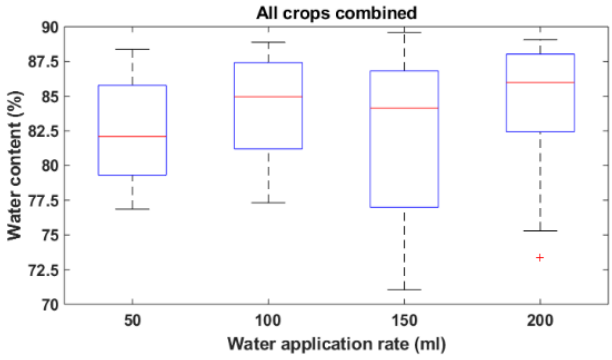
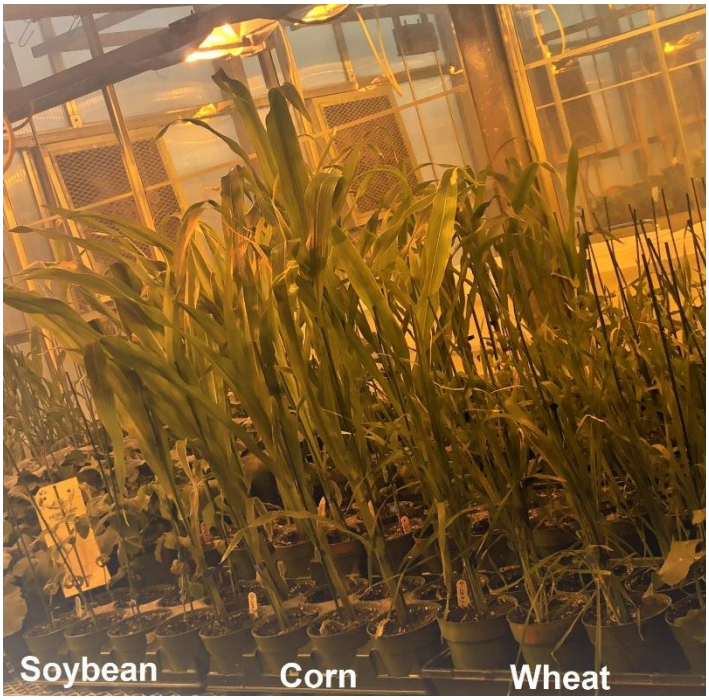


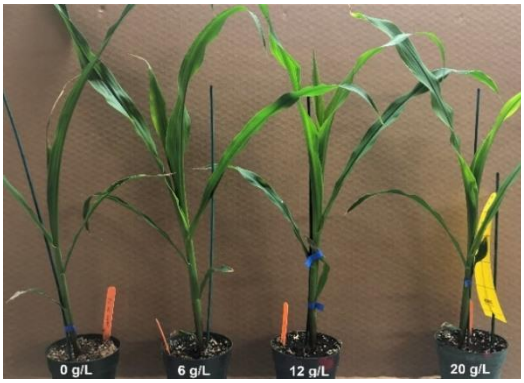
Figure 3-21: Boxplot of leaf water content of combined samples of four species

96 pots were separated into four different sections making an equal number of plants from each species. After that, high P fertilizer (15-30-15) was applied three times in a week at four concentrations 0 g/L, 6 g/L, 12 g/L, and 20 g/L for four separated sections. Figure 3-22 a-d shows the greenhouse setup and three different species of plants. Data collection started on the 19th of March, 2019 and continued till the 26th of March, 2019. Data were collected one leaf at a time, and it took 5 seconds to collect one measurement. Both sensor1 and sensor2 are calibrated each day by capturing the reflectance from a white reflector. In this experiment, a white mirror paper was used as the white surface. Henceforward, reflectance at 12 wavelengths (450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm) was collected from the sample leaf surfaces. During data collection, six plants of soybean and one plant of wheat from 20 g/L fertilizer were not used because they were intoxicated for over-fertilization. So,

finally, a total of 267 leaf samples including 96 leaves of corn, 78 leaves of soybean and 93 leaves of wheat were used for data modeling.



(a)



(c)

(d)

(b)

Figure 3-22: (a) Greenhouse-controlled environment. Sample pots of (b) soybean, (c) corn and (d) wheat, subjected to four levels of P fertilization.

During capturing reflectance, the device was scanned at three different positions around the midrib of a leaf by scanning 15 times. Later, the K-means clustering method is used to cluster the data. From the three clusters, three centroids selected representing the reflectance of the three positions. Then these centroid reflectance data captured at 12 wavelengths are considered as features for classifying the P levels. Here, the intention of doing classification is to build a model by statistically processing the data from the sensor to predict the four categories of responses from

four P treatments (0, 6, 12 and 20 g/L). Next, the data set is randomly shuffled and normalized using z-score [82]. Finally, the KNN classifier is utilized to train a model on 75% of the data and tested on 25%. The process flow diagram is the same as Experiments 1 and 2, shown in Figure 3-12.

In all the five experiments discussed above, the same device, and the same data collection procedure is used. Data modeling pipeline including preprocessing and modeling vary a little bit depending on the purpose. In each case, the procedures that exhibit the best performance are reported.

Chapter 4: Results and Discussion

This chapter describes the results and discussion separately for each experiment using the proposed device and algorithms mentioned in the methodology. Section 4.1-4.5 shows the results of experiments from 1-5, respectively. Also, each section discusses the comparative analysis of the existing methods with the proposed methods.

4.1. Experiment 1

At first, we investigate the reflectance characteristics of the different intensity levels of N fertilized leaf in the greenhouse experiment. To serve the purpose, the average reflectance from each of the four N treatments at 12 wavelengths are plotted in Figure 4-1 and Figure 4-2. Figure 4-1 shows the average reflectance versus wavelength drawn from Sensor1 having six wavelengths in the visible range, while Figure 4-2 illustrates the data value of Sensor2 at the NIR range.

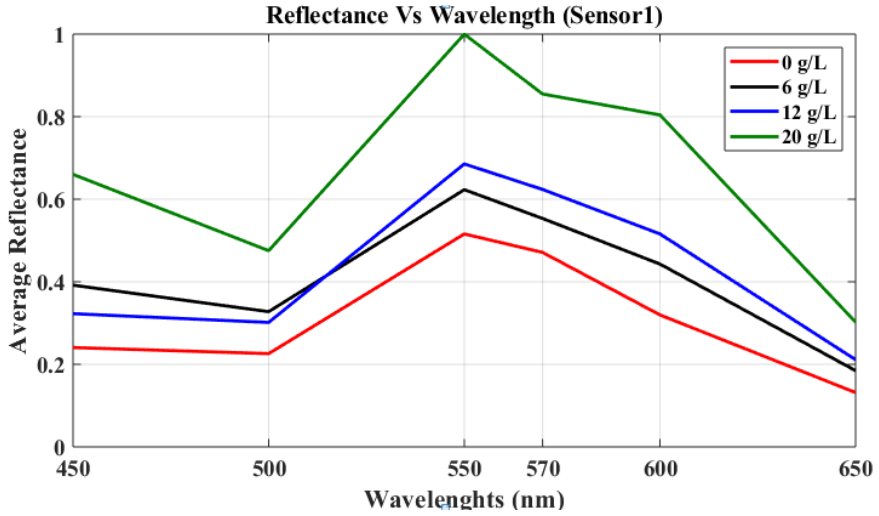


Figure 4-1: Average reflectance versus wavelength (Sensor1) of leaves subjected to four N fertilization regimes under visible range. The Red line indicates the reflectance from 0 g/L plant, the black line represents 6 g/L, blue represents 12 g/L and the green represents 20 g/L N rates. All the reflectance is scaled to the 20g/L reflectance at 550 nm.

The figures show that the average reflectance values of 20 g/L treatments are greater than other treatments. Also, the reflectance curves are distinctive in respect of N treatments, which suggest that the four N levels can be classified.

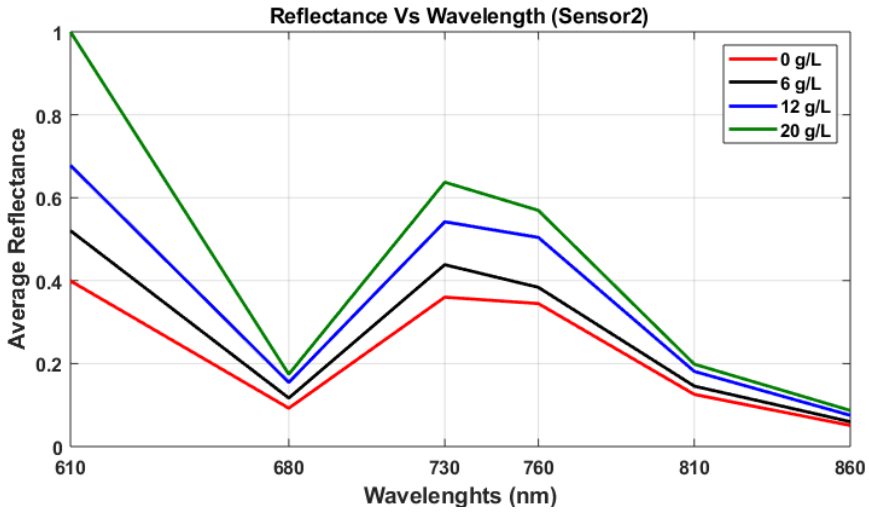


Figure 4-2: Average reflectance versus wavelength (Sensor2) of canola leaves subjected to four N fertilization regimes under the NIR range. The Red line indicates the reflectance from 0 g/L plant, the black line represents 6 g/L, blue represents 12 g/L and the green represents 20 g/L N rates. All the reflectance is scaled to the 20g/L reflectance at 610 nm.

For classifying the captured data, the data set is divided into two parts: training set (75%) and testing set (25%), and the model is trained and tested five times with shuffling data. Finally, the average with standard deviation is reported in Table 4-1. It represents the class-wise (for each category) and overall results on the test set for the greenhouse growing canola plants.

From the classification results shown in Table 4-1, it can be found that Category4 has the best accuracy and Category2 has the least compared to others. One of the reasons might be the fact that there should some samples from Category2 whose reflectance properties are nearly the same as Category1. Later, a binary classification (two-class classification) is performed where Category1 and Category2 are combined, at the same time Category3 and Category4 are combined. It means the classification is performed only in two classes. The results of the binary classification combining Category1 and Category2, and Category3 and Category4, are shown in Table 4-2.

Table 4-1: Category wise/class wise results on the test set for the greenhouse experiment five times running. The average ± standard deviation is reported.

	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
Category1	92.8 ± 1.5	81.2 ± 2.1	92.8 ± 1.2	92.1 ± 1.1	86.6 ± 2.2
Category2	71.4 ± 1.1	99.9 ± 0.2	71.4 ± 2.0	100 ± 0.0	83.3 ± 2.5
Category3	90.9 ± 3.1	99.9 ± 0.1	90.9 ± 1.6	100 ± 0.0	95.2 ± 1.6
Category4	99.9 ± 0.1	81.2 ± 1.5	99.9 ± 0.3	92.3 ± 1.3	89.6 ± 1.8
Total	88.4 ± 3.0	90.6 ± 2.3	88.8 ± 2.9	96.1 ± 2.0	88.7 ± 2.6

Category 1, Category 2, Category 3, and Category 4 represent four N treatments (0, 6, 12, and 20 g/L) in the greenhouse experiment.

Table 4-2: Results on the test set for the greenhouse experiment for Binary classification (Category1+Category2) and (Category3+Category4) for five times running. The average ± standard deviation is reported.

Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
94.2 ± 1.8	90.9 ± 2.1	100.0 ± 0.0	90.9 ± 1.9	95.2 ± 1.8

The results reveal that the accuracy, precision, recall, specificity, and F1-score have improved compared to distinct classes classification. Moreover, the standard deviation is decreased which means more consistent result has been acquired. In this section, some other states of art machine learning algorithms such as Decision Tree, Support Vector Machine (SVM), Ensemble Bagged Tree are compared with the results obtained from the KNN algorithm. For this comparison, the same procedure is applied, and the testing accuracy is shown in Table 4-3. Among them, KNN showed the best accuracy (underlined). The next best result is shown by the Ensemble Bagged Tree followed by SVM and Decision Tree. So, using the KNN algorithm for creating a model for this type of data is the best choice.

The comparison results shown in Table 4-3 reveals that KNN performs better than SVM, Decision Tree, and Ensemble Bagged Tree. To the best of our knowledge, the performance of a classifier over others depends on the data on which the model is built. However, in some cases,

KNN shows better performance, when features of the dataset are much less than training samples [93]. Previously, in some articles, researchers showed similar comparative findings like ours. In a study of classifying acoustic signal, the comparison of SVM and KNN algorithm reports better performance than SVM [94]. Moreover, SVM is outperformed by KNN in weather classification data [93]. In addition, with comparison to Decision Tree for classifying Irish national forest inventory data, KNN achieves better accuracy [95].

Table 4-3: Comparison of the test results of different machine learning algorithms.

Training Algorithms	Testing Accuracy (%)
Decision Tree	65.1 ± 2.7
Support Vector Machine (SVM)	65.0 ± 2.0
Ensemble Bagged Tree	73.3 ± 2.2
K-Nearest Neighbor (KNN)	88.4 ± 3.0

In this research, the reflectance at 12 different wavelengths has been used as features to train the proposed method. From the feature weight optimization analysis, as described in the methodology, the features are ranked based on the weights gained from PSO. The whole process is run five times. Using the weights in five runs, box plots are plotted in Figure 4-3. From the figure, the importance of features can be ranked based on the median value of the weights after multiple runs. The most important feature is found to be 450 nm followed by 500, 860, 680, 570, 650, 600, 550, 760, 810, 730, and 610 nm in descending order.

The proposed multi-spectral sensors, along with the suitable machine learning algorithm, have been shown to be very cost-effective for determining leaf N status. This is the first approach to design a low-cost system for measuring the leaf N level using the reflectance at 12 wavelengths. The proposed method is verified by the control environment. One of the limitations of this experiment is that the actual contents of N are not measured by the correlation, which will be the next phase of the project. Applying the technique to other crop species at various growth stages has great potential for future research.

In this research, a low-cost and portable multi-spectral sensing system is proposed, which can capture the reflectance at 12 wavelengths ranging from 450nm to 860 nm. The KNN algorithm

has been employed to model the collected data set. By applying different algorithms, the best testing accuracy of 88.4% was found for the greenhouse. The testing results also reveal that the proposed low-cost multispectral sensor array can measure the leaf's N level with decent accuracy.

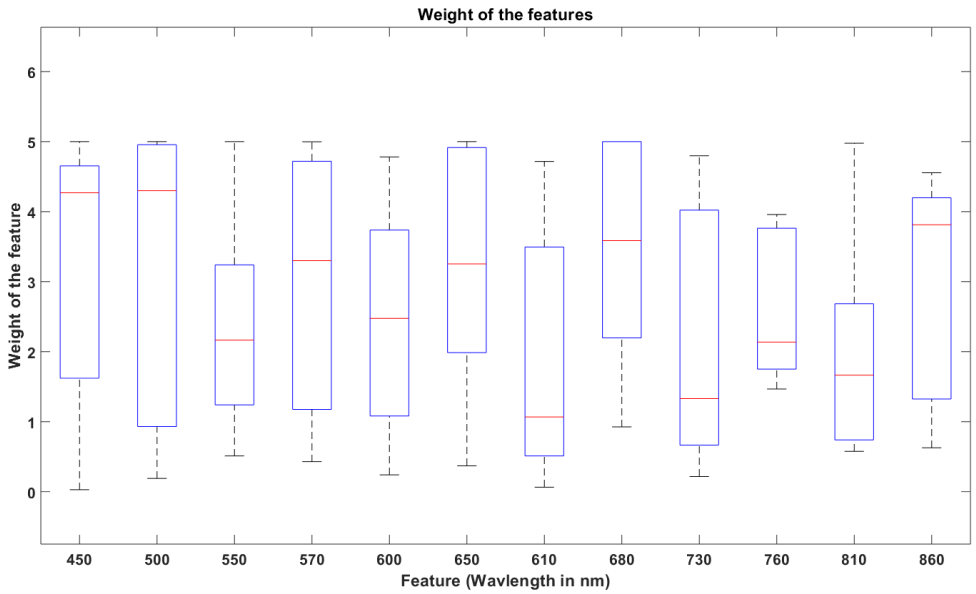


Figure 4-3: Box plots of the weights achieved by the features in 5 runs. Here, the 12 features are 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm.

4.2. Experiment 2

The field experiment results are summarized in Table 4-4. It shows the 25% test set results of two categories (low N and high N). The average accuracy of the test set is 79.2%, which is as less compared to the test results achieved in the greenhouse experiment. Although the field experiment is all about classifying two categories of N, the combined dataset has high variance as it consists of 42 canola cultivars. In some cases, the model might get confused identifying each category for similar responses from a different category. As a result, the model gets disorderly, in some cases, to differentiate two N levels.

Table 4-4: Category wise results on the test set for the field experiment five times running. The average ± standard deviation is reported.

Accuracy (%)	**Precision (%)**	**Recall (%)**	**Specificity (%)**	**F1-Score (%)**
79.2 ± 2.5	80.1 ± 2.7	79.6 ± 2.1	80.2 ± 2.3	79.3 ± 2.4

In this section, some other states of art machine learning algorithms such as Decision Tree, Support Vector Machine (SVM), Ensemble Bagged Tree are compared with the results obtained from the KNN algorithm. For this comparison, the same procedure is applied, and the testing accuracy is showed in Table 4-5. Among them, KNN showed the best accuracy (underlined). The next best result is shown by the Ensemble Bagged Tree followed by SVM and Decision Tree.

Table 4-5: Comparison of the test results of different machine learning algorithms.

Training Algorithms	Testing Accuracy (%)
Decision Tree	73.6 ± 2.6
Support Vector Machine (SVM)	70.1 ± 2.1
Ensemble Bagged Tree	75.0 ± 2.3
K-Nearest Neighbor (KNN)	79.2 ± 2.5

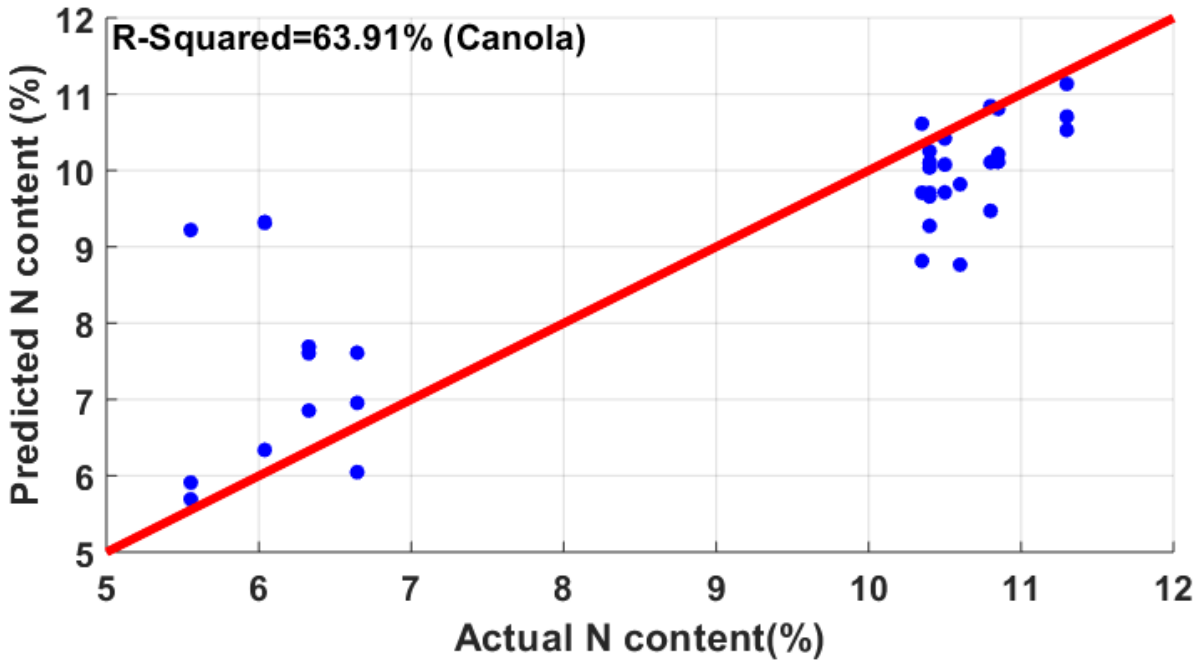
The proposed method is verified in the field plants with an accuracy of 79.2%. One of the limitations of this work is that the actual contents of N are not measured by the correlation, which will be the next phase of our project.

4.3. Experiment 3

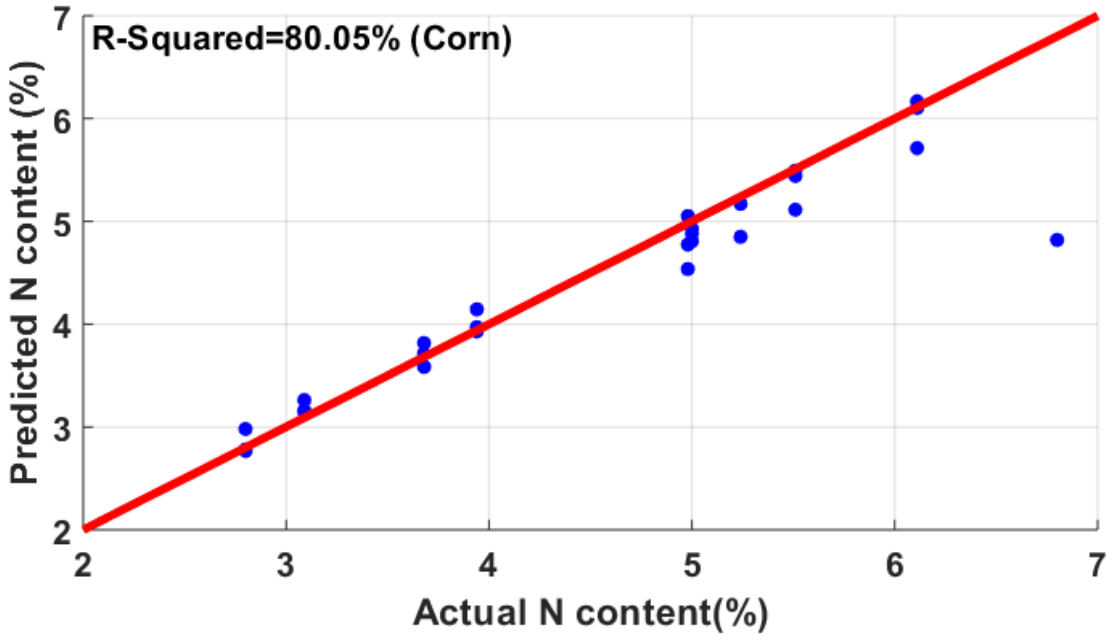
In this experiment, first, a t-test was performed before regression modeling. With a 5% p-value, the t-test selected all the 12 variables as important features. After normalizing, as discussed in the methodology section, GPR is used for correlating the reflectance data with crop measurements. In this section, the average results of 5-fold cross-validation are reported. The regression is performed on individual species wise (Figure 4-4 a-d) as well as a combination of them (Figure 4-5). In the N experiment, the best correlation was found in soybean (R^2 =82.29%) followed by corn (R^2 =80.05%). Wheat showed the least correlation among all having an R^2 of 63.21%. The coefficient of determination shown by canola was 63.91%. Another regression analysis was performed combining all the samples of four species. The combined model showed an overall R^2 of 73.96% with RMSE of 1.13. Figure 4-5 shows the correlation plot, and the summary of the results is shown in Table 4-6.

