

Article

Low-Cost Multispectral Sensor Array for Determining Leaf Nitrogen Status

Mohammad Habibullah ^{1,*}, Mohammad Reza Mohebbian ¹, Raju Soolanayakanahally ²,
Ali Newaz Bahar ¹, Sally Vail ², Khan A. Wahid ¹ and Anh Dinh ¹

¹ Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada; mr.mohebbian@usask.ca (M.R.M.); ali.bahar@usask.ca (A.N.B.); khan.wahid@usask.ca (K.A.W.); anv252@mail.usask.ca (A.D.)

² Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK S7N 0X2, Canada; raju.soolanayakanahally@canada.ca (R.S.); sally.vail@canada.ca (S.V.)

* Correspondence: moh220@mail.usask.ca; Tel.: +1-306250-3165

Received: 11 June 2020; Accepted: 18 August 2020; Published: 25 August 2020



Abstract: A crop's health can be determined by its leaf nutrient status; more precisely, leaf nitrogen (N) level, is a critical indicator that carries a lot of worthwhile nutrient information for classifying the plant's health. However, the existing non-invasive techniques are expensive and bulky. The aim of this study is to develop a low-cost, quick-read multi-spectral sensor array to predict N level in leaves non-invasively. The proposed sensor module has been developed using two reflectance-based multi-spectral sensors (visible and near-infrared (NIR)). In addition, the proposed device can capture the reflectance data at 12 different wavelengths (six for each sensor). We conducted the experiment on canola leaves in a controlled greenhouse environment as well as in the field. In the greenhouse experiment, spectral data were collected from 87 leaves of 24 canola plants, subjected to varying levels of N fertilization. Later, 42 canola cultivars were subjected to low and high nitrogen levels 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 greenhouse experiment and 79.2% for the field experiment. Overall, the result concludes that the proposed cost-effective sensing system can be viable in determining leaf nitrogen status.

Keywords: plant phenotyping; non-invasive; spectrometer; reflectance; machine learning

1. Introduction

Nitrogen (N) is one of the most indispensable elements for plants owing to its vital role in major physiological processes [1]. Various enzymatic proteins that regulate plant growth processes are formed basically from N [2]. During vegetative growth, plants take up soil N as mineral and transport it to the leaves. Due to the limitation of N supply in soils, the optimal use of N fertilizer at the different growth stages of plants is a challenging issue in the existing agricultural system [3]. Besides, to achieve high yield, sometimes farmers over-fertilize N, which can be leached below the root zone or might get lost in run-off. These occurrences may cause the rise of nitrate ion in water, which creates human health issues [4]. Furthermore, excess N fertilization results in adverse environmental effects such as soil denitrification, greenhouse effect, and water pollution [5]. On the other hand, photosynthesis and crop yield are negatively affected by N deficiency; therefore, N content monitoring during the vegetative growth stage is needed for optimal fertilizer application [6].

Optimal N management can minimize N losses to the environment and also ensure high production. The conventional approach for N management is based on soil analysis, which is time-consuming and not rapid enough to make fertilizer recommendations on time [6].

Several crop-based N status measurement techniques have been proposed over time. Most of the current techniques can be categorized as chemical analysis, optical spectroscopy, image-based analysis, and electrical impedance-based analysis. The widely used chemical-based methods are Kjeldahl digestion and Dumas combustion; these are destructive and labor-intensive processes [7]. Optical spectroscopy is widely used for non-destructive measurements of N status in leaves by utilizing Soil Plant Analysis Development (SPAD) (Spectrum Technologies, Inc., 3600 Thayer Court, Aurora, IL, USA), the spectrometer FieldSpec 3 (Analytical Spectral Devices, Boulder, CO, USA), GreenSeeker (Trimble, Sunnyvale, CA, USA), or QuickBird (Ball Aerospace and Technologies, Boulder, CO, USA). These are basically called proximal optical sensors as they are a form of remote sensing where these are placed either in contact or close to the crop [4]. However, these spectral-based measurements have some drawbacks. For instance, the SPAD meter's functionality is centered on only one wavelength at 650 nm for chlorophyll determination and can become saturated with high N application. According to Xiong et al. (2015) [5], the part of N allocated to chlorophyll is very small, and mostly constitutes photosynthetic proteins. Moreover, the SPAD meter is also affected by the crop species, and environmental factors such as shading, and light-dependent chloroplast movements. Although a correlation coefficient $R^2 = 0.86$ is reported by using the FieldSpec 3 device [8], this device is expensive and not feasible to carry around in the field. Additionally, GreenSeeker is also a costly device, and the estimate can become saturated with increasing areas of biomass/leaf.

Technical development in precision agriculture for facilitating non-invasive measurements such as image-based measurement is becoming popular among farmers. The ways the images are captured can be categorized into three: satellites, airplanes, and unmanned aerial vehicles (UAVs) [9]. The imagery from the QuickBird satellite is expensive and affected by atmospheric interference [10]. Further, using aerial devices needs specific training, and the weather conditions can also harm image quality. Red green blue (RGB) images are based on three optical bands that correspond to three visual cones of the eye. Mercado et al. (2008) [11] used a digital RGB camera and could show a correlation between N in tomato with color images. However, in recent years, the use of the hyperspectral camera is becoming more popular than RGB cameras for plant phenotyping due to the high throughput. This device can provide nitrogen distribution in spatial coordinates [12]. However, hyperspectral imaging (HSI) is generally very expensive, limiting its use in the field, and the usage is limited to a very specific application. A lot of spaceborne image spectrometer data has become accessible in recent years, and this potential supply of data would decrease the cost of research in the future [13–15]. Researchers are also trying to correlate the electrical properties of plants (electrical impedance) with macronutrients [16]. These electrical techniques are still in the initial phase of development. Concisely, these methods suffer from electrode polarization effects. Figure 1 summarizes the overall techniques related to N estimation in crops.

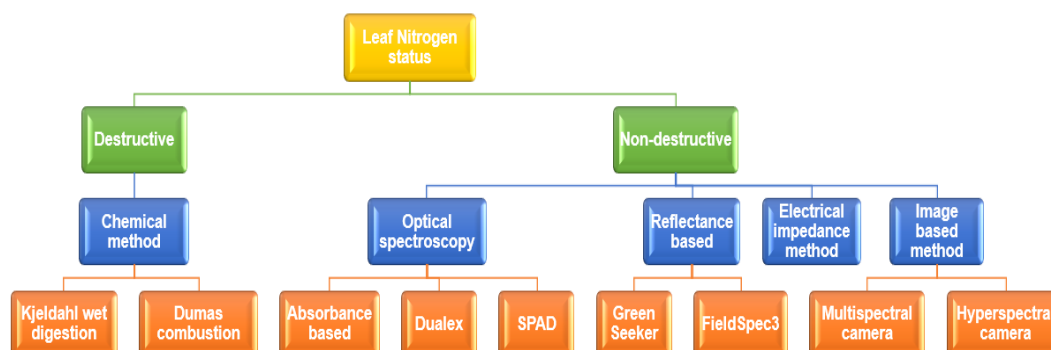


Figure 1. Destructive and non-destructive methods used for estimating leaf nitrogen status.

