*remote sensing*

MDPI

*Article*

# Determination of Leaf Nitrogen Concentrations Using Electrical Impedance Spectroscopy in Multiple Crops

**Rinku Basak *, Khan Wahid and Anh Dinh**

Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada; khan.wahid@usask.ca (K.W.); anh.dinh@usask.ca (A.D.)

* Correspondence: rib595@mail.usask.ca; Tel.: +1-306-203-6124

Received: 16 December 2019; Accepted: 7 February 2020; Published: 8 February 2020

**Abstract:** In this work, crop leaf nitrogen concentration (LNC) is predicted by leaf impedance measurements made by electrical impedance spectroscopy (EIS). This method uses portable equipment and is noninvasive, as are other available nondestructive methods, such as hyperspectral imaging, near-infrared spectroscopy, and soil-plant analyses development (SPAD). An EVAL-AD5933EBZ evaluation board is used to measure the impedances of four different crop leaves, i.e., canola, wheat, soybeans, and corn, in the frequency range of 5 to 15 kHz. Multiple linear regression using the least square method is employed to obtain a correlation between leaf nitrogen concentrations and leaf impedances. A strong correlation is found between nitrogen concentrations and measured impedances for multiple features using EIS. The results are obtained by PrimaXL Data Analysis ToolPak and validated by analysis of variance (ANOVA) tests. Optimized regression models are determined by selecting features using the backward elimination method. After a comparative analysis among the four different crops, the best multiple regression results are found for canola with an overall correlation coefficient ($R$) of 0.99, a coefficient of determination ($R^2$) of 0.98, and root mean square (RMSE) of 0.54% in the frequency range of 8.7–12 kHz. The performance of EIS is also compared with an available SPAD reading which is moderately correlated with LNC. A high correlation coefficient of 0.94, a coefficient of determination of 0.89, and RMSE of 1.12% are obtained using EIS, whereas a maximum correlation coefficient of 0.72, a coefficient of determination of 0.53, and RMSE of 1.52% are obtained using SPAD for the same number of combined observations. The proposed multiple linear regression models based on EIS measurements sensitive to LNC can be used on a very local scale to develop a simple, rapid, inexpensive, and effective instrument for determining the leaf nitrogen concentrations in crops.

**Keywords:** leaf nitrogen concentration; EIS; correlation coefficient; coefficient of determination; ANOVA; SPAD

## 1. Introduction

Nitrogen is the most required mineral nutrient of a crop due to its importance in several plant cell components; its concentration in plant tissue is the highest of all mineral nutrients. Leaf nitrogen concentration (LNC) provides valuable information about the physiological status of plants which is directly linked to photosynthetic potential and primary production [1–3]. Precise timing and the rate of nitrogen fertilizer application play a major role in plant nutrition. Nitrogen deficiency significantly reduces the photosynthetic yield of crops while excessive application of nitrogen fertilizer causes stress to the crop and environmental pollution. Therefore, the prediction of nitrogen requirements is necessary for efficient utilization of nitrogen fertilizers [1]. Soil tests or tissue tests are possible ways

to predict the nitrogen in plants, but these are expensive, laborious, and time-consuming. Applications of handheld chlorophyll meters, for example, soil-plant analyses development (SPAD), near-infrared spectroscopy, and hyperspectral imaging are mostly desirable as rapid, nondestructive, and noninvasive methods for predicting nitrogen in the leaves [4–6]. Leaf chlorophyll content is a key indicator of plant physiological status. It is correlated with leaf nitrogen concentration, although the correlation depends on soil condition, plant species, and the stage of growth [7–12]. The nitrogen status in the leaves can also be determined by analyzing the distribution of the color components of an image of a single leaf or group of plants.

In one study, leaf nitrogen status was determined by hyperspectral indices which were based on various algorithms [4]. In other studies, hyperspectral remote sensing images and spectral indices were used to assess the leaf nitrogen status and reflectance spectra of a wheat and reed canopy, as shown by the variation in wavelengths [13–15]. Hyperspectral data, fluorescence, and near-infrared spectroscopy, detected via digital cameras and satellite-mounted hyperspectral sensors, have been developed and studied for the detection of the nutritional status in crop fields. In addition, plant nitrogen status can also be correlated with laser-induced chlorophyll fluorescence.