Table 4-6: 5-fold cross-validation results of the N experiment. Here, the average of the metrics, mentioned in Table 3-2, from the five folds are reported.

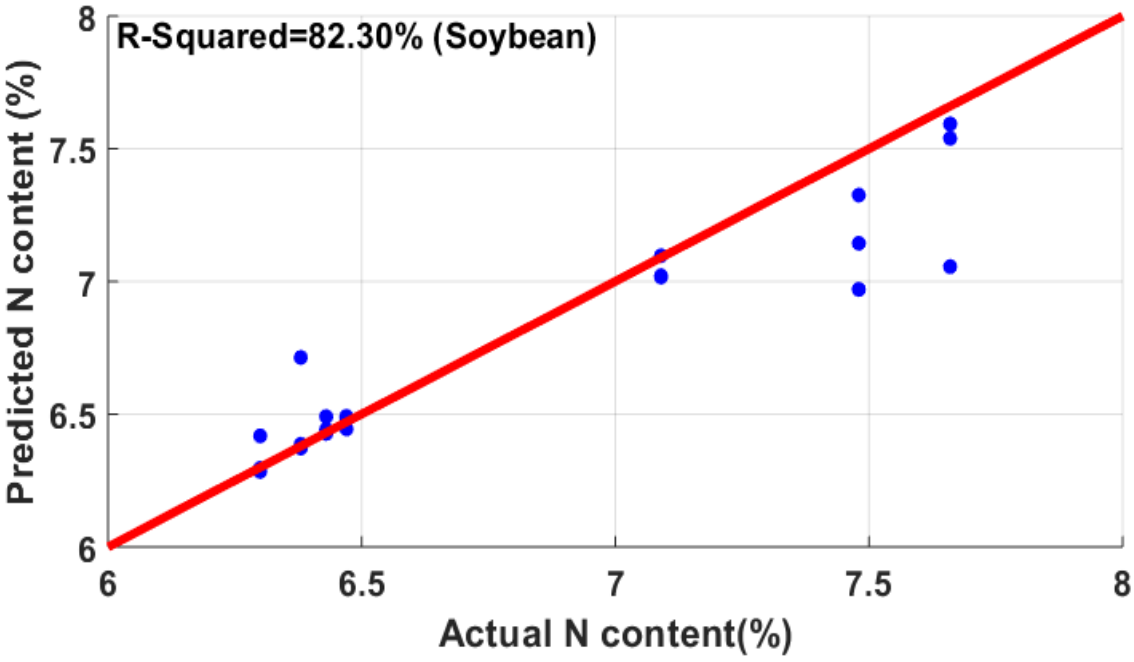
Plant species	R^2 (%)	RMSE	MAE
Canola	63.91	1.28	0.87
Corn	80.05	0.50	0.31
Soybean	82.29	0.21	0.12
Wheat	63.21	0.57	0.37
All crops combined	73.96	1.13	0.72



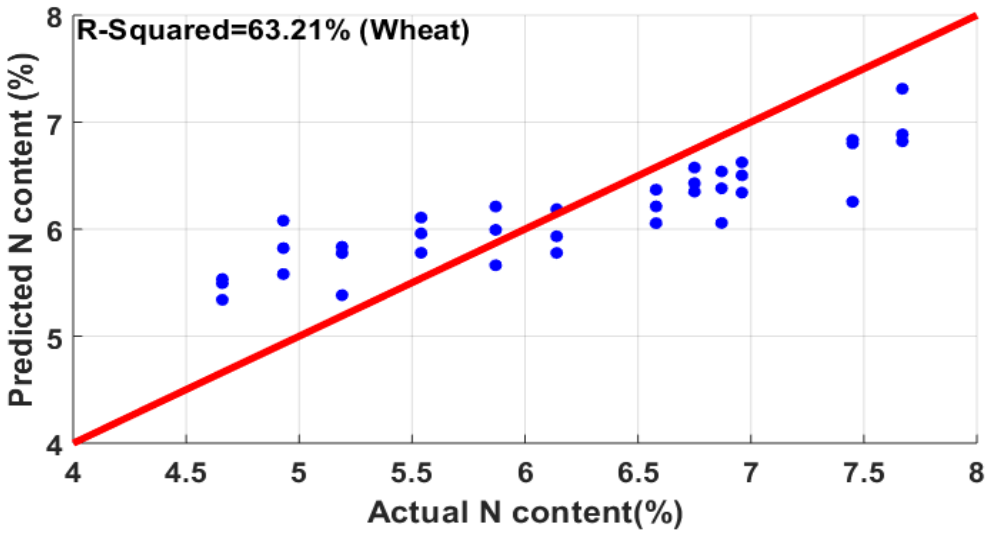
(a)


R-Squared=80.05% (Corn)
Predicted N content (%)
Actual N content(%)
2
3
4
5
6
7

(b)


R-Squared=82.30% (Soybean)
Predicted N content (%)
Actual N content(%)
6
6.5
7
7.5
8

(c)



(d)

Figure 4-4: N estimation in (a) canola, (b) corn, (c) soybean, and (d) wheat. These figures show the correlation between the predicted N content and actual N content. Here, soybean shows the best correlation of 82.3%, whereas wheat shows the least (63.21%).

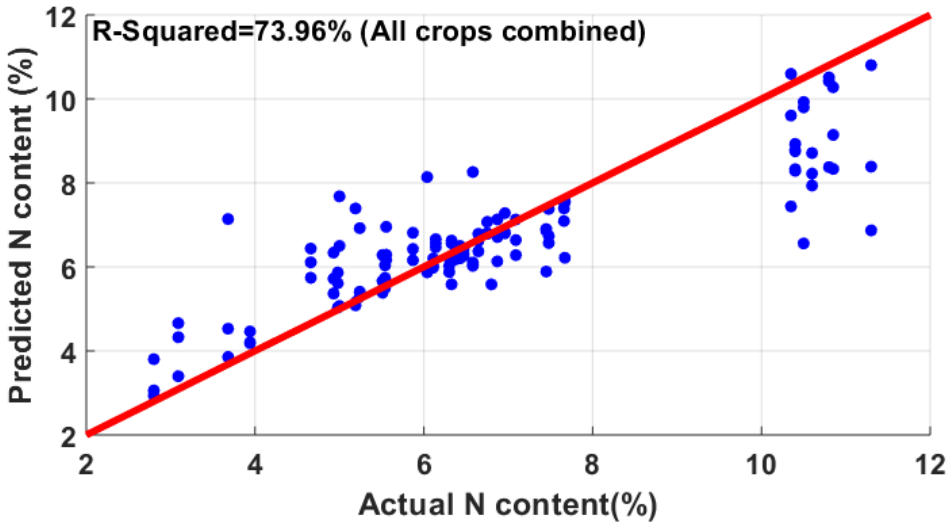


Figure 4-5: N content estimation combining canola, corn, soybean, and wheat. Here, all the samples from these four species were combined during performing correlation.

In this part, the importance of the features (wavelengths) is determined by calculating the increase of MSE of 5-fold cross-validation after each featured being removed. For example, in the case of N determination, when the feature of 810 nm is removed and the whole regression analysis

is performed with 11 other features, the average mean square error of 10 runs is 1.7311. On the other hand, for the 600 nm wavelength feature, the MSE is 1.54, which is much less than 810 nm. That means, the decrease of the correlation performance by removing the 810 nm wavelength as the feature is much more significant than 600 nm. So, reflectance at 810 nm wavelength is more important than 600 nm in respect of determining N. After performing the same technique on 12 wavelengths individually, the importance of wavelengths is ranked and shown in Figure 4-6. It is found that the most three significant wavelengths, in this analysis, are found to be 810 nm, 650 nm, and 610 nm.

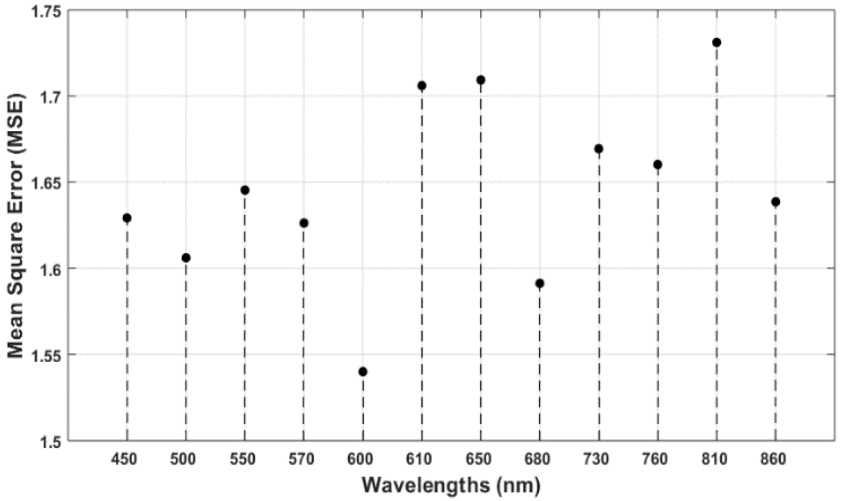


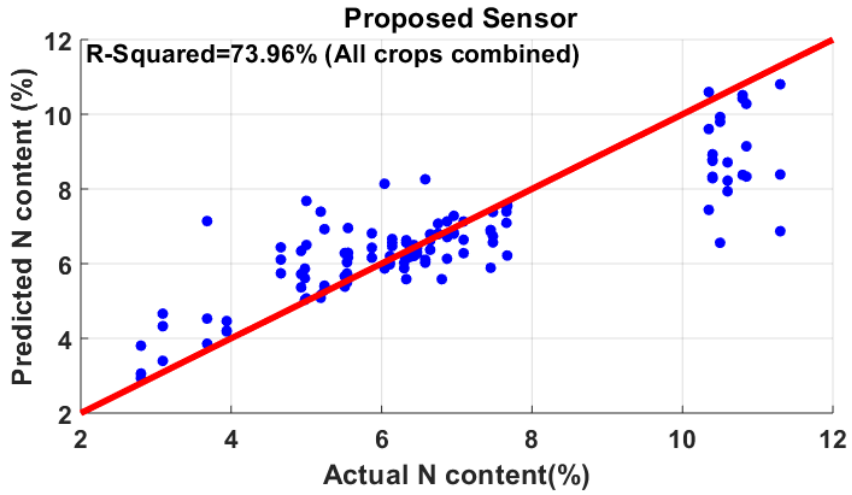
Figure 4-6: Important wavelengths for N are shown, based on the increase of MSE after the wavelength being removed. The most three significant wavelengths, in this analysis, are found to be 810 nm, 650 nm, and 610 nm.