Various chemical forms of N are sensitive to visible and near-infrared (NIR) regions. The reflectance around 550 nm wavelength is significant for differentiating various N treatments [17]. Besides, isotopes of N have also been measured using spectral analysis [18]. Moreover, the total N in

plants can be determined at 671 nm and 780 nm wavelength [19]. More precisely, for hyperspectral imaging 440, 473, 513, 542, 659, 718, 744, 865, 928, 965, 986, and 1015 nm wavelengths are effective for determining N level in leaves, which could obtain the correlation $R^2 = 0.882$ with N content [20].

Machine learning (ML) allows effective decision-making and rational behavior without (or with limited) human interference in real-world scenarios. ML creates a comprehensive and scalable platform for not only data-driven decision making but also for the integration of specialist expertise into the method. According to the literature [21–23], machine learning approaches can help to create a precise model for estimating the leaf N status. Probabilistic machine learning models, such as Gaussian and Buffet processes [24,25], are able to reduce sensor noises when the fusion of information is used. On the other hand, since multi-spectral data is high-dimensional [26], non-linear methods such as support vector machine, k-nearest neighbor, neural networks, and decision tree have been regarded as acceptable methods with the ability to solve the issue of overfitting [6].

In the present study, a multi-spectral sensor is proposed that can be used to detect the leaf N levels. The proposed method is applied and tested on the canola leaves. The primary objectives of this study are (a) develop a low-cost sensor, (b) validate the portability for in-field measurement, and (c) build an accurate machine learning model for level identification. To achieve our research objectives, we tested our model on a single canola variety in greenhouse, and on several canola varieties grown in field with variable nitrogen levels.

2. Materials and Methods

2.1. Designed Hardware for N Sensing System

The designed hardware prototype was composed of two sensors named as Sensor1 and Sensor2. Sensor1 (SparkFun number SEN-14347, SparkFun Electronics, Niwot, CO, USA) was a visible, multi-spectral sensor, and Sensor2 (SparkFun number SEN-14351, SparkFun Electronics, Niwot, CO, USA) was a NIR multi-spectral sensor. Other essential components were a Qwiic mux breakout board (SparkFun number BOB-14685 SparkFun Electronics, Niwot, CO, USA) and a Raspberry Pi version 3 model B (Raspberry Pi Foundation; Cambridge, UK).

Sensor1 (Figure 2a,b) was a six-channel multi-spectral sensor in the visible spectrum. With a full-width half-max (FWHM) of 40 nm, the spectral bands ranged from around 430 to 670 nm. It had a built-in aperture that provided a controlled light entrance. The I2C register set was used to store the spectral data. The six visible channels were 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 as shown in Figure 2c. Moreover, it had a 16-bit ADC (analog to digital converter), and it operated at low voltage of around 2.7 to 3.6 V with the I2C interface. In addition, the sensor's field of view was ± 20 degree. Here, each channel was tested with GAIN = $16\times$ at ambient temperature (25°C), and the calibration and measurements were made using diffused light. The measurement unit of the channel was $\mu\text{W}/\text{cm}^2$ with an accuracy of 12%. The sensor-on-chip had a built-in source light (5700 K white LED, L130-5780HE1400001, Lumileds, CA, USA) having a color rendering index (CRI) of 80. The energy at each channel was calculated with a ± 40 nm bandwidth around the center wavelengths.

Sensor2 was a digital six-channel spectrometer in the NIR light region (Figure 3a,b). The six independent optical filters had a spectral response defined in the NIR wavelengths, approximately from 600–870 nm and 20 nm of FWHM. Additionally, it had capabilities similar to Sensor1 of integrating Gaussian filters into CMOS (complementary metal-oxide-semiconductor) silicon via nano-optic deposited interference filter technology. As mentioned, it provided a six-channel spectrometry solution at wavelengths 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 (Figure 3c). In addition, the onboard source light was a 2700 K warm LED (L130-2790001400001; Lumileds; CA, USA) with a CRI of 90. The energy at each channel was calculated with a ± 33 nm bandwidth around the center wavelengths. The other configurations that are like Sensor1 are not mentioned.

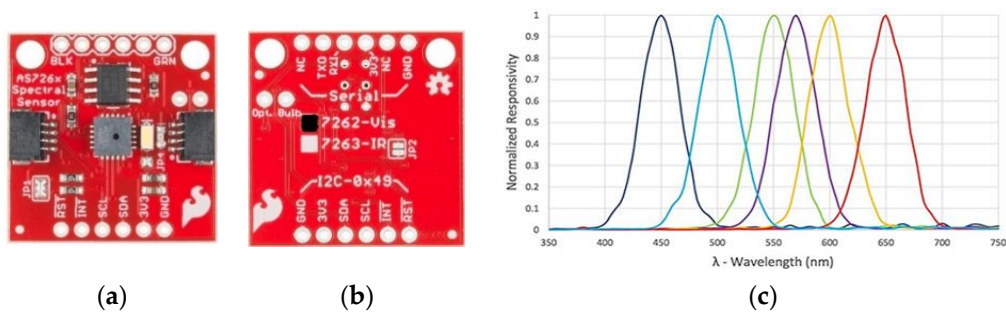


Figure 2. (a) Front view and (b) back view of Sensor1 [27]. (c) Normalized spectral responsivity versus wavelength of Sensor1, which detected at six visible channels—450 nm (channel V), 500 nm (channel B), 550 nm (channel G), 570 nm (channel Y), 600 nm (channel O), and 650 nm (channel R), each with 40 nm FWHM [28].

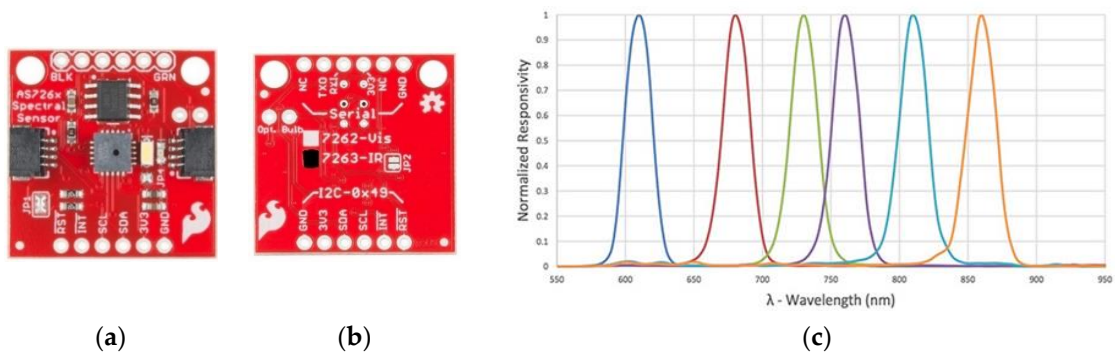


Figure 3. (a) Front view and (b) back view of AS7263 spectral breakout board [29]. (c) Normalized spectral responsivity versus wavelength of Sensor2, which detected at six channels—610 nm (channel R), 680 nm (channel S), 730 nm (channel T), 760 nm (channel U), 810 nm (channel V), and 860 nm (channel W), each with 20 nm FWHM [28].