Noninvasive methods involving leaf or canopy reflectance properties have been studied and applied mostly to determine crop N status. Canopy-level sensors are capable of measuring crop N status in larger areas by analyzing different reflectance spectra. The normalized differential vegetation index (NDVI) calculated using near-infrared (NIR) canopy reflectance has been studied to determine crop N requirements [13–15]. Plant-based N measurements and modeling approaches have been reported in various studies. The chlorophyll index and leaf nitrogen of canola have been evaluated under a wide range of soil moisture using SPAD [16]. The area- and mass-based leaf nitrogen of wheat have been estimated using continuous wavelet analysis [17]. The leaf nitrogen and chlorophyll of soybean plants have been measured using SPAD [18]. The leaf nitrogen of corn has been assessed from digital images using the dark green color index (DGCI) [19]. Although there are many advantages of these noninvasive methods, they have some limitations as to their environmental sensitivity and confounding factors (i.e., soil condition, light intensity, canopy shape, and color). Hyperspectral imaging creates images using hundreds of thousands of narrow bands. Although complete field imaging and estimation are done very rapidly using hyperspectral imaging, in most cases, the primary disadvantages are cost and complexity. Fast computers, sensitive detectors, and large data storage capacities are needed to analyze hyperspectral data. However, SPAD has limitations in the measurement of leaf nitrogen concentrations because the measurement is indirect and not linear, and the device is not cost effective. SPAD-based leaf nitrogen estimation is affected by environmental factors and the characteristics of individual crop species [20].

Several researchers have reported methodologies based on electrical impedance measurements to determine plant physiological status, such as nitrogen nutrition stress in tomato leaves [21], N status estimation in lettuce [22], citrus fruit acidity (pH measurement) [23], tea leaf growth [24], and other biological analyses [25]. The impedance measurement is done using electrical impedance spectroscopy (EIS), which is less sensitive to environmental variables than other available noninvasive methods. The nutrition status of trifolium subterraneum and tomato plants was also determined by electrical measurements using EIS [26,27]. EIS is a fast, nondestructive, easily implemented, and inexpensive method which could be an attractive alternative to optical spectroscopy for applications in plant science [28–30]. Impedance is very sensitive to the variation of frequencies set by the EIS tool, which is both convenient and easy to implement, but the computation is complex and model dependent. EIS works in a large range of frequencies operated by an electrical source and is easier to control than other noninvasive methods. EIS is proposed in this work to collect in situ data locally and directly on the leaf, which, then, is used for the prediction and validation of N status.

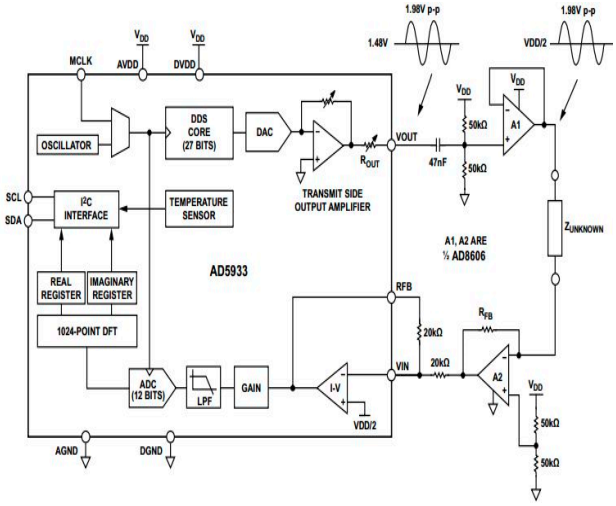
Therefore, the objectives of this study are the following: (i) To find the correlations between leaf nitrogen concentrations and leaf impedances of canola, wheat, soybeans, and corn, using simple and multiple regression analysis; (ii) to predict or determine the leaf nitrogen concentrations of the four different plant species using EIS with the help of multiple regression analysis; and (iii) to compare the performance of EIS with the SPAD measurement for the determination of leaf nitrogen concentrations.