In this section, the features of other devices, used in previous works related to N content estimation in leaves, are compared with the proposed multispectral sensor (Table 4-7). In several previous publications [10][12][96] that utilized hyperspectral camera, found good correlation around 86%-92% between images and leaf N content. The other popular devices are Field Spec 3 with R^2 77-86% [19], imagery from Quick Bird with R^2 79-83% [21], Green Seeker with R^2 of 57-74% [20], Multiplex with $R^2$of 73%-86% [97], and SPAD with R^2 of 60.21%. All the above-mentioned device accuracies are from in-field experiments, whereas the proposed sensor is tested on the greenhouse control environment. However, most of the devices except hyperspectral camera, and Field Spec 3, are based on either chlorophyll (SPAD, Multiplex, atLEAF) or relative greenness (Green Seeker, Quick Bird imagery) instead of N. Moreover, the proposed sensor is

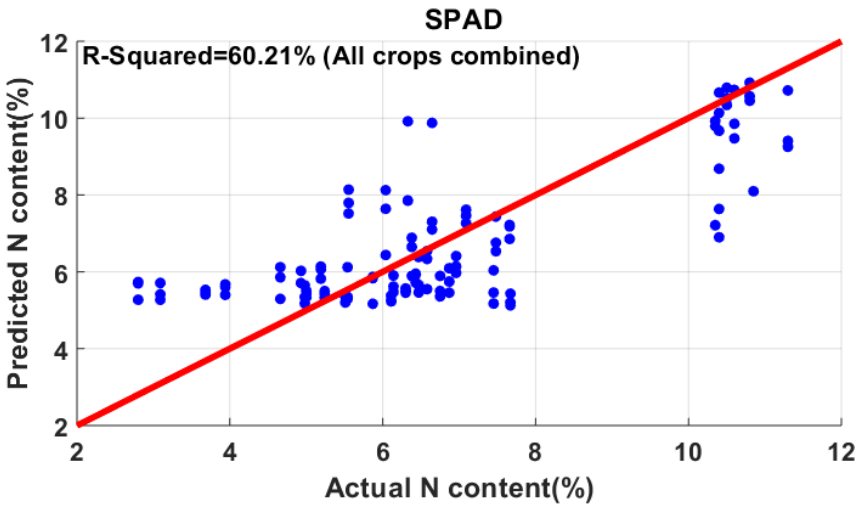
useful in estimating N as, it operates on more features (12 wavelengths) than SPAD (two wavelengths), atLEAF (two wavelength), Green Seeker (two wavelengths), and Multiplex (four wavelengths). Although, hyperspectral camera, and Field Specs are the most accurate in N determination, these devices are very expensive: hyper spectral camera ~$15,000-$50,000, Field Spec series around ~$10,000-$20000, Green Seeker ~$700, atLEAF ~$250, and SPAD ~$1,500-$2,500; whereas the total estimated cost of the proposed prototype, including all the components is $150. By incorporating the manufacturing labor cost and overhead cost, the approximate estimation of the device might be $200. The complete breakdown of the cost is shown in Table 3-1. However, this is the first version of the prototype, which can be further modified to reduce the cost. For example, without using the raspberry pi (control circuit) development board, the control circuit can be implemented in a custom-made Printed Circuit Board (PCB), according to the design requirements. Moreover, the cost of the device will be significantly reduced if it is mass manufactured. On a different note, the proposed sensor is comparatively light weight (350 g) than other devices (Hyperspectral camera: 1.3 kg - 4.5 kg, FieldSpec Series: 5.4 kg). Furthermore, the technologies like hyperspectral camera, and satellite imagery are normally used in the developed countries; and for the farmers from low resource countries, these expensive devices are out of question. On top of that, as the device is cheap, and portable; it can be manufactured in large quantity and deployed in sensor arrays more flexibly that are unfeasible for expensive, and heavyweight devices. It can also be mounted to custom IoT (Internet of Things) platforms for wireless monitoring of the N level. So, the proposed sensor is very effective as a low-cost, portable, quick, and light-weight device for monitoring N content.

SPAD (Soil Plant Analysis Development) is a commercially used meter for indirect measurement of N. In this work, we also took measurements using SPAD. The intent of performing this analysis is to see how profoundly SPAD readings correlate with N compared to the proposed sensor. For this analysis, the same methodology including preprocessing, modeling, 5-fold cross-validation was applied for both. It is to be noted that the dataset used for the comparison is the one that has the combination of all four types of plants (121 leaves). It was observed that SPAD predicted the N contents with a R^2 of 60.21%, whereas the proposed sensor showed 73.96% (Figure 4-7 a-b). It is to be noted that the atLeaf chlorophyll meter is another cheap alternative version of SPAD. The technical difference between those is, SPAD uses two wavelengths 650 nm and 940 nm, whereas atLEAF works at 660 nm and 940 nm [98]. These wavelengths (650 nm or 660 nm)

are basically sensitive to chlorophyll. In contrast, the proposed sensor works on 11 more wavelengths (450 nm, 500 nm, 550 nm, 570 nm, 600 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm) in addition to 650 nm. As it is adding more features to the statistical learning process, the proposed sensor performs better than SPAD in determining N content in leaves.



(a)



(b)

Figure 4-7: Comparison between (a) proposed sensor, and (b) SPAD. The proposed sensor shows a better correlation (73.96%) with N than SPAD (60.21%).

Cost-effective and non-destructive estimation of these contents is challenging. Most of the noninvasive approaches include expensive equipment. In this research, a low cost, portable optical sensor is proposed that can be effectively modeled with an appropriate regression algorithm to

determine leaf N. Using the multispectral sensor, we have correlated leaf reflectance at 12 wavelengths with crop measurements. The 5-fold cross-validation results of the GPR model reveal that the best correlation of N is found in soybean (R^2 of 82.29%). After comparing the N estimation result with a commercially used device, SPAD; we have found that the proposed multispectral sensor shows a better correlation with N than SPAD (R^2 of 60.21%). It is worthwhile to mention that the overall cost of our proposed sensor is $200, which is very cheap compared to other technologies. Also, the accuracy of the device can be further improved by experimenting on more samples and making a robust model for N estimation. Besides, there are a lot of future scopes to use this device to correlate other nutrients such as P and K. In addition, it will be interesting to see the correlation performance of the device in other crop species which have great potential for future work.

Table 4-7: Comparison of the existing techniques for N sensing

Features	**SPAD**	**Multiplex**	**FieldSpec Series**	**Green Seeker**	**QuickBird Satellite Imagery**	**Hyper Spectral Camera**	**Our Proposed System**
Basis	Chlorophyll	Chlorophyll and polyphenol	Nitrogen	Greenness	Greenness	Nitrogen	Nitrogen
Technique	Transmittance	Fluorescence	Reflectance	Reflectance	Reflectance	Hyper Spectral Image	Reflectance
No of Wavelength	Two (640 nm, 940 nm)	Four at excitation (375 nm, 450 nm, 530 nm, 630 nm) Four at Detection (447 nm, 590 nm, 665 nm, 735 nm)	200-500	Two (660 nm, 970 nm)	Four bands 450-900 nm, 450-520 nm, 520-600 nm, 630-690 nm, 760-900 nm	400-500	12 (450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm)
Cost (USD)	$1,500 -$2,500	$1,500-$2,500	~$10,000-$60,000	$700	---	$25,000-$10,0000	$200

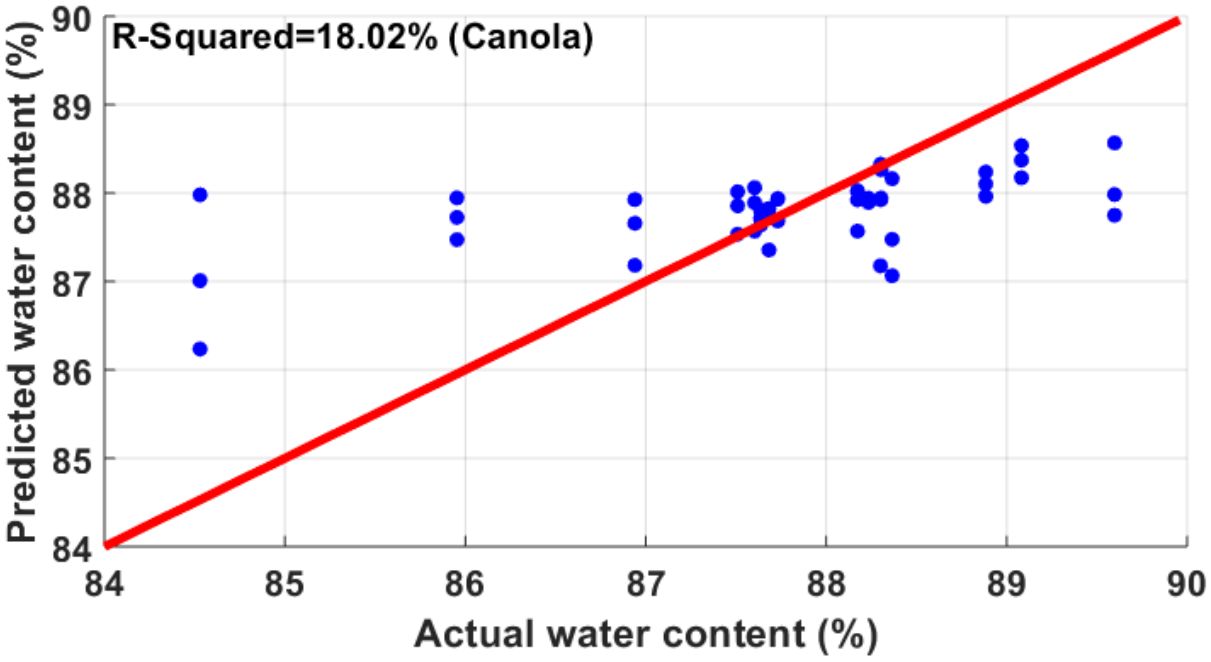
4.4. Experiment 4

The 5-fold cross-validation results of water content estimation are shown in Table 4-8. Observing the results, it is found that, the reflectance from canola does not correlate well with the leaf water content having an R^2 of 18.02%. In the case of wheat, there were some outliers, which were outside the range of two standard deviations from the median. Those outliers have been removed and correlated, showing a correlation of 64.58%.