Raspberry Pi 3 was used for controlling the sensors. The main processing unit consisted of a 1.2 GHz ARM processor and a 1 GB LPDDR2 main memory [30]. This device is used in various applications such as image processing [31], IOT systems [32], and automatic control, for example in smart irrigation [33].

The system consisted of two sensors that had the same I2C address. For simplicity, a multiplexer (mux) was used to communicate with Raspberry Pi on the same bus. It had eight configurable addresses of its own that offered 64 I2C buses. In the proposed setup, channel 1 and channel 2 were used. After collecting the data, the modeling and processing were performed on a computer using MATLAB 2018b. The summary of the hardware components mentioned above is given in Table 1. Moreover, the overall setup, along with the physical 3D view of the system, is shown in Figure 4a–c.

Table 1. Summary of the technical details of the hardware components.

Hardware Components	Summary of the Technical Specifications
Sensor1 (SEN-14347)	6 channel multi-spectral sensor in the visible range (450, 500, 550, 570, 600, 650 nm). The measurement unit of the channel is $\mu\text{W}/\text{cm}^2$. Built-in source light (5700 K white LED)
Sensor2 (SEN-14351)	6 channel multi-spectral sensor in the NIR range (610, 680, 730, 760, 810, and 860 nm). The measurement unit of the channel is $\mu\text{W}/\text{cm}^2$. Onboard source light is a 2700 K warm LED
Microprocessor-based control circuitry	Raspberry Pi 3 model B. 1.2 GHz AR processor. 1 GB LPDDR2 main memory
Multiplexer	8 configurable addresses

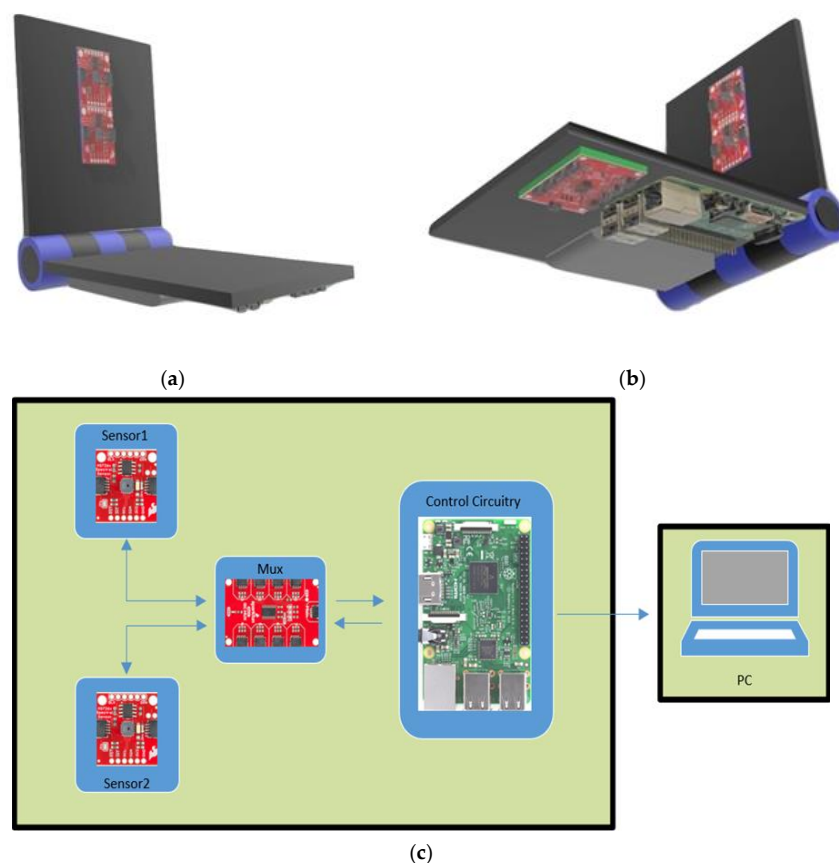


Figure 4. Side view (a) and bottom view (b) of the physical design, and the connection diagram of the devices (c). Here, Sensor1 and Sensor2 are connected to the multiplexer's channel 1 and channel 2, and the mux is connected to the control circuit (Raspberry Pi 3 Model B). All read and write communications are through I2C protocol.

2.2. Greenhouse Experimental Setup

Canola seeds were sown on 2 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% nitrogen, 30% phosphorus, and 15% potassium) fertilizer at a rate of 4 g/L to ensure uniform establishment. Later, the twenty-four pots were divided into four N concentrations, each containing six replicate plants (4 N levels \times 6 pots = 24 plants, Figure 5). Henceforward, only nitrogen fertilizer 30-0-0 (30% nitrogen, 0% phosphorus, and 0% potassium) was applied three times a week at four concentration levels: 0, 6, 12, and 20 g/L. Two panels of 45 w grow lights were used, which were illuminated 16 h a day. The day temperature was kept at 25 °C and night at 20 °C with a relative humidity of ~45%.



Figure 5. One random canola plant from each of the four nitrogen fertilization concentrations. The photograph was taken during week seven.

2.3. Field Experiment Setup

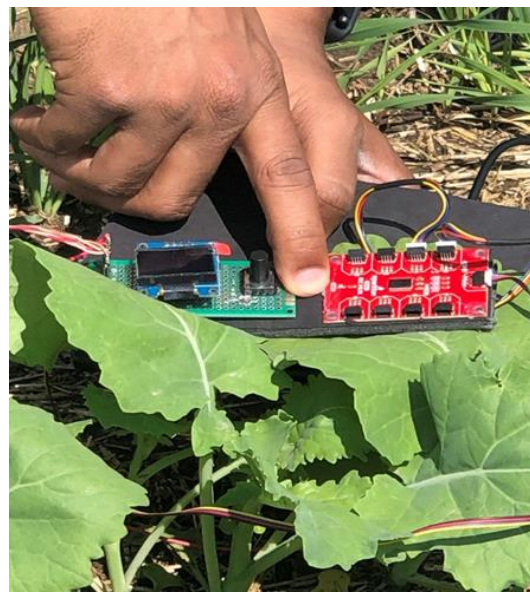
This experiment was conducted at Lewellyn Farm, Saskatoon, Saskatchewan, Canada, during the summer of 2019 (Figure 6a). In this experiment, a total of 42 canola varieties were subjected to low-N and high-N levels. Here, a total of 336 plots (42 varieties \times 4 replicates per N levels) were subjected to two levels of N treatments. In this work, three leaves (one of each plant) from three central and south side plants were used from each plot. Finally, the total number of leaf samples was 1008 (3×336 ; Figure 6b,c).