## 2. Materials and Methods

The EVAL-AD5933EBZ evaluation board (Analog Devices Inc.) is a high-precision impedance converter system that combines an on-board frequency generator with a 12-bit, 1 mega sample per second (MSPS) analog-to-digital converter (ADC), and an internal temperature sensor. The schematic diagram of the board is shown in Figure 1a [28]. Both the excitation signal and response signal are sampled by the ADC. The frequency range of the board is from 5 to 100 kHz without external components. Frequencies lower than 5 kHz are achievable using an external divider. The device has a master clock of 16.77 MHz. Although the device is model dependent, it offers high accuracy and versatility for a well-fitted model, which makes it suitable for electrochemical analysis, corrosion monitoring, automotive sensors, proximity sensing, and bio-impedance measurements.

The experiments were carried out at the greenhouse of the Agriculture and Agri-Food Canada (AAFC), Research and Development Centre, Saskatoon, Saskatchewan, Canada, as shown in Figure 1b. The experimental setup of the EIS data acquisition system, as shown in Figure 1c, was connected to a graphical user interface of the supporting software. For impedance spectroscopy measurements, a $2V_{P\text{-}P}$ generator voltage was used. The AC signal injected into the sample was generated by a built-in function generator of the evaluation board. The frequency generator allows an external complex impedance to be excited with a known frequency. This portable impedance converter network analyzer was used in EIS for measuring the impedances of the four different plant leaves (e.g., canola, wheat, soybeans, and corn) by varying the frequency in a range of 5 to 15 kHz. A pair of electrodes for an electrocardiogram (ECG) were connected to the evaluation board and used to measure the impedance of the leaf samples noninvasively. A separation of 3 cm between the two electrodes was maintained for all the measurements. Although the EIS method is time-consuming for large crops, the test duration is short because of on-board implementation which enables measurements at particular frequencies.

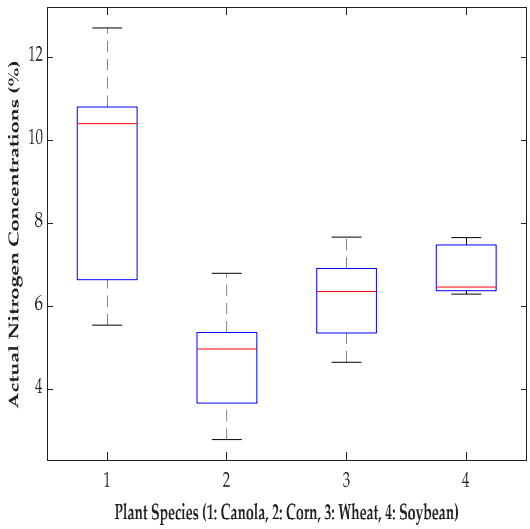
The leaf impedance measurement was done on a selected number of observations or samples ($n$) of the four different plant species. The plants were fertilized with different nitrogen levels of 0, 6, 12, and 20 g/liter with a constant water regime. A total of 111 samples were selected as follows: canola 26, wheat 36, soybeans 21, and corn 28. The experiments were carried out with the available number of samples of each plant species at the AAFC greenhouse 5 to 6 weeks after sowing. The measurements were performed in the vegetative growth stage of the crops. The impedance ($Z$) of a leaf sample was measured at 100 Hz intervals within the 5 to 15 kHz frequency range. A total of 101 features ($k$) were selected at different frequency points: *f1* (5.1 kHz), *f2* (5.2 kHz), *f3* (5.3 kHz), …, *f101* (15 kHz). The impedance at each frequency point is considered a feature. Therefore, the whole dataset of a particular plant species consisted of 101 features for the given samples.




(**a**)

(**b**)

(**c**)




(**d**)

**Figure 1.** (**a**) Schematic diagram of EVAL-AD5933EBZ evaluation board [28]; (**b**) plants in the Agriculture and Agri-Food Canada (AAFC) greenhouse; (**c**) impedance measurement of plant leaves using electrical impedance spectroscopy (EIS); and (**d**) boxplots of actual nitrogen concentrations, measured by the laboratory experiments, for four different plant species.