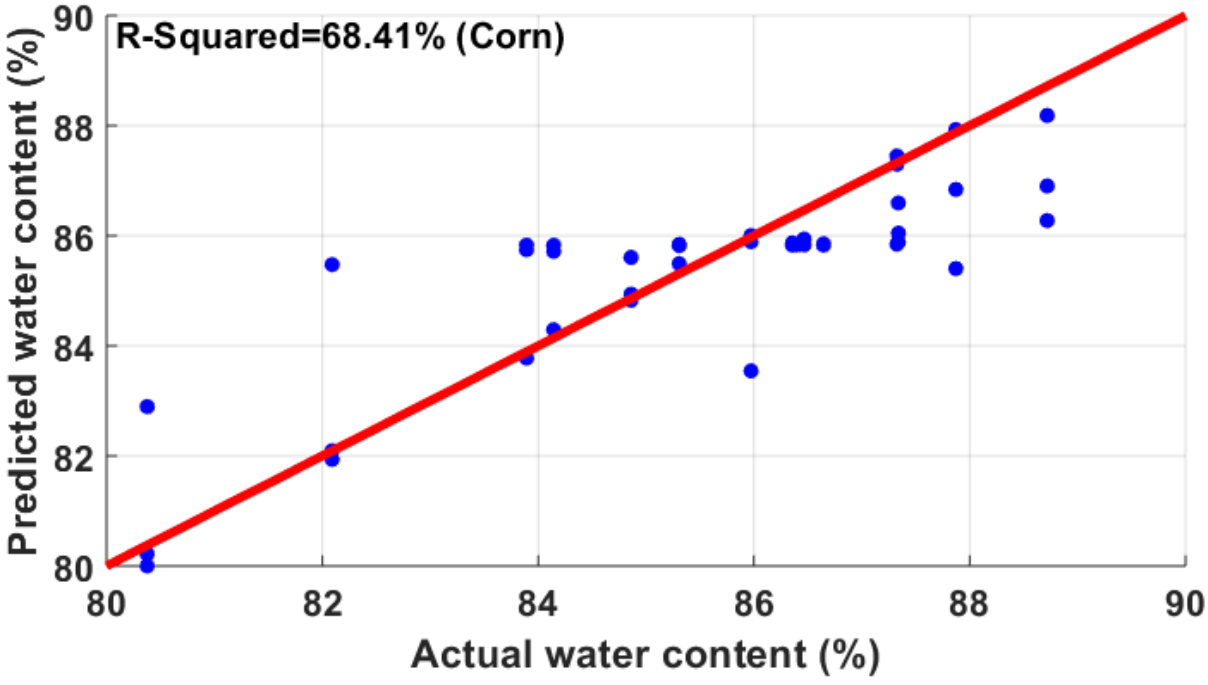
Table 4-8: 5-fold cross-validation results of water. Here, the average of the metrics from the five folds is reported.

Plant species	$\boldsymbol{R^2}$ (%)	RMSE	MAE
Canola	18.02	1.06	0.76
Corn	68.41	1.17	0.75
Soybean	46.38	3.50	2.11
Wheat	64.58	1.16	0.85
All crops combined	46.08	3.97	2.75

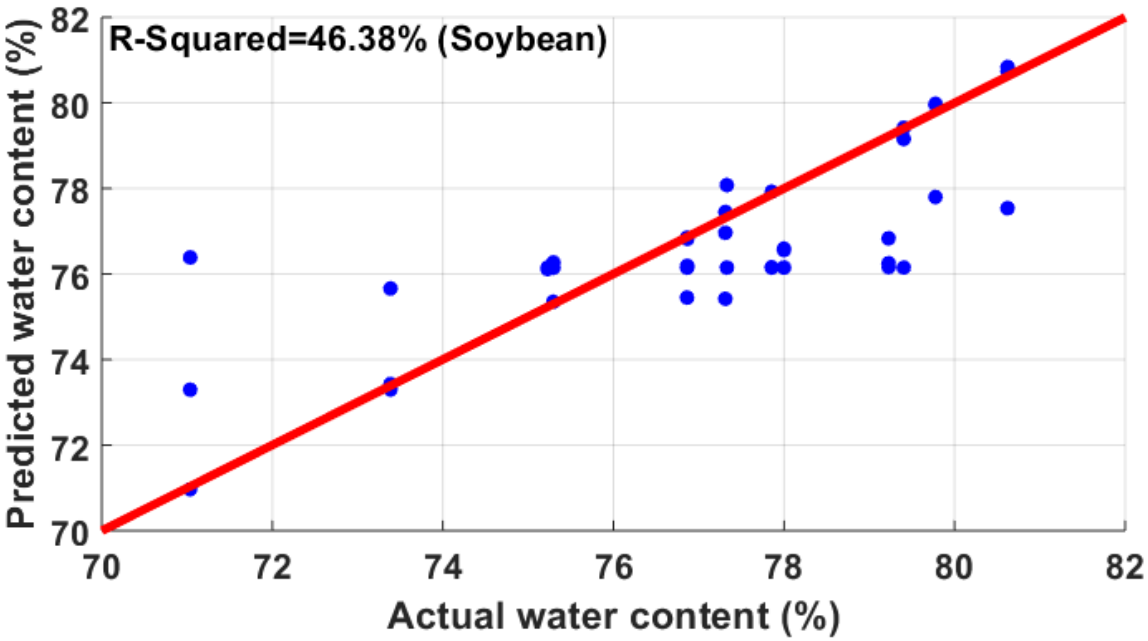
The best correlation is observed in corn leaves (68.41%). Although, individual R^2 of canola and soybean is less, a combination of canola and soybean shows a better estimation of 61.08%. The regression analysis combining corn, wheat, canola, and soybean shows an overall co-efficient of determination of 46.08%. So, unlike N correlation, the combined result in water estimation is not found to be satisfactory. Species wise actual vs predicted water content plots are shown in Figure 4-8 a-d.


R-Squared=18.02% (Canola)
Predicted water content (%)
Actual water content (%)
84
85
86
87
88
89
90

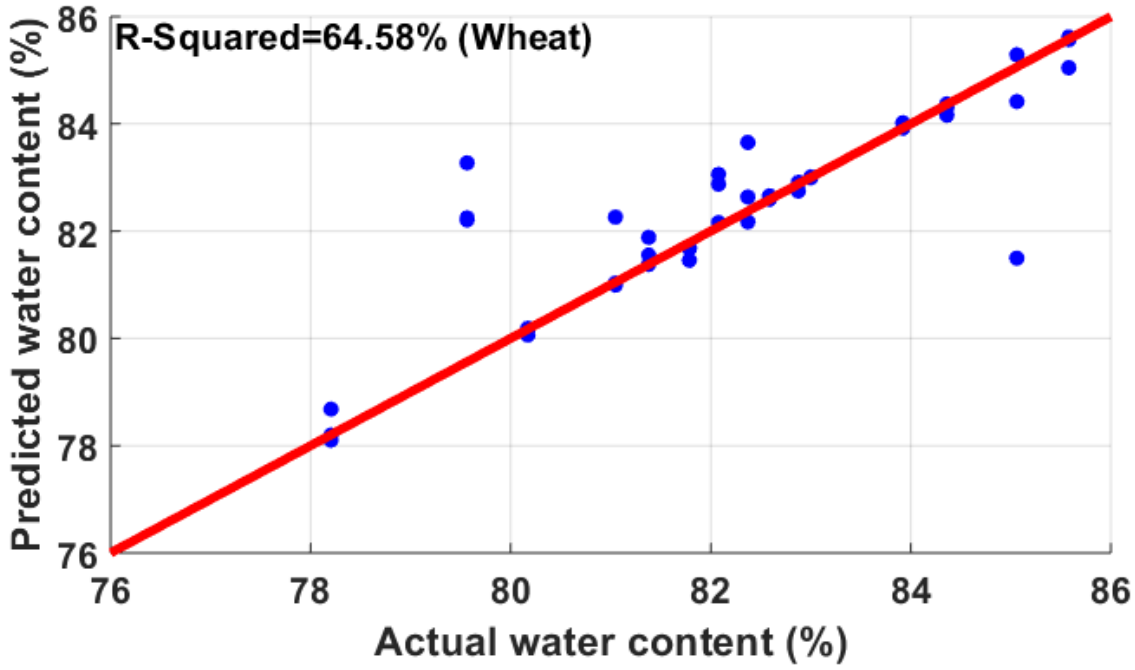
(a)


R-Squared=68.41% (Corn)
Predicted water content (%)
Actual water content (%)
80
82
84
86
88
90

(b)



(c)



(d)

Figure 4-8: Water content estimation in (a) canola, (b) corn, (c) soybean, and (d) wheat. These figures show the correlation between the predicted water content and actual water content. Here, corn shows the best correlation of 68.41%, wheat shows 64.58%; whereas canola and soybean do not correlate well.

Analysis for determining wavelength importance is performed for water contents estimation in leaves. It was found that when the 760 nm wavelength is omitted and the Gaussian process regression algorithm is applied with 11 other features, the average (after 10 runs) MSE of 5-fold cross-validation is calculated to be 22.9037. Similarly, the MSE of the other features is measured and shown in Figure 4-9. The most significant wavelength for water content determination is found to be 760 nm followed by 730 nm and 860 nm in descending order. The significance of the other wavelengths is observed to be nearly similar.

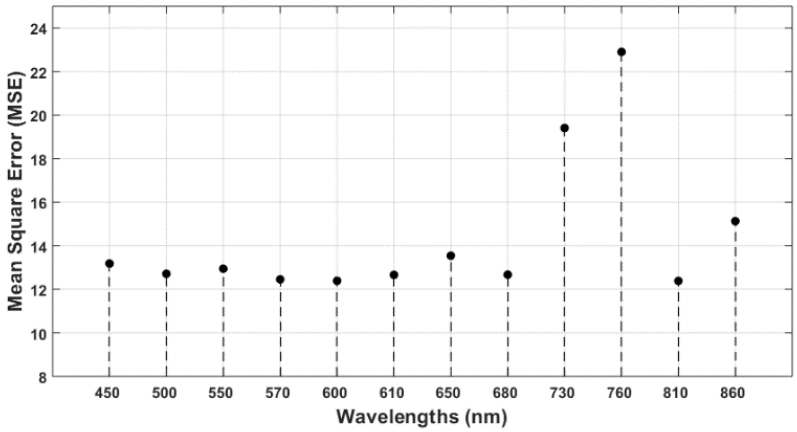


Figure 4-9: Important wavelengths for water are shown, based on the increase of MSE after the wavelength being removed. The most significant wavelength for water content determination is found to be 760 nm followed by 730 nm and 860 nm in descending order. The significance of the other wavelengths is observed to be nearly similar.

In this study, N estimation results are better than water estimation in terms of correlation. One of the main reasons is that plant water content is mainly sensitive to thermal or short-wave infrared regions, but the proposed sensor operates in the visible, and NIR regions. So, more spectral bands in the short-wave infrared regions are perhaps necessary to improve the water correlation results. Moreover, the values of the water content in the samples are in a small range from 60% - 90%, for which the model is not robust enough to exhibit good correlation. So, the utility of the proposed device in water content estimation needs further consideration.

4.5. Experiment 5

First, the reflectance characteristics of the different intensity levels of P are investigated for corn, soybean, and wheat. So, the average reflectance from each of the four P treatments at 12

wavelengths are plotted in Figure 4-10. In this case, the four types of responses resulting from four P treatments (0, 6, 12 and 20 g/L) are shown for sensor1 (left) and sensor2 (right). The figure reveals that the average reflectance increases with corresponding P treatments at almost every wavelength. From 450 nm to 650 nm, the average reflectance for 20 g/L treatment is highest, while 0 g/L exhibits the lowest. However, different reflectance profile is found between 680 nm to 860 nm for different crop species. For example, highest reflectance is observed for 6 g/L treatment in Corn, lowest is seen for the same treatment in Soybean. However, these four reflectance categories representing four levels of P responses distinguishable at most of the wavelengths, especially 450 nm, 500 nm, 730 nm, and 810 nm.