(a)



(b)



(c)

Figure 6. Field view of different canola plots grown under two nitrogen levels at Lewellyn Farm, Saskatoon, Saskatchewan, Canada during the summer of 2019. Here, a total of 42 canola varieties were subjected to low-N and high-N levels, totaling 336 plots. (a). The proposed sensor, (b), is used for taking measurements in the field, (c).

2.4. Data Collection and Modeling

The proposed prototype measures the reflectance at different wavelengths. The device was calibrated by taking reflectance data from a white surface each time before the data were collected.

A regular white mirror paper [34] was used as a reference for calibration for all measurements. The reflectance data at 12 different wavelengths (450, 500, 550, 570, 600 nm, and 650, 610, 680, 730, 760, 810, and 860 nm) were collected for each of the leaves. In both the experiments, reflectance data collection started at week seven after the sowing. For the greenhouse experiment, 87 fresh leaves were selected from 24 plants while 1008 leaves from the 336 plots were taken for the field experiment. Figure 6b,c shows the proposed setup being used for collecting spectral data. All the data were collected from three different positions around the midrib of a leaf by scanning 15 times. For data processing, the K-means clustering technique, which is widely used for pattern recognition, and plant stress phenotyping [35] was applied. K-means clustering was performed on the 15 scans by using the value of K as 3. From the three most different clusters, three centroids were selected corresponding to the reflectance from the three different positions. The reflectance data collected at 12 wavelengths from the leaves were used as features to classify N levels. The purpose of the classification was 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 in the greenhouse experiment and two categories in the field experiment.

Feature selection can be categorized into the filter, wrapper, and embedded methods [36]. Filter methods select sub-sets of variables without having information about the classifier. By contrast, wrapper and embedded methods use a fitting process such as the machine learning procedure for finding significant features. To shed light on it, different feature combinations were evaluated based on a predictive model in wrapper methods. On the other hand, embedded methods learn which set of features can feed to a model to achieve the best accuracy. In this paper, we used the embedded method in which features are weighted based on the particle swarm optimization (PSO) algorithm during learning. A similar approach was taken by Pourmohammadali et al. [37], wherein they utilized the genetic algorithm.

In the next step, the normalization technique was applied to data to bin the data in a similar domain when they were not in the same quantity domain. Normalization helps the classifier work more effectively since it changes the underlying probability distribution of features. In this paper, a modified standard score (z-score) was used for normalization [38]. Instead of using average, the median operator was also implemented to make the normalization robust against outliers [39].

The k-nearest neighbor (KNN) is a lazy learning classification method that works based on the closest training examples in feature space [40]. It is the simplest technique when there is no prior knowledge about the distribution of the dataset. The K was set to 1 in this research, founded by trial and error. KNN has been applied in many fields, such as leaf N and water content classification [38], plant leaf classification [41], and gene expression analysis [42]. Various machine learning approaches, including decision tree (DT), support vector machine (SVM), and ensemble bagged tree, are also used for comparison. SVM relies on a set of hyperplanes in a high or infinite-dimensional space. The radial basis function (RBF) is used as a kernel of SVM, and soft-margin and RBF kernel parameters are optimized using the Wang [43] method. It has been widely used in several research domains like plant disease detection [44] and optical remote sensing [45]. Furthermore, DT divides the dataset to small subsets based on the entropy of features, and creates a tree with leaf nodes, incrementally. In this research, decision tree C4.5 [46] is used. Moreover, ensemble bagged decision tree is also implemented in which a bag of the decision trees uses ensemble technique for aggregating results [47].

2.5. Evaluation Metrics

The final classification model needs a multiclass evaluation. The systematic analysis of classification, which was introduced by Sokolova and Lapalme [48], is used, and the validation metrics, including accuracy, sensitivity/recall, specificity, precision, and F1-score, are utilized. The definitions of these metrics are defined in the Supplementary Materials. For evaluation, data is divided into a train (75%), and test (25%) sets using the hold-out method. After training the system, the designed model is applied to the test set. Figure 7 shows the process flow diagram of this study, starting from the calibrating phase to the training and testing phase.

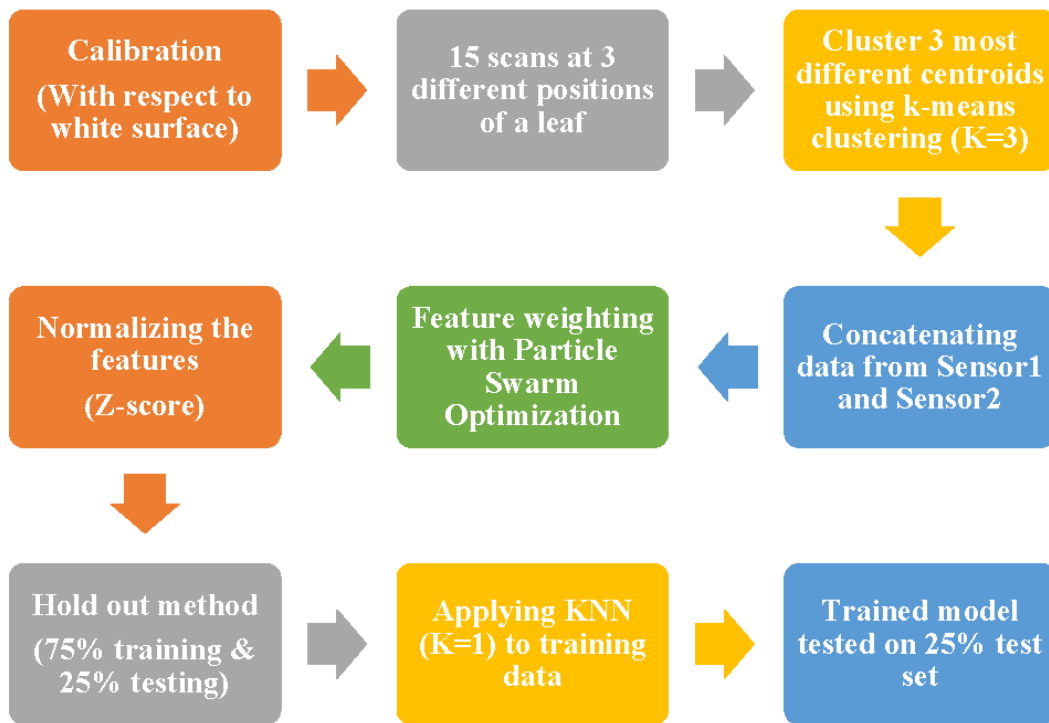


Figure 7. Process flow diagram of the methodology starting from calibrating the data with respect to the white surface to hold-out testing.