The impedance is related to the gain factor as follows:

$$Impedance, Z(Ohm) = \frac{1}{Gain\ Factor \times Magnitude} \quad (1)$$

The magnitude of the impedance can be calculated as

$$Magnitude = \sqrt{R^2 + X^2} \quad (2)$$

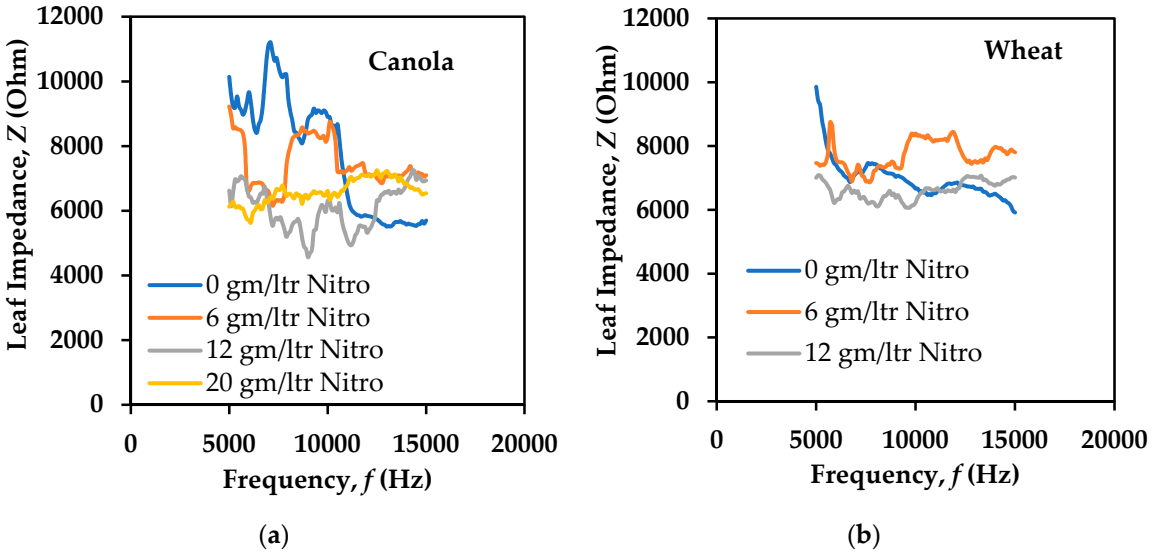
where $R$ is the resistance and $X$ is the reactance, and the gain factor is determined by the calibration using a known resistance of 7.5 kΩ [31].

After the impedances were measured, the samples with nitrogen concentrations were dried in a 60 °C incubator for 2 days. Then, the dry samples were weighted and made into powder. The actual percentages of nitrogen concentration (i.e., (nitrogen mg/mass) × 100) were measured from the powdered samples with the help of laboratory experiments using a LECO TruMac nitrogen analyzer,

where nitrogen mg = ((area × calibration) – blank) × drift × sensitivity factor. The obtained results for the different plant species are represented by the boxplots, as shown in Figure 1d. It was determined that the nitrogen concentrations are different in the different plant species. The size and area of the leaf samples vary with the different plant species, as well as their physiological properties. In the example, as shown in Figure 1d, canola has high nitrogen concentrations as compared with the other plant species.

The examination of two or three leaf samples from individual plants of each species shows the average impedance profiles of the samples at varying frequencies for different nitrogen fertilization levels (see Figure 2). It was determined that the leaf impedance of the plants decreases with an increase of frequency. Leaf impedance also decreases with an increase of nitrogen levels in the plants. The average impedance profile varies from 6 to 10 kohm with the variation of nitrogen fertilization levels from 0 to 20 gm/liter for a frequency range of 5 to 15 kHz. A high impedance profile is obtained for canola and corn as compared with soybeans and wheat. The measured impedances were examined to obtain correlations with the leaf nitrogen concentrations.