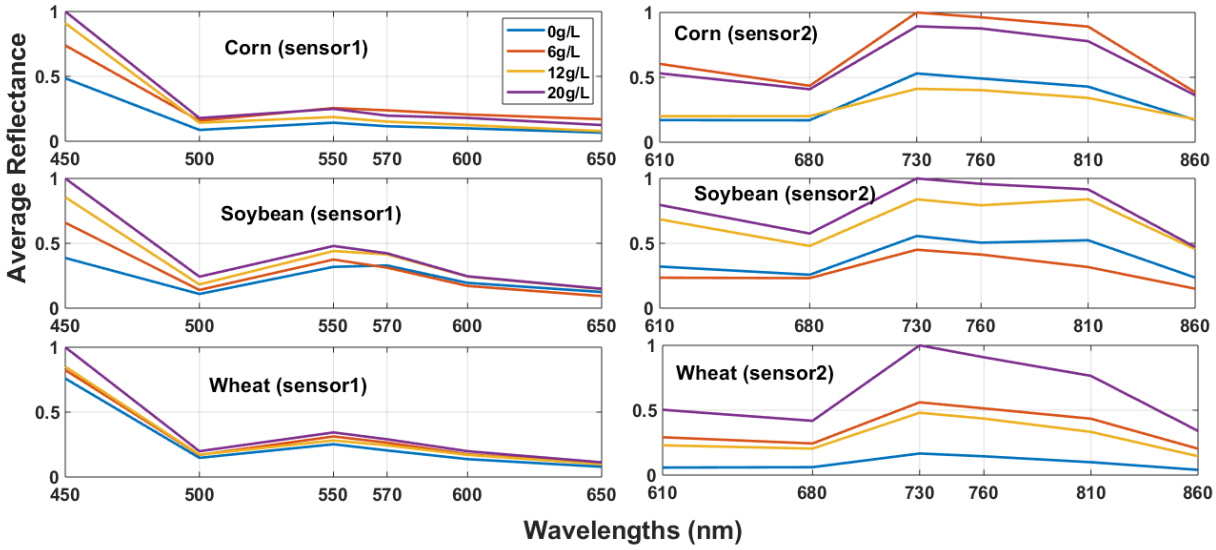


Figure 4-10: Average reflectance versus wavelength (sensor1 in left and sensor2 in right) of leaves subjected to different levels of P fertilization. The blue line indicates the reflectance from 0 g/L plant, the red line represents 6 g/L, yellow represents 12 g/L and the violet represents 20 g/L P rates. For sensor1 (left) all the reflectance values are scaled to the 20 g/L reflectance at 450 nm, and to 730 nm for sensor2. For corn in sensor2, all are scaled to 6 g/L at 730 nm.

In this section, the results of classifications are reported. The data set is divided into two parts: the training set (75%) and testing set (25%). This train-test split is performed five times randomly, and each time a model is built by the KNN algorithm, using the training set. After that, the trained models are tested on the corresponding test sets. Finally, the averages and the standard deviations of the validation metrics for five test sets are reported in Table 4-9 to Table 4-12. Here, Table 4-9, Table 4-10, and

Table 4-11 represent the four-class classification results of corn, soybean, and wheat respectively. Also, Table 4-12 shows the results of a combined three species. Both class-wise (for each category) and overall (average of each class results) results are reported in each table. In the following tables from Table 4-9 to Table 4-12, Category 1, Category 2, Category 3, and Category 4 represent four P treatments (0, 6, 12, and 20 g/L) in the greenhouse experiment.

Table 4-9: Classifying P responses for Corn

	Accuracy	**Precision**	**Recall**	**Specificity**	**F1-Score**
Category1	62.9±8.1	59.3±15.4	62.9±8.1	86.7±7.4	59.3±4
Category2	74.7±4.3	71.3±12.5	74.7±4.3	88.8±4.9	72.7±6.5
Category3	63.7±3.8	74.5±8.8	63.7±3.8	92.1±2.7	68.4±4.4
Category4	69.7±4.6	70.6±14.1	69.7±4.6	90.2±5.6	69.4±6.9
Overall	67.7±5.2	68.9±12.7	67.7±5.2	89.4±5.1	67.4±5.4

Table 4-10: Classifying P responses for Soybean

	Accuracy	**Precision**	**Recall**	**Specificity**	**F1-Score**
Category1	84.6±9.2	86.7±12.6	84.6±9.2	92.9±6.8	84.7±4.1
Category2	71.7±9.5	78.3±11.1	71.7±9.5	90.8±2.5	74.5±8.5
Category3	86.3±9.6	68.3±17.4	86.3±9.6	85.3±9.1	74.6±8.6
Category4	51.3±9.6	89.3±15.4	51.3±9.6	99.3±1.1	63.7±4.3
Overall	73.5±9.5	80.7±14.1	73.5±9.5	92.1±4.9	74.4±6.4

Table 4-11: Classifying P responses for Wheat

	Accuracy	**Precision**	**Recall**	**Specificity**	**F1-Score**
Category1	83.7± 9.1	71.1±17.5	83.7±10.7	88.3±7.4	75.3±10.1
Category2	65.1±3.3	66.1±4.1	65.0±7.8	87.5±2.8	65.5±3.1
Category3	73.7±13.0	78.2±14.6	73.7±14.7	92.3±6.5	74.6±9.7
Category4	62.5±23.2	70.1±21.3	62.5±28.1	92.7±6.2	61.5±12.1
Overall	71.2±12.1	71.4±14.4	71.2±15.9	90.2±5.7	69.2±8.7

Table 4-12: Classifying P responses for all the three species combined

	Accuracy	Precision	Recall	Specificity	F1-Score
Category1	68.2±3.9	61.6±5.1	68.2±3.9	84.7±2.5	64.5±1.9
Category2	68.1±5.9	63.3±4.8	68.1±5.9	85.9±2.6	65.5±4.1
Category3	66.9±6.3	70.4±5.1	66.9±6.3	89.3±3.2	68.6±5.0
Category4	58.1±10.2	69.1±8.3	58.1±10.2	93.7±1.3	62.6±7.1
Overall	65.3±6.6	66.1±5.8	65.3±6.6	88.4±2.4	65.3±4.6

Now, the results by using KNN algorithms are compared with other states of art machine learning algorithms- Decision Tree, Support Vector Machine (SVM), and Ensemble Bagged. It is to be noted that the exact same procedure is implemented for fairness of the comparisons. The results are compared based on the average accuracy and standard deviation; reported in

Table 4-13. The comparison reveals that KNN showed the best accuracy among all of them. This justifies the use of KNN for classifying P levels.

Among the all reported result from Table 4-9 to Table 4-12, the best results are found in soybean (73.5%), followed by wheat (71.2%), and corn (67.7%). However, the results of the combined species model in Table 4-12 show the lowest (65.3%) among all. Here, the combined dataset has high variance as it consists of three different types of plants. In some cases, the model might get confused determining each category for similar responses from a different category. So, the model gets disorderly sometimes to differentiate categories. Also, it is found that the results of different categories vary with plant species. Therefore, species wise models should be used.

Table 4-13: Accuracy comparisons with other algorithms

	Ensemble Bagged	**SVM**	**Decision Tree**	**KNN**
Corn	53.9±4.4	53.1±9.2	44.7±6	67.7±5.2
Soybean	67.2±6.9	62.1±4.4	62.6±13.7	73.5±9.5
Wheat	66.4±5.1	65.2±5.2	52.8±4.5	71.2±12.1
All species combined	53.9±2.3	42.7±2.2	43.7±2.5	65.3±6.6

Now, the importance of the wavelengths to predict P status is investigated. Based on the optimized feature weights from PSO, as described in the methodology, the importance of the wavelength can be ranked based on the weights. After running the process five times, the box plots of the feature weights are shown in Figure 4-11. Here, only the weights of the model based on the combined dataset (Table 4-12) are analyzed. According to the median values of the weights from five runs, the most important wavelength is found to be 610 nm followed by 730, 570, 860, 500, 450, 760, 810, 650, 680, 600 and 550 nm.

In this section, the comparisons between the proposed system with other existing noninvasive methods related to P measurements are discussed based on cost and weight. Hyperspectral devices and FieldSpec series operate on more optical bands (Hyperspectral 400-500, FieldSpec 200-500) than the proposed device (12 bands). So, in any setting, their measurement accuracy will be higher. However, in terms of cost, these devices are very expensive (hyperspectral camera ~$25,000-$50,000, Field Spec series around ~$10,000-$60000) compared

to our proposed device, which is very cheap approximately $200. In addition, the proposed sensor is also very useful, as it is comparatively light-weight (350 g) than the hyperspectral camera (1.3 kg – 4.5 kg) and FieldSpec series (5.4 kg). Moreover, as the proposed device is cheap and portable, it can be implemented on a large scale for P monitoring, which is impractical for expensive and bulky devices.

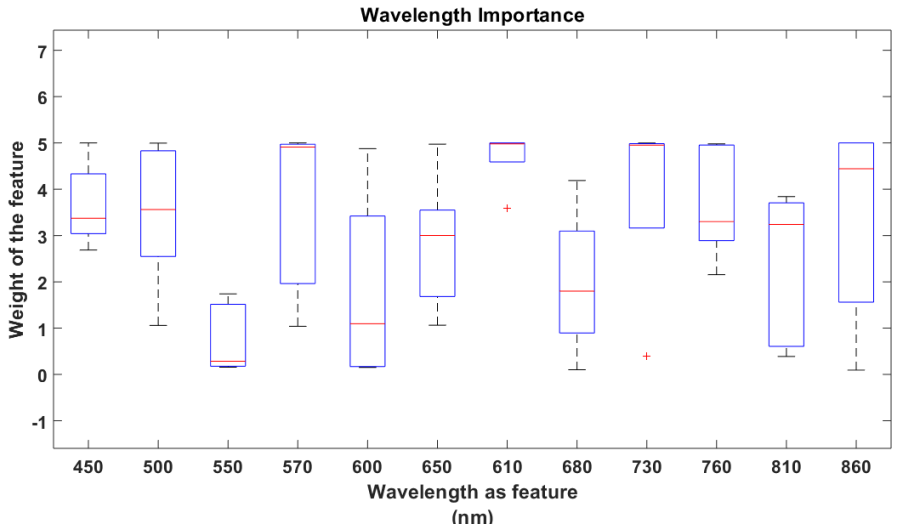


Figure 4-11: Box plots of the weights achieved by the features in 5 runs. Here, the 12 features are 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm.

In this work, we propose a low-cost multi-spectral sensing device, which can determine the P status in leaves. The sensor module detects reflectance at 12 wavelengths from 450nm to 860 nm, and predicts the P level, using the KNN algorithm. To investigate the performance of the proposed method, we performed an experiment on corn, soybean, and wheat in a greenhouse-controlled environment, and found an average accuracy of 71.2%, 73.5%, and 67.7% for corn, soybean, and wheat respectively. Moreover, the model built on the combined dataset shows 65.3% accuracy. So, the species-wise model is suggested. In addition, it is worth to refer that the cost of the proposed device is $200, which is very cheap compared to existing methods.

The multispectral system we propose in this work has been demonstrated as a very effective device to determine leaf P status in terms of cost, weight, and portability. However, the device has not been correlated with actual ground truth P content, and the performance of this sensing system has not tested in the field. These two limitations will be the next phase project. Moreover, only

four levels of P have been classified in this experiment. Applying the technique to other crop species at different growth stages will need more investigation in future research.

Overall, investigating the performance of the proposed device in all the five experiments, it can be said that the proposed techniques are an effective option for monitoring N and P in plant leaves with decent accuracies. Although the existing costly techniques like hyperspectral imaging might show better estimation is this regard, the proposed cost-effective technique also shows decent results. It is basically a cost feature tradeoff.

Chapter 5: Conclusion and Recommendations for the Future Work

5.1. Conclusion

The ever-growing population calls for improvement of the current food production techniques to resolve the global food security bottleneck. To tackle this hindrance, effective implementation of the fast-paced technologies is needed for high throughput plant phenotyping. However, the available methods are expensive, bulky, and sometimes inaccessible that demand significant improvements.