3. Result and Discussion

At first, the reflectance properties of the different intensity levels of fertilized nitrogen leaf in the greenhouse experiment were investigated. To serve the purpose, the average reflectance from each of the four N treatments at 12 wavelengths was plotted (Figures 8 and 9). These figures show that the average reflectance values of 20 g/L treatments are greater than other treatments. Additionally, the reflectance curves are distinctive respecting to N treatments, which suggests that the four N levels can be classified.

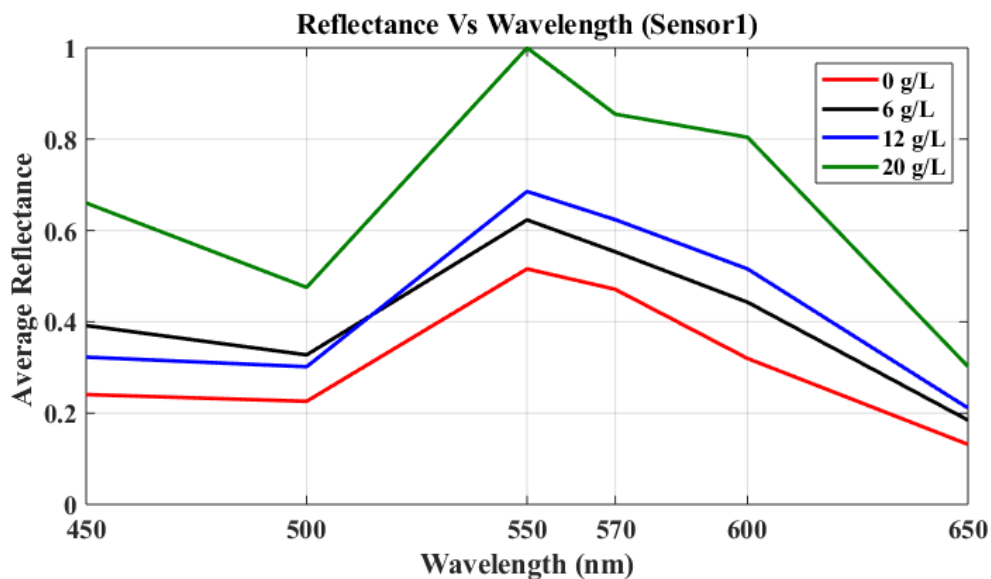


Figure 8. Average reflectance versus wavelength (Sensor1) of leaves subjected to four nitrogen fertilization regimes under visible range. The red line indicates the reflectance from 0 g/L plant, the black line represents 6 g/L, blue represents 12 g/L, and the green represents 20 g/L nitrogen rates. All the reflectance is scaled to the 20 g/L reflectance at 550 nm.

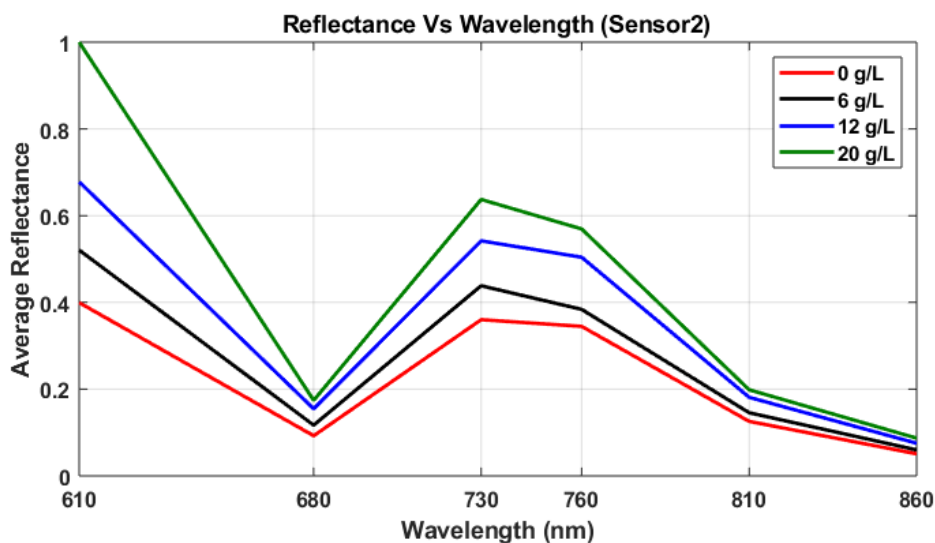


Figure 9. Average reflectance versus wavelength (Sensor2) of canola leaves subjected to four nitrogen fertilization regimes under the NIR range. The red line indicates the reflectance from 0 g/L plant, the black line represents 6 g/L, blue represents 12 g/L, and the green represents 20 g/L nitrogen rates. All the reflectance is scaled to the 20 g/L reflectance at 610 nm.

For classifying the captured data, the data set was divided into two parts: training set (75%) and testing set (25%), and the model was trained and tested five times with shuffling data. Finally, the average with a standard deviation of evaluation metrics was reported in Table 2, representing the class-wise (for each category) and overall results on the test set for the greenhouse growing canola plants.

Table 2. Category-wise/class-wise results of the test set for the greenhouse experiment from five test runs. The average \pm standard deviation is reported.

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

From the classification results shown in Table 2, it can be found that Category 4 has the best accuracy, and Category 2 has the least compared to others. One of the reasons might be the fact that there should some samples from Category 2 whose reflectance properties are nearly the same as Category 1. Later, a binary classification (two-class classification) was performed where Category 1 and Category 2 were combined; at the same time, Category 3 and Category 4 were combined. This means the classification is performed only in two classes. The results of the binary classification combining Category 1 and Category 2, and Category 3 and Category 4, are shown in Table 3.

Table 3. Results of the test set for the greenhouse and field experiments. For the greenhouse experiment, we use binary classification (Category 1 + Category 2) and (Category 3 + Category 4) five times running. The average \pm standard deviation is reported.

Environment	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
Greenhouse	94.2 \pm 1.8	90.9 \pm 2.1	100.0 \pm 0.0	90.9 \pm 1.9	95.2 \pm 1.8
Field	79.2 \pm 2.5	80.1 \pm 2.7	79.6 \pm 2.1	80.2 \pm 2.3	79.3 \pm 2.4

The binary classification of greenhouse results reveals that the accuracy, precision, recall, specificity, and F1-score have improved compared to distinct classes classification. Moreover, the standard deviation is decreased, which means a more consistent result has been acquired.