In this work, simple and multiple linear regressions using the least square method were applied to determine any correlations between plant leaf nitrogen concentrations and leaf impedances. The results were obtained by XLMiner and PrimaXL Analysis ToolPaks and validated by analysis of variance (ANOVA) tests. In multiple regression, the number of features was considered along with the observations in a given frequency range. In order to obtain optimized regression models, either the feature selection or dimensionality reduction (DR) method can be applied to reduce the number of features in a dataset. The feature selection method using backward elimination was selected and applied in this work. For different observations of the plant species, the nitrogen concentrations are predicted accordingly. The correlation coefficient ($R$) between leaf impedance and nitrogen concentration was determined. The corresponding coefficient of determination ($R^2$), adjusted $R^2$, and root mean square error (RMSE) were also determined along with the ANOVA F-test and T-test, using multiple linear regression. In this work, the training and tests were performed with the same dataset, using statistical analysis. After taking the whole dataset, the features within the highest $p$-value (i.e., greater than 0.05) were removed. The prediction was confirmed for all the trained datasets by the obtained $p$-values less than or equal to 0.05 in both the F-test and T-test. After a few iterations, the multiple regression models were obtained for the selected features of different crops.



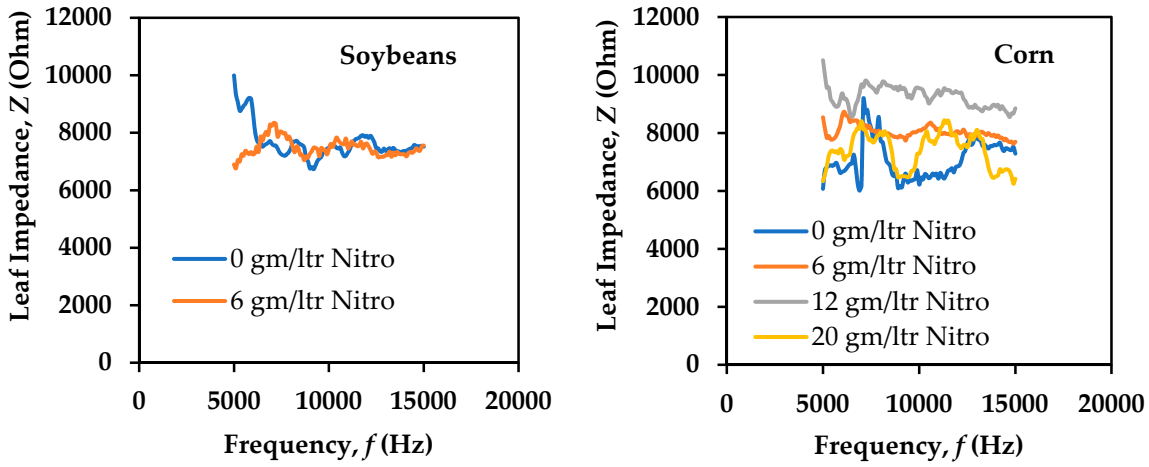
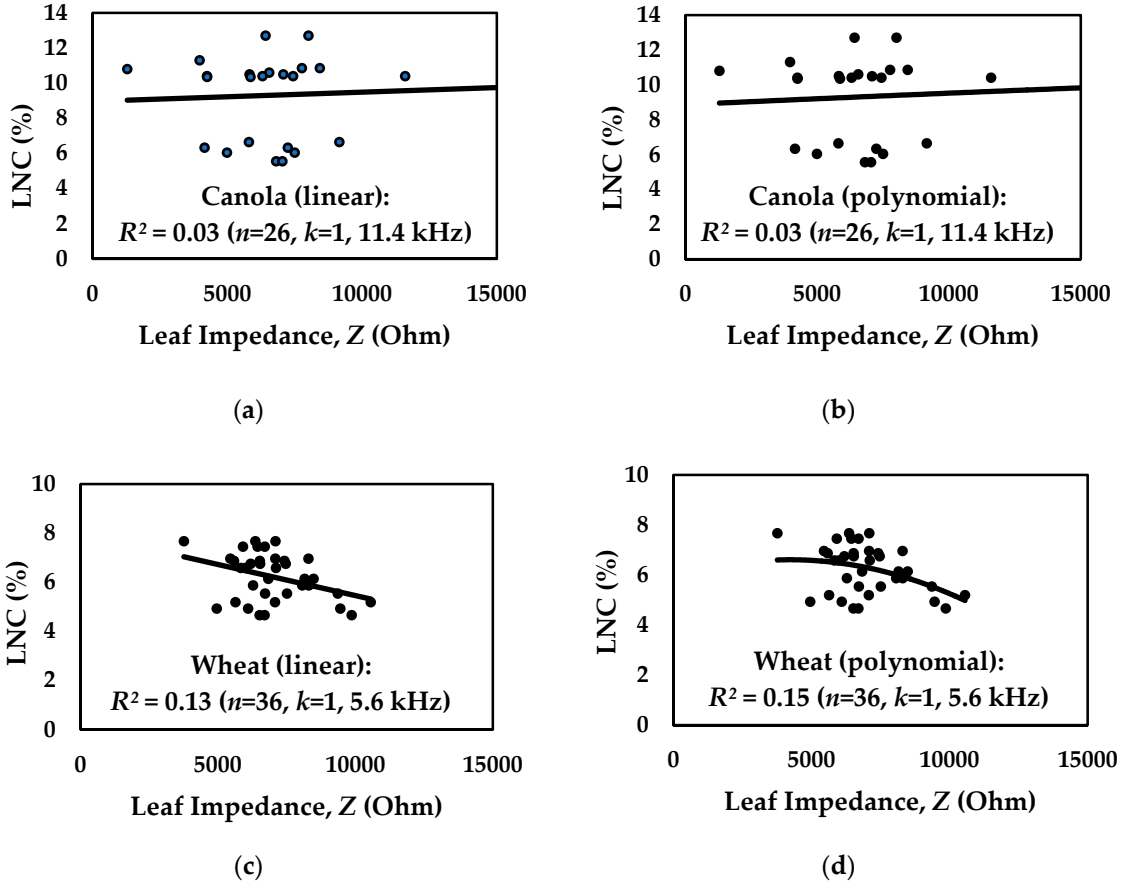