In this thesis, a low-cost multi-spectral sensing system is proposed, which can sense leaf-level nutrients, specifically N and P by capturing the reflectance at 12 wavelengths ranging from 450nm to 860 nm. The major two parts of the system are - a visible sensor that can capture the reflectance at six wavelengths from 450 - 650 nm and the other six wavelengths are covered by a NIR sensor range of 610 - 860 nm. The features of the proposed sensor are summarized below:

- Multifunctional (N and P sensing)
- Overall cost of the device is $200
- Portable to be carried around the field
- Lightweight (350 g)
- Multispectral sensing (12 optical bands centered at 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm in VIS/NIR regions)

To examine the performance of the proposed device, five experiments were conducted. Data collected from leaf samples are modeled through a machine learning pipeline. The overall results of each experiment are summarized below:

- Classifying four N levels in leaves with an average accuracy of 88.4% on the test set (Experiment 1).

- Classifying high N and low N in leaves with an average accuracy of 79.2% in the field (Experiment 2).
- In experiment 3, the results reveal that the best correlation of N is found in soybean (R^2 of 82.29%) followed by corn (80.05%), canola (63.91%), and wheat (63.21%). After comparing the N estimation result with a commercially used device, SPAD; it was found that the proposed multispectral sensor shows a better correlation with N (73.96%) than SPAD (R^2 of 60.21%).
- For estimating leaf water content, corn shows the best correlation of 68.41%, followed by wheat of 64.58%, soybean of 46.38%, and canola shows 18.02% (Experiment 4).
- Finally, in the fifth experiment of classifying four P responses, an average accuracy of about 71.2%, 73.5%, and 67.7% for corn, soybean, and wheat respectively was observed.

Overall, the result concludes that the proposed system can estimate N and P with decent accuracy which is comparable to existing devices. But less correlation is observed in water estimation. So, the utility of the proposed device in water content estimation needs further consideration. Moreover, the device has not been correlated with actual ground truth P content. In addition, most of the experiments were conducted on the green-house controlled environment rather in the actual field setting. Furthermore, in this thesis, only four different species of plants: canola, corn, soybean, and wheat, are used as test plants. It was beyond the scope of this thesis to investigate how the results vary on other species of plants.

5.2. Recommendations for the Future Work

Although the proposed system has many benefits, there are considerable scopes and opportunities to explore new possibilities and improve the system. Some of the recommendations for future works are:

- The proposed device is composed of six-channel visible and six-channel NIR detectors. So, more channels can be incorporated into this device for performance improvement. In this respect, 18-channel sensor AS7265 (AMS) [99] can be considered.

- Water experiments can be explored more considerably. A more robust model might be built by creating a large variety of water content in leaf samples.
- The proposed system can be further explored to estimate actual P content by correlating the spectral information with chemically measured P.
- The accuracy of the device can be further improved by experimenting on more samples and making a robust model for N estimation.
- A similar technique can be applied to other crop species.
- The performance variations because of different field parameters like wind, dust, shadow, and different growth stage of plants can be explored.
- Besides, the future scope includes using this device to correlate with other nutrients, such as P and K.

The exploration of the above-mentioned future scopes may create new opportunities for determining leaf-level chemical traits and making plant phenotyping a high throughput.

References

[1] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, “Global food demand and the sustainable intensification of agriculture,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 50, pp. 20260–20264, Dec. 2011.

[2] H. C. J. Godfray *et al.*, “Food security: The challenge of feeding 9 billion people,” *Science*. 2010.

[3] A. Walter, F. Liebisch, and A. Hund, “Plant phenotyping: from bean weighing to image analysis,” *Plant Methods*, vol. 11, no. 1, p. 14, 2015.

[4] L. Li, Q. Zhang, and D. Huang, “A review of imaging techniques for plant phenotyping.,” *Sensors (Basel).*, vol. 14, no. 11, pp. 20078–111, Oct. 2014.

[5] F. Fiorani and U. Schurr, “Future Scenarios for Plant Phenotyping,” *Annu. Rev. Plant Biol.*, vol. 64, no. 1, pp. 267–291, Apr. 2013.

[6] “Plant Phenotyping and Imaging Research Centre - U of S Plant Phenotyping and Imaging Research Centre - University of Saskatchewan.” [Online]. Available: https://p2irc.usask.ca/. [Accessed: 19-Feb-2020].

[7] “IPPN Home.” [Online]. Available: https://www.plant-phenotyping.org/. [Accessed: 19-Feb-2020].

[8] “Plant phenomics. Plant phenotyping tools and research.” [Online]. Available: https://www.plantphenomics.org.au/. [Accessed: 19-Feb-2020].

[9] S. Dhondt, N. Wuyts, and D. Inzé, “Cell to whole-plant phenotyping: the best is yet to come,” *Trends Plant Sci.*, vol. 18, no. 8, pp. 428–439, Aug. 2013.

[10] P. Pandey, Y. Ge, V. Stoerger, and J. C. Schnable, “High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging,” *Front. Plant Sci.*, vol. 8, no. August, p. 1348, Aug. 2017.

[11] N. Fahlgren, M. A. Gehan, and I. Baxter, “Lights, camera, action: high-throughput plant phenotyping is ready for a close-up,” *Curr. Opin. Plant Biol.*, vol. 24, pp. 93–99, Apr. 2015.

[12] K. Q. Yu, Y. R. Zhao, X. L. Li, Y. N. Shao, F. Liu, and Y. He, “Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant,” *PLoS One*, vol. 9, no. 12, pp. 1–19, 2014.

[13] X. Zhang, F. Liu, Y. He, and X. Gong, “Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging,” *Biosyst. Eng.*, vol. 115, no. 1, pp. 56–65, May 2013.

[14] Y. Ge, G. Bai, V. Stoerger, and J. C. Schnable, “Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging,” *Comput. Electron. Agric.*, vol. 127, pp. 625–632, Sep. 2016.

[15] J. Praveen Kumar and S. Domnic, “Image based leaf segmentation and counting in rosette plants,” *Inf. Process. Agric.*, vol. 6, no. 2, pp. 233–246, Jun. 2019.

[16] Y. Zhai, L. Cui, X. Zhou, Y. Gao, T. Fei, and W. Gao, “Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: Comparison of partial least-square regression and support vector machine regression met,” *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2502–2518, 2013.

[17] S. L. Osborne, J. S. Schepers, D. D. Francis, and M. R. Schlemmer, “Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements,” *Agron. J.*, vol. 94, no. 6, p. 1215, 2002.

[18] W. A. Wahono, D. Indradewa, B. H. Sunarminto, E. Haryono, and D. Prajitno, “Evaluation of the use of SPAD-502 chlorophyll meter for non-destructive estimation of nitrogen status of tea leaf (Camellia sinensis L. Kuntze),” *Indian J. Agric. Res.*, vol. 53, no. of, pp. 333–337, May 2019.

[19] J. Wang *et al.*, “Non-destructive evaluation of the leaf nitrogen concentration by in-field visible/near-infrared spectroscopy in Pear Orchards,” *Sensors*, vol. 17, no. 3, p. 538, Mar. 2017.

[20] R. Basyouni, B. Dunn, and C. Goad, “Use of nondestructive sensors to assess nitrogen status in potted Dianthus (Dianthus chinensis L.) production,” *Can. J. Plant Sci.*, p. CJPS-2016-

0059, Jul. 2016.

[21] J. Wu, D. Wang, C. J. Rosen, and M. E. Bauer, "Comparison of petiole nitrate concentrations, SPAD chlorophyll readings, and QuickBird satellite imagery in detecting nitrogen status of potato canopies," *F. Crop. Res.*, vol. 101, no. 1, pp. 96–103, Feb. 2007.

[22] D. Xiong *et al.*, "SPAD-based leaf nitrogen estimation is impacted by environmental factors and crop leaf characteristics," *Sci. Rep.*, vol. 5, no. July, pp. 1–12, 2015.

[23] R. F. Muñoz-Huerta, R. G. Guevara-Gonzalez, L. M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez, and R. V. Ocampo-Velazquez, "A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances," *Sensors (Switzerland)*, vol. 13, no. 8, pp. 10823–10843, 2013.

[24] H. Liu, H. Zhu, Z. Li, and G. Yang, "Quantitative analysis and hyperspectral remote sensing of the nitrogen nutrition index in winter wheat," *Int. J. Remote Sens.*, vol. 00, no. 00, pp. 1–24, Aug. 2019.

[25] C. Ni, D. Wang, and Y. Tao, "Variable weighted convolutional neural network for the nitrogen content quantization of Masson pine seedling leaves with near-infrared spectroscopy," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 209, pp. 32–39, Feb. 2019.

[26] T. Blackmer and J. S. Schepers, "Techniques for monitoring crop nitrogen status in corn," *Commun. Soil Sci. Plant Anal.*, vol. 25, no. 9-10 : Soil testing and plant Analysis:Precision nutrient management part II, pp. 1791–1800, 1994.

[27] K. Kusnierek and A. Korsaeth, "Simultaneous identification of spring wheat nitrogen and water status using visible and near infrared spectra and powered partial least squares regression," *Comput. Electron. Agric.*, vol. 117, pp. 200–213, Sep. 2015.

[28] F. Li *et al.*, "Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices," *F. Crop. Res.*, vol. 157, pp. 111–123, Feb. 2014.

[29] L. Li *et al.*, "Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy," *Ind. Crops Prod.*, vol. 91, pp. 194–204,

Nov. 2016.

[30] K. G. Raghothama and A. S. Karthikeyan, “Phosphate Acquisition,” *Plant Soil*, vol. 274, no. 1–2, pp. 37–49, Jul. 2005.

[31] S. D. Guthe and R. R. Deshmukh, “Estimation of phosphorus content in leaves of plants using PLSRand SVMR : A review,” *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 132–135, 2017.

[32] L. Yanli *et al.*, “Prediction of nitrogen and phosphorus contents incitrus leaves based on hyperspectral imaging,” *Int. J. Agric. Biol. Eng.*, vol. 8, no. 2, pp. 80–88, 2015.

[33] J. V. Sinfield, D. Fagerman, and O. Colic, “Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils,” *Comput. Electron. Agric.*, vol. 70, no. 1, pp. 1–18, 2010.

[34] G. Sun, Y. Ding, X. Wang, W. Lu, Y. Sun, and H. Yu, “Nondestructive determination of nitrogen, phosphorus and potassium contents in greenhouse Tomato plants based on multispectral three-dimensional imaging,” *Sensors*, vol. 19, no. 23, p. 5295, Dec. 2019.

[35] P. Mishra, M. S. M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, and P. Scheunders, “Close range hyperspectral imaging of plants: A review,” *Biosyst. Eng.*, vol. 164, pp. 49–67, Dec. 2017.

[36] J. G. A. Barbedo, “Detection of nutrition deficiencies in plants using proximal images and machine learning: A review,” *Comput. Electron. Agric.*, vol. 162, no. December 2018, pp. 482–492, Jul. 2019.

[37] G. ElMasry, N. Mandour, S. Al-Rejaie, E. Belin, and D. Rousseau, “Recent applications of multispectral imaging in seed phenotyping and quality monitoring—an overview,” *Sensors*, vol. 19, no. 5, p. 1090, Mar. 2019.

[38] G. ElMasry, D.-W. Sun, and P. Allen, “Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef,” *J. Food Eng.*, vol. 110, no. 1, pp. 127–140, May 2012.

[39] H. Huang, M. Chen, and Y. Duan, “Dimensionality Reduction of Hyperspectral Image

Using Spatial-Spectral Regularized Sparse Hypergraph Embedding," *Remote Sens.*, vol. 11, no. 9, p. 1039, May 2019.

[40] Y. Ge *et al.*, "High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel," *Plant Methods*, vol. 15, no. 1, p. 66, Dec. 2019.