The field experiment results are also summarized in Table 3. 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 less compared to the test results achieved in the greenhouse experiment. Although the field experiment is all about classifying two categories of nitrogen, the combined dataset has high variance as it consists of 42 canola varieties. 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, in differentiating two nitrogen levels.

In this section, some other state-of-the-art machine learning algorithms such as decision tree, SVM, and 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. 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. Therefore, using the KNN algorithm for creating a model for this type of data is the best choice.

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

Training Algorithms	Testing Accuracy (%) of the Greenhouse Experiment	Testing Accuracy (%) of Field Experiment
Decision Tree	65.1 ± 2.7	73.6 ± 2.6
Support Vector Machine (SVM)	65.0 ± 2.0	70.1 ± 2.1
Ensemble Bagged Tree	73.3 ± 2.2	75.0 ± 2.3
K-Nearest Neighbor (KNN)	88.4 ± 3.0	79.2 ± 2.5

The comparisons in Table 4 illustrate that KNN outperforms decision tree, SVM, and ensemble bagged tree. This is not surprising because, in several previous comparative studies, researchers have shown similar results. For example, in a comparative study of SVM and KNN on classifying of an acoustic signal, KNN classifier shows better performance than SVM [49]. Additionally, KNN outperforms SVM in weather classification data [50]. Moreover, better accuracy is achieved by KNN compared to decision tree in the classification of Irish national forest inventory data [51]. In general, KNN can achieve higher accuracy when the number of training samples is much larger than the number of features [50]. However, there is no universal classifier which is consistently better at any task than others.

In this research, the reflectance at 12 different wavelengths has been used 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 10. 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.

Later, the comparisons between the proposed system and other similar devices are discussed. In terms of the number of wavelengths, most of them are based on 2–4 wavelengths. For example, SPAD operates only at one (650 nm), Multiplex at four (447, 590, 665, 735 nm), GreenSeeker at two (970 and 660 nm); whereas the proposed device detects at 12 wavelengths. Moreover, it is cheaper (USD \$200: Sensor1 \$25, Sensor2 \$25, multiplexer \$20, Raspberry Pi \$50, manufacturing overhead \$50 and misc. cost \$30; as of 13 March 2020) than all other comparable devices (GreenSeeker ~\$700, SPAD ~\$1500–\$2500).

The proposed multi-spectral sensors, along with the suitable machine learning algorithm, have been shown to be very cost-effective for determining leaf nitrogen status. This is the first time that a low-cost system was designed for measuring leaf nitrogen level using the reflectance at 12 wavelengths.

The proposed method was verified in greenhouse and field conditions. One of the limitations of this research is that the actual contents of nitrogen are not measured by the correlation, which will be the next phase of our project. Additionally, only four levels (greenhouse experiment) and two levels (field experiment) of nitrogen were considered in these experiments, and the effects with increasing nitrogen levels need further investigation. Furthermore, N is found basically as proteins and chlorophyll forms in leaf cells; however, nitrogen-containing biochemical constituents sensitive in visible and NIR spectral domains have been used in this study. One of the reasons of neglecting protein-related N forms is they are most sensitive in the SWIR spectral domain [52,53] (1020, 1510, 1980, 2060, 2130, 2180, 2300 nm, as confirmed by other research [54–56]), and the detectors in those regions are very expensive and bulky to incorporate into the existing setup. So, continuous exploration for cheaper detector technology is needed. In addition, applying the technique to other crop species at various growth stages has great potential for future research.

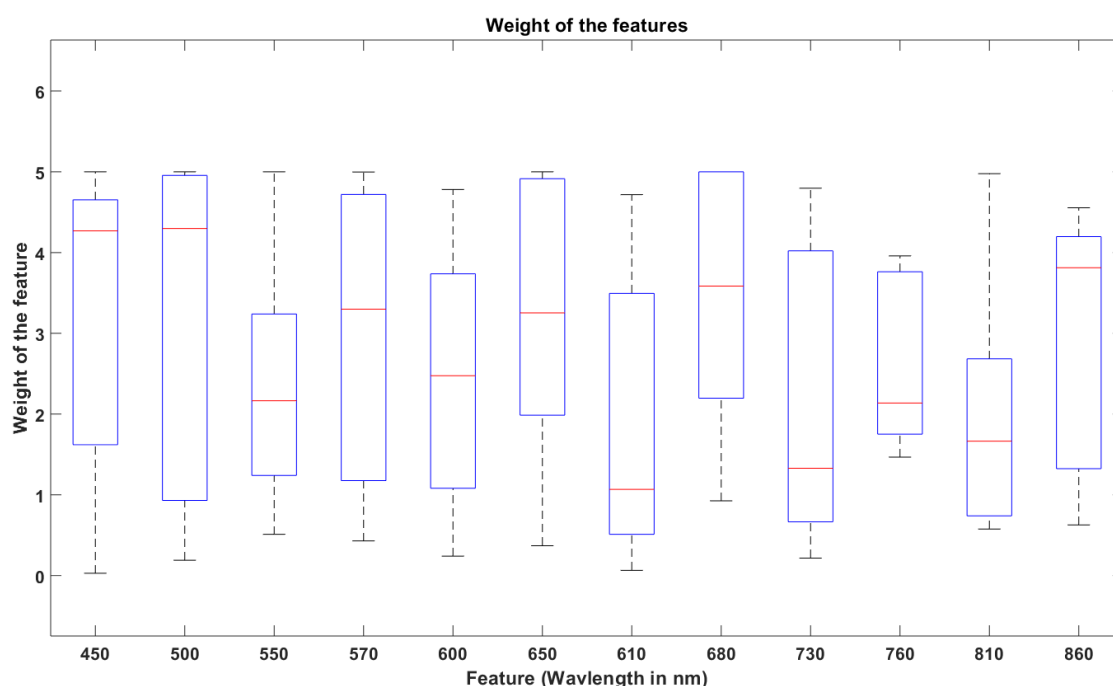


Figure 10. Box plots of the weights achieved by the features in five runs. Here, the 12 features are 450, 500, 550, 570, and 600 nm, and 650, 610, 680, 730, 760, 810, and 860 nm.

4. Conclusions

In this research, a low-cost and portable multi-spectral sensing system is proposed for classifying the N status of plants. The device can capture the reflectance at 12 wavelengths ranging from 450 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. To examine the performance of the proposed device, two individual experiments have been conducted. The KNN algorithm has been employed to model the collected data set. The model achieved 88.4% and 79.2% accuracy for the greenhouse and the field experiments, respectively. The results reveal that the proposed low-cost multi-spectral sensor array can classify leaves' nitrogen level with decent accuracy. It is worth mentioning that the overall cost of the proposed device is USD \$200, which is very cheap compared to existing technologies. Moreover, this device can be further modified to measure the other plant nutrients such as phosphorus (P), potassium (K), etc. Additionally, the device accuracy can be improved if more optical bands in the SWIR regions are incorporated.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2504-3129/1/1/7/s1>, Table S1: Validation metrics.