**Figure 2.** Plots of frequency versus leaf impedance. (**a**) canola; (**b**) wheat; (**c**) soybeans; and (**d**) corn at different nitrogen fertilization levels. The impedance profile for a few samples of wheat and soybeans could not be taken because of the effects of high nitrogen fertilization.

## 3. Results

First, with the use of simple regression analysis, the maximum correlations between leaf impedance (*Z*) and LNC for a single feature were found for canola, wheat, soybeans, and corn, as shown in Figure 3. A positive correlation for canola was obtained at 11.4 kHz, a negative correlation for wheat at 5.6 kHz, a positive correlation for soybeans at 7.7 kHz, and a positive correlation for corn at 11.9 kHz. The results are shown in Table 1.



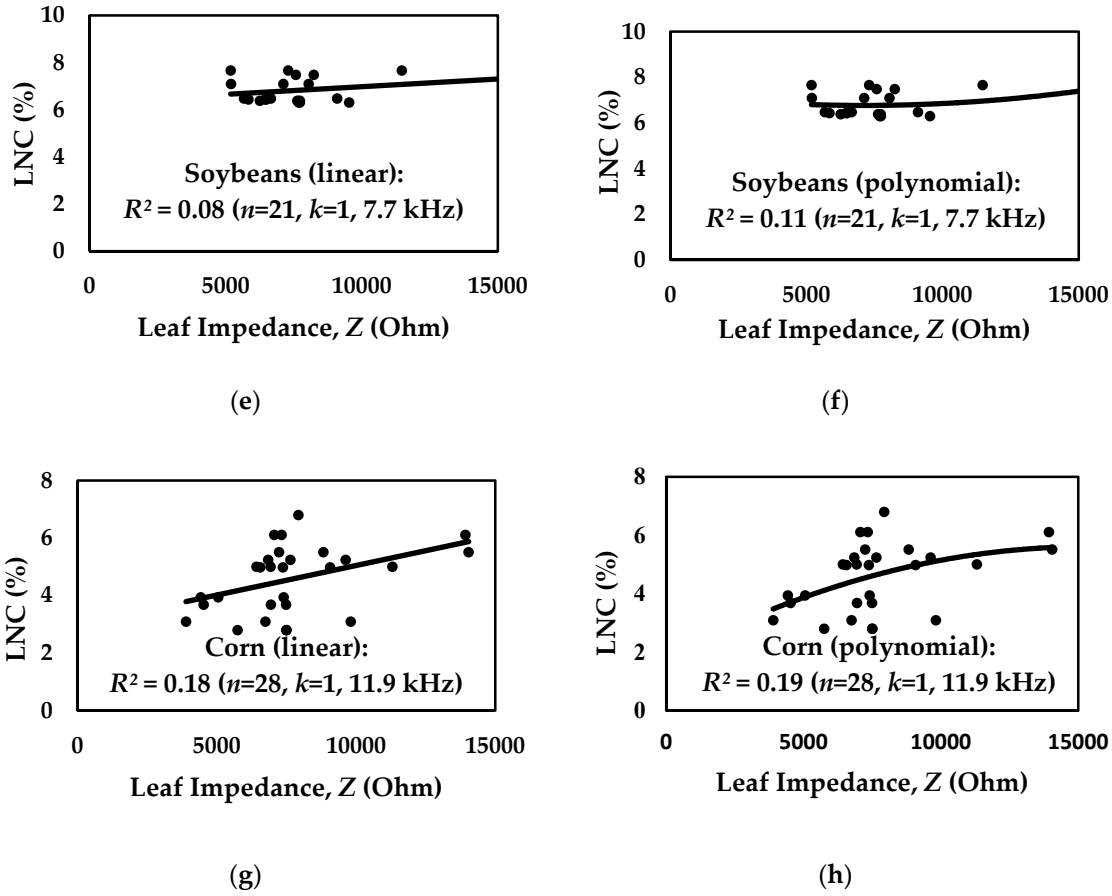



**Figure 3.** Correlations between leaf impedance and leaf nitrogen concentration (LNC) for four different plant species. The coefficient of determination ($R^2$) is extracted for canola (**a**) linear 0.03; (**b**) polynomial 0.03; for wheat (**c**) linear 0.13; (**d**) polynomial 0.15; for soybeans (**e**) linear 0.08; (**f**) polynomial 0.11; and for corn (**g**) linear 0.18; (**h**) polynomial 0.19.

Linear and polynomial (order 2) curve fitting methods and simple regression models were used for the different plant species at the highest correlation point. A better correlation was found for polynomial curve fitting as compared with linear in different frequencies of simple regression of the plant species. A maximum correlation coefficient ($R$) of 0.44 was obtained for corn, 0.39 for wheat, 0.34 for soybeans, and 0.19 for canola. On the one hand, based on a single feature, a moderate correlation was found for corn and wheat, on the other hand, the correlation was weak for soybeans and canola. Overall, the correlation results for different plant species are not satisfactory with simple regression analysis.

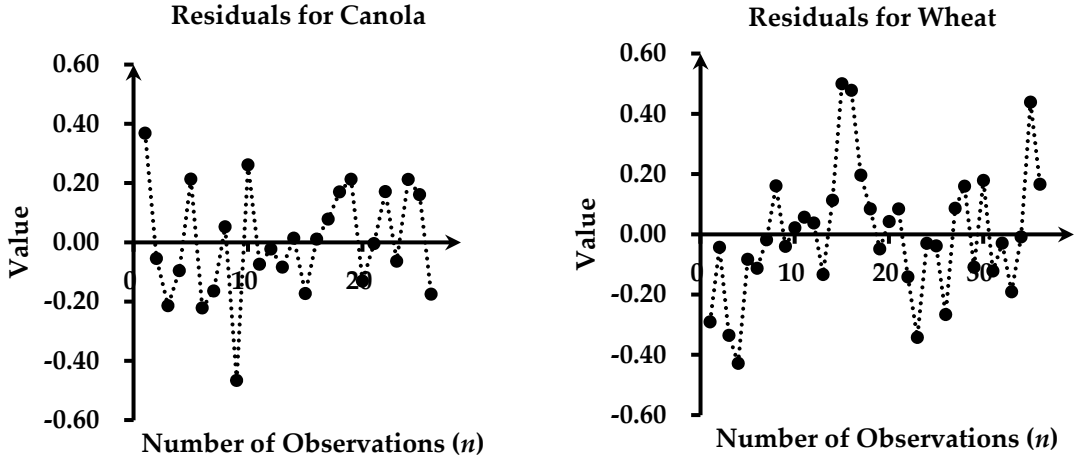
**Table 1.** Correlations between leaf nitrogen concentration (LNC) and leaf impedance ($Z$).for different plant species.