[41] S. K. Arndt, A. Irawan, and G. J. Sanders, "Apoplastic water fraction and rehydration techniques introduce significant errors in measurements of relative water content and osmotic potential in plant leaves," *Physiol. Plant.*, vol. 155, no. 4, pp. 355–368, Dec. 2015.

[42] E. R. Hunt and B. N. Rock, "Detection of changes in leaf water content using near- and middle-infrared reflectances," *Remote Sens. Environ.*, 1989.

[43] X. Jin, C. Shi, C. Y. Yu, T. Yamada, and E. J. Sacks, "Determination of leaf water content by visible and near-infrared spectrometry and multivariate calibration in Miscanthus," *Front. Plant Sci.*, vol. 8, no. May, pp. 1–8, May 2017.

[44] R. Gente *et al.*, "Determination of leaf water content from terahertz time-domain spectroscopic data," *J. Infrared, Millimeter, Terahertz Waves*, vol. 34, no. 3–4, pp. 316–323, Apr. 2013.

[45] Q. Zhang, Q. Li, and G. Zhang, "Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection," *Spectrosc. An Int. J.*, vol. 27, no. 2, pp. 93–105, 2012.

[46] "AS7262-BLGT ams | Sensors, Transducers | DigiKey." [Online]. Available: https://www.digikey.com/product-detail/en/ams/AS7262-BLGT/AS7262-BLGTCT-ND/6219609#images. [Accessed: 27-Aug-2019].

[47] "Single-Chip Spectrometers | DigiKey." [Online]. Available: https://www.digikey.be/en/articles/techzone/2017/jun/optical-sensor-on-chip-ics-simplify-handheld-spectrometer-design. [Accessed: 27-Aug-2019].

[48] "SparkFun Spectral Sensor Breakout - AS7262 Visible (Qwiic) - SEN-14347 - SparkFun Electronics." [Online]. Available: https://www.sparkfun.com/products/14347. [Accessed:

25-Feb-2020].

[49] “AS7263 6-Channel Spectrometer - ams | Mouser Canada.” [Online]. Available: https://www.mouser.ca/new/ams/ams-as7263-spectral-id-device/. [Accessed: 27-Aug-2019].

[50] “SparkFun Spectral Sensor Breakout - AS7263 NIR (Qwiic) - SEN-14351 - SparkFun Electronics.” [Online]. Available: https://www.sparkfun.com/products/14351. [Accessed: 25-Feb-2020].

[51] Senthilkumar G, Gopalakrishnan K, and Satish Kumar, “Embedded image capturing system using raspberry pi system,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 2, pp. 213–215, 2014.

[52] A. Imteaj, T. Rahman, M. K. Hossain, M. S. Alam, and S. A. Rahat, “An IoT based fire alarming and authentication system for workhouse using Raspberry Pi 3,” in *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2017, no. February 2010, pp. 899–904.

[53] “Amazon.com: Raspberry Pi 3 Model B Board: Computers & Accessories.” [Online]. Available: https://www.amazon.com/Raspberry-Pi-MS-004-00000024-Model-Board/dp/B01LPLPBS8. [Accessed: 25-Jan-2020].

[54] “SparkFun Qwiic Mux Breakout Board - 8 Channel (TCA9548A) - RobotShop.” [Online]. Available: https://www.robotshop.com/ca/en/sparkfun-qwiic-mux-breakout-board-8-channel-tca9548a.html?gclid=CjwKCAiAjrXxBRAPEiwAiM3DQqz8kDvYReuEt8ffmXgu35HIizTiqAb-wGAoVtqG_cTCvzyTe6uo8xoCD7wQAvD_BwE. [Accessed: 26-Jan-2020].

[55] “blackweb Slim Powerbank | Walmart Canada.” [Online]. Available: https://www.walmart.ca/en/ip/blackweb-slim-powerbank/6000197020355. [Accessed: 26-Jan-2020].

[56] “Yosoo DIY Display Module, 1pc 1.3 inch OLED Display Module IIC I2C Communicate 128X64 White Text Color | Walmart Canada.” [Online]. Available: https://www.walmart.ca/en/ip/Yosoo-DIY-Display-Module-1pc-1-3-inch-OLED-Display-

Module-IIC-I2C-Communicate-128X64-White-Text-Color/PRD4YQ4LAM30XYR. [Accessed: 26-Jan-2020].

[57] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[58] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm," *2006 IEEE Southwest Symp. Image Anal. Interpret.*, pp. 61–65, 2006.

[59] R. R. Kumar, P. Viswanath, and C. S. Bindu, "Nearest neighbor classifiers: Reducing the computational demands," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 45–50.

[60] K. Mochida *et al.*, "Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective," *Gigascience*, vol. 8, no. 1, Jan. 2019.

[61] Balakrishna K. and M. Rao, "Tomato plant leaves disease classification using KNN and PNN," *Int. J. Comput. Vis. Image Process.*, vol. 9, no. 1, pp. 51–63, Jan. 2019.

[62] S. R. M. Amin and R. Awang, "Automated detection of nitrogen status on plants: Performance of image processing techniques," in *2018 IEEE Student Conference on Research and Development (SCOReD)*, 2018, pp. 1–4.

[63] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, "Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance," *Comput. Electron. Agric.*, vol. 74, no. 1, pp. 91–99, Oct. 2010.

[64] M. Habibullah, M. A. M. Oninda, A. N. Bahar, A. Dinh, and K. A. Wahid, "NIR-spectroscopic classification of blood glucose level using machine learning approach," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–4.

[65] T. Seehapoch and S. Wongthanavasu, "Speech emotion recognition using Support Vector Machines," in *2013 5th International Conference on Knowledge and Smart Technology (KST)*, 2013, pp. 86–91.

[66] K.-P. Wu and S.-D. Wang, “Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space,” *Pattern Recognit.*, vol. 42, no. 5, pp. 710–717, May 2009.

[67] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[68] S. L. Salzberg, “C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993,” *Mach. Learn.*, 1994.

[69] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, “Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid,” *IEEE Trans. Ind. Informatics*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.

[70] J. Chopda, H. Raveshiya, S. Nakum, and V. Nakrani, “Cotton Crop Disease Detection using Decision Tree Classifier,” in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, 2018, pp. 1–5.

[71] L. Prado Osco *et al.*, “Predicting Canopy Nitrogen Content in Citrus-Trees Using Random Forest Algorithm Associated to Spectral Vegetation Indices from UAV-Imagery,” *Remote Sens.*, vol. 11, no. 24, p. 2925, Dec. 2019.

[72] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, “A comparison of decision tree ensemble creation techniques,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2007.

[73] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[74] K. Machová, F. Barčák, and P. Bednár, “A bagging method using decision trees in the role of base classifiers,” *Acta Polytech. Hungarica*, vol. 3, no. 2, pp. 121–132, 2006.

[75] D. Lavanya, “Ensemble Decision Tree Classifier For Breast Cancer Data,” *Int. J. Inf. Technol. Converg. Serv.*, vol. 2, no. 1, pp. 17–24, Feb. 2012.

[76] U. Lee, S. Chang, G. A. Putra, H. Kim, and D. H. Kim, “An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis,” *PLoS One*, vol. 13, no. 4, p. e0196615, Apr. 2018.

[77] J. Sun *et al.*, “Estimating Rice Leaf Nitrogen Concentration: Influence of Regression Algorithms Based on Passive and Active Leaf Reflectance,” *Remote Sens.*, vol. 9, no. 9, p. 951, Sep. 2017.

[78] M. Seeger, “Gaussian processes for machine learning.,” *Int. J. Neural Syst.*, vol. 14, no. 2, pp. 69–106, Apr. 2004.

[79] S. Lunderman, G. M. Fioroni, R. L. McCormick, M. R. Nimlos, M. J. Rahimi, and R. W. Grout, “Screening fuels for autoignition with small-volume experiments and gaussian process classification,” *Energy & Fuels*, vol. 32, no. 9, pp. 9581–9591, Sep. 2018.

[80] M. Lin *et al.*, “Robust gaussian process regression for real-time high precision GPS signal enhancement,” *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD ’19*, pp. 2838–2847, Jun. 2019.

[81] T. Chen and B. Wang, “Bayesian variable selection for Gaussian process regression: Application to chemometric calibration of spectrometers,” *Neurocomputing*, vol. 73, no. 13–15, pp. 2718–2726, Aug. 2010.

[82] F. S. Firouzabadi, A. Vard, M. Sehhati, and M. Mohebian, “An optimized framework for cancer prediction using immunosignature,” *J. Med. Signals Sens.*, vol. 8, no. 3, pp. 161–169, 2018.

[83] R. P. Borda and J. D. Frost, “Error reduction in small sample averaging through the use of the median rather than the mean,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 25, no. 4, pp. 391–392, Oct. 1968.

[84] M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mañanas, and F. Mokarian, “A Hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning,” *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 75–85, 2017.

[85] C. A. Boneau, “The effects of violations of assumptions underlying the t test.,” *Psychol. Bull.*, vol. 57, no. 1, pp. 49–64, 1960.

[86] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in *MHS’95.*

Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995, pp. 39–43.

[87] D. Y. Sha and H.-H. Lin, "A multi-objective PSO for job-shop scheduling problems," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1065–1070, Mar. 2010.

[88] S. Lakshmi and R. Sivakumar, "Plant phenotyping through Image analysis using nature inspired optimization techniques," in *Intelligent Systems Reference Library*, 2019, pp. 165–187.

[89] A. Modiri and K. Kiasaleh, "Permittivity estimation for breast cancer detection using particle swarm optimization algorithm," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 1359–1362.

[90] K. A. Ross *et al.*, "Cross-Validation," in *Encyclopedia of Database Systems*, Boston, MA: Springer US, 2009, pp. 532–538.

[91] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.

[92] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology," *Interdiscip. J. Information, Knowledge, Manag.*, vol. 14, pp. 045–076, Sep. 2018.

[93] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting," *Appl. Sci.*, vol. 8, no. 1, p. 28, Dec. 2017.

[94] R. Palaniappan, K. Sundaraj, and S. Sundaraj, "A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals," *BMC Bioinformatics*, vol. 15, no. 1, p. 223, Dec. 2014.

[95] D. O. McInerney and M. Nieuwenhuis, "A comparative analysis of k NN and decision tree methods for the Irish National Forest Inventory," *Int. J. Remote Sens.*, vol. 30, no. 19, pp. 4937–4955, Sep. 2009.

[96] N. Vigneau, M. Ecarnot, G. Rabatel, and P. Roumet, "Potential of field hyperspectral

imaging as a non destructive method to assess leaf nitrogen content in Wheat," *F. Crop. Res.*, vol. 122, no. 1, pp. 25–31, 2011.

[97] J. W. Li *et al.*, "Use of fluorescence-based sensors to determine the nitrogen status of paddy rice," *J. Agric. Sci.*, vol. 151, no. 6, pp. 862–871, Dec. 2013.

[98] R. O. Mendoza-Tafolla, P. Juarez-Lopez, R. E. Ontiveros-Capurata, M. Sandoval-Villa, I. Alia-Tejacal, and G. Alejo-Santiago, "Estimating nitrogen and chlorophyll status of romaine lettuce using SPAD and at LEAF readings," *Not. Bot. Horti Agrobot. Cluj-Napoca*, vol. 47, no. 3, pp. 751–756, 2019.

[99] "SparkFun Triad Spectroscopy Sensor - AS7265x (Qwiic) - SEN-15050 - SparkFun Electronics." [Online]. Available: https://www.sparkfun.com/products/15050. [Accessed: 29-Apr-2020].