Author Contributions: Conceptualization, M.H., K.A.W., and A.D.; methodology, M.H. and R.S.; validation, M.H. and M.R.M.; formal analysis, M.H. and M.R.M.; investigation, M.H., R.S., and K.A.W.; resources, R.S., K.A.W., S.V., and A.D.; data curation, M.H. and M.R.M.; writing, M.H.; writing, review, and editing, M.R.M., A.N.B., K.A.W., R.S., S.V., and A.D.; supervision, K.A.W. and A.D.; funding acquisition, K.A.W. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is financially supported by The Canada First Research Excellence Fund through the Global Institute for Food Security, University of Saskatchewan to K.A.W. and by Agriculture and Agri-Food Canada (AAFC) to R.S.

Acknowledgments: The authors thank Shankar Pahari, AAFC and Mohammad Abdul Moin Oninda, Rakibul Chowdhury, and Rinku Basak from the University of Saskatchewan for assistance during the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Macko, S.A. Constrained preferences in nitrogen uptake across plant species and environments. *Plant. Cell Environ.* **2011**, *34*, 525–534. [[CrossRef](#)] [[PubMed](#)]
2. Sinfield, J.V.; Fagerman, D.; Colic, O. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Comput. Electron. Agric.* **2010**, *70*, 1–18. [[CrossRef](#)]
3. Vigneau, N.; Ecarnot, M.; Rabatel, G.; Roumet, P. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in wheat. *Field Crop. Res.* **2011**, *122*, 25–31. [[CrossRef](#)]
4. Padilla, F.M.; Gallardo, M.; Peña-Fleitas, M.T.; de Souza, R.; Thompson, R.B. Proximal optical sensors for nitrogen management of vegetable crops: A review. *Sensors* **2018**, *18*, 2083. [[CrossRef](#)] [[PubMed](#)]
5. Xiong, D.; Chen, J.; Yu, T.; Gao, W.; Ling, X.; Li, Y.; Peng, S.; Huang, J. SPAD-based leaf nitrogen estimation is impacted by environmental factors and crop leaf characteristics. *Sci. Rep.* **2015**, *5*, 1–12. [[CrossRef](#)]
6. Berger, K.; Verrelst, J.; Féret, J.-B.; Wang, Z.; Woche, M.; Strathmann, M.; Danner, M.; Mauser, W.; Hank, T. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* **2020**, *242*, 111758. [[CrossRef](#)]
7. Osborne, S.L.; Schepers, J.S.; Francis, D.D.; Schlemmer, M.R. Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements. *Agron. J.* **2002**, *94*, 1215. [[CrossRef](#)]
8. Wang, J.; Shen, C.; Liu, N.; Jin, X.; Fan, X.; Dong, C.; Xu, Y. Non-destructive evaluation of the leaf nitrogen concentration by in-field visible/near-infrared spectroscopy in Pear Orchards. *Sensors* **2017**, *17*, 538. [[CrossRef](#)]
9. Barbedo, J.G.A. Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Comput. Electron. Agric.* **2019**, *162*, 482–492. [[CrossRef](#)]
10. Muñoz-Huerta, R.F.; Guevara-Gonzalez, R.G.; Contreras-Medina, L.M.; Torres-Pacheco, I.; Prado-Olivarez, J.; Ocampo-Velazquez, R.V. A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances. *Sensors* **2013**, *13*, 10823–10843. [[CrossRef](#)]
11. Mercado-Luna, A.; Rico-García, E.; Lara-Herrera, A.; Soto-Zarazúa, G.; Ocampo-Velázquez, R.; Guevara-González, R.; Herrera-Ruiz, G.; Torres-Pacheco, I. Nitrogen determination on tomato (*Lycopersicon esculentum* Mill.) seedlings by color image analysis (RGB). *Afr. J. Biotechnol.* **2010**, *9*, 5326–5332.
12. Yu, K.Q.; Zhao, Y.R.; Li, X.L.; Shao, Y.N.; Liu, F.; He, Y. Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant. *PLoS ONE* **2014**, *9*, e116205. [[CrossRef](#)] [[PubMed](#)]
13. Krutz, D.; Müller, R.; Knodt, U.; Günther, B.; Walter, I.; Sebastian, I.; Säuberlich, T.; Reulke, R.; Carmona, E.; Eckardt, A.; et al. The Instrument Design of the DLR Earth Sensing Imaging Spectrometer (DESI). *Sensors* **2019**, *19*, 1622. [[CrossRef](#)] [[PubMed](#)]
14. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sens.* **2015**, *7*, 8830–8857. [[CrossRef](#)]
15. Staenz, K.; Mueller, A.; Heiden, U. Overview of terrestrial imaging spectroscopy missions. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, VIC, Australia, 21–26 July 2013; pp. 3502–3505.
16. Tomkiewicz, D.; Piskier, T. A plant based sensing method for nutrition stress monitoring. *Precis. Agric.* **2012**, *13*, 370–383. [[CrossRef](#)]
17. Blackmer, T.; Schepers, J.S. Techniques for monitoring crop nitrogen status in corn. *Commun. Soil Sci. Plant Anal.* **1994**, *25*, 1791–1800. [[CrossRef](#)]

18. Wang, L.; Okin, G.S.; Wang, J.; Epstein, H.; Macko, S.A. Predicting leaf and canopy 15N compositions from reflectance spectra. *Geophys. Res. Lett.* **2007**, *34*, L02401.
19. Stone, M.L.; Solie, J.B.; Raun, W.R.; Whitney, R.W.; Taylor, S.L.; Ringer, J.D. Use of spectral radiance for correcting in-season fertilizer nitrogen deficiencies in winter wheat. *Trans. Am. Soc. Agric. Eng.* **1996**, *39*, 1623–1631. [CrossRef]
20. Zhang, X.; Liu, F.; He, Y.; Gong, X. Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosyst. Eng.* **2013**, *115*, 56–65. [CrossRef]
21. Lawes, R.A.; Oliver, Y.M.; Huth, N.I. Optimal nitrogen rate can be predicted using average yield and estimates of soil water and leaf nitrogen with infield experimentation. *Agron. J.* **2019**, *111*, 1155–1164. [CrossRef]
22. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [CrossRef]
23. Sun, J.; Yang, J.; Shi, S.; Chen, B.; Du, L.; Gong, W.; Song, S. Estimating Rice Leaf Nitrogen Concentration: Influence of Regression Algorithms Based on Passive and Active Leaf Reflectance. *Remote Sens.* **2017**, *9*, 951. [CrossRef]
24. Pattern Recognition and Machine Learning. *J. Electron. Imaging* **2007**, *16*, 049901. [CrossRef]
25. Griffiths, T.L.; Ghahramani, Z. The Indian buffet process: An introduction and review. *J. Mach. Learn. Res.* **2011**, *12*, 1185–1224.
26. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]
27. SparkFun Spectral Sensor Breakout AS7262 Visible (Qwiic)-SEN-14347-SparkFun Electronics. Available online: <https://www.sparkfun.com/products/14347> (accessed on 25 February 2020).
28. Single-Chip Spectrometers | DigiKey. Available online: <https://www.digikey.be/en/articles/techzone/2017/jun/optical-sensor-on-chip-ics-simplify-handheld-spectrometer-design> (accessed on 27 August 2019).
29. AS7263 6-Channel Spectrometer-Ams | Mouser Canada. Available online: <https://www.mouser.ca/new/ams/ams-as7263-spectral-id-device/> (accessed on 27 August 2019).
30. Amazon.com. Raspberry Pi 3 Model B Board: Computers & Accessories. Available online: <https://www.amazon.com/Raspberry-Pi-MS-004-00000024-Model-Board/dp/B01LPLPBS8> (accessed on 25 January 2020).
31. Senthilkumar, G.; Gopalakrishnan, K.; Satish, K. Embedded image capturing system using raspberry pi system. *Int. J. Emerg. Trends Technol. Comput. Sci.* **2014**, *3*, 213–215.
32. Imteaj, A.; Rahman, T.; Hossain, M.K.; Alam, M.S.; Rahat, S.A. An IoT based fire alarming and authentication system for workhouse using Raspberry Pi 3. In Proceedings of the 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 16–18 February 2017; pp. 899–904.
33. Agrawal, N.; Singhal, S. Smart drip irrigation system using raspberry pi and arduino. *Int. Conf. Comput. Commun. Autom.* **2015**, 928–932.
34. Paper Mirror. Available online: https://www.amazon.ca/Cloakroom-Decorative-Background-Removable-Self-Adhesive/dp/B07W312F98/ref=asc_df_B07W312F98/?tag=googleshopc0c-20&linkCode=df0&hvadid=335179893508&hvpos=1o2&hvnetw=g&hvrnd=6497894200531816537&hvpone=&hvpstwo=&hqvmt=&hvdev=c&hvdvcmdl= (accessed on 4 February 2020).
35. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [CrossRef]
36. Rodriguez-Galiano, V.F.; Luque-Espinar, J.A.; Chica-Olmo, M.; Mendes, M.P. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Sci. Total Environ.* **2018**, *624*, 661–672. [CrossRef]
37. Pourmohammadali, B.; Salehi, M.H.; Hosseinifard, S.J.; Esfandiarpour Boroujeni, I.; Shirani, H. Studying the relationships between nutrients in pistachio leaves and its yield using hybrid GA-ANN model-based feature selection. *Comput. Electron. Agric.* **2020**, *172*, 105352. [CrossRef]
38. Habibullah, M.; Mohebian, M.R.; Soolanayakanahally, R.; Wahid, K.A.; Dinh, A. A cost-effective and portable optical sensor system to estimate leaf nitrogen and water contents in crops. *Sensors* **2020**, *20*, 1449. [CrossRef] [PubMed]
39. Borda, R.P.; Frost, J.D. Error reduction in small sample averaging through the use of the median rather than the mean. *Electroencephalogr. Clin. Neurophysiol.* **1968**, *25*, 391–392. [CrossRef]
40. Kumar, R.R.; Viswanath, P.; Bindu, C.S. Nearest neighbor classifiers: Reducing the computational demands. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 27–28 February 2016; pp. 45–50.

41. Mallah, C.; Cope, J.; Orwell, J. Plant leaf classification using probabilistic integration of shape, texture and margin features. In Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications, Innsbruck, Austria, 12–14 February 2013.
42. Parry, R.M.; Jones, W.; Stokes, T.H.; Phan, J.H.; Moffitt, R.A.; Fang, H.; Shi, L.; Oberthuer, A.; Fischer, M.; Tong, W.; et al. K-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *Pharm. J.* **2010**, *10*, 292–309. [[CrossRef](#)]
43. Wu, K.-P.; Wang, S.-D. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recognit.* **2009**, *42*, 710–717. [[CrossRef](#)]
44. Rumpf, T.; Mahlein, A.-K.; Steiner, U.; Oerke, E.C.; Dehne, H.W.; Plümer, L. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Comput. Electron. Agric.* **2010**, *74*, 91–99. [[CrossRef](#)]
45. Verrelst, J.; Camps-Valls, G.; Muñoz-Mari, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.P.W.; Moreno, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290. [[CrossRef](#)]
46. Salzberg, S.L. *C4.5: Programs for Machine Learning* by J. Ross Quinlan; Morgan Kaufmann Publishers: Burlington, MA, USA, 1994; Volume 16, pp. 235–240.
47. Banfield, R.E.; Hall, L.O.; Bowyer, K.W.; Kegelmeyer, W.P. A comparison of decision tree ensemble creation techniques. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 173–180. [[CrossRef](#)]
48. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
49. Palaniappan, R.; Sundaraj, K.; Sundaraj, S. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. *BMC Bioinform.* **2014**, *15*, 223. [[CrossRef](#)]
50. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. *Appl. Sci.* **2017**, *8*, 28. [[CrossRef](#)]
51. McInerney, D.O.; Nieuwenhuis, M. A comparative analysis of k NN and decision tree methods for the Irish National Forest Inventory. *Int. J. Remote Sens.* **2009**, *30*, 4937–4955. [[CrossRef](#)]
52. Homolová, L.; Malenovský, Z.; Clevers, J.G.P.W.; García-Santos, G.; Schaepman, M.E. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* **2013**, *15*, 1–16. [[CrossRef](#)]
53. Herrmann, I.; Karnieli, A.; Bonfil, D.J.; Cohen, Y.; Alchanatis, V. SWIR-based spectral indices for assessing nitrogen content in potato fields. *Int. J. Remote Sens.* **2010**, *31*, 5127–5143. [[CrossRef](#)]
54. Curran, P.J. Remote sensing of foliar chemistry. *Remote Sens. Environ.* **1989**, *3*, 271–278. [[CrossRef](#)]
55. Kumar, L.; Schmidt, K.; Dury, S.; Skidmore, A. Imaging Spectrometry and Vegetation Science. In *Imaging Spectrometry*; Van Der Meer, F.D., De Jong, S.M., Eds.; Kluwer Academic, Springer: Dordrecht, Germany, 2006; pp. 111–155.
56. Fourty, T.; Baret, F.; Jacquemoud, S.; Schmuck, G.; Verdebout, J. Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sens. Environ.* **1996**, *56*, 104–117. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).