| Plant Species | Correlation Coefficient ($R$), Coefficient of Determination ($R^2$) and Simple Regression Model | |
|---|---|---|
| | **Linear** | **Polynomial (order 2)** |
| Canola | $R$ = 0.19, $R^2$ = 0.03 at 11.4 kHz<br>$LNC = 5E^{-05}Z + 8.9604$ | $R$ = 0.19, $R^2$ = 0.03 at 11.4 kHz<br>$LNC = -3E^{-10}Z^2 + 7E^{-05}Z$+8.8636 |
| Wheat | $R$ = -0.37, $R^2$ = 0.13 at 5.6 kHz<br>$LNC = -0.0003Z + 7.9957$ | $R$ = –0.39, $R^2$ = 0.15 at 5.6 kHz<br>$LNC = -4E^{-08}Z^2 + 0.0003Z + 5.8849$ |
| Soybeans | $R$ = 0.29, $R^2$ = 0.08 at 7.7 kHz<br>$LNC = 7E^{-05}Z + 6.3237$ | $R$ = 0.34, $R^2$ = 0.11 at 7.7 kHz<br>$LNC = 1E^{-08}Z^2 - 0.0001Z$ +7.2717 |
| Corn | $R$ = 0.42, $R^2$ = 0.18 at 11.9 kHz | $R$ = 0.44, $R^2$ = 0.19 at 11.9 kHz |

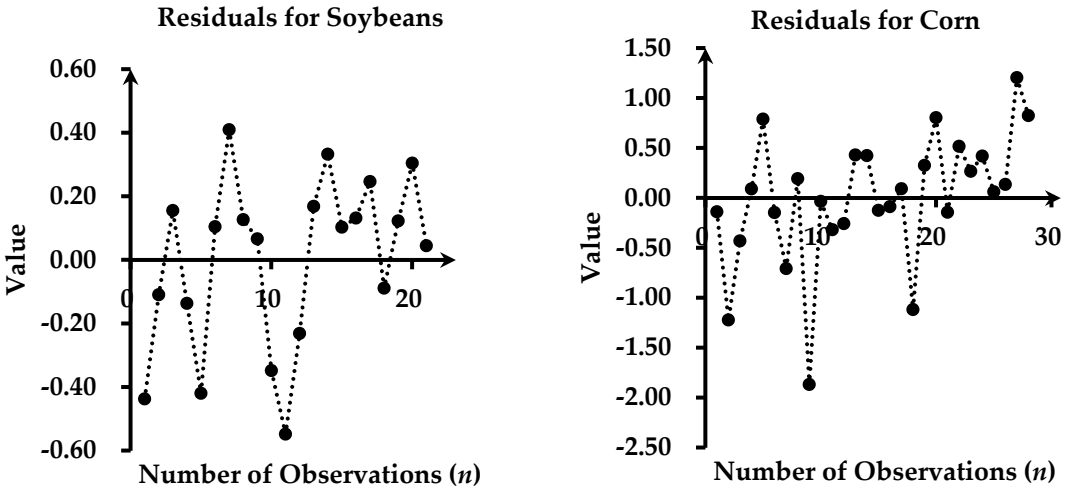
| | |
|---|---|
| $LNC = 0.0002Z + 2.9853$ | $LNC = -2E^{-08}Z^2 + 0.0005Z + 1.7835$ |

Next, multiple regression analysis was used to obtain better correlation results. Principal component analysis (PCA) is a popular dimensionality reduction (DR) approach of multiple regression and mostly applicable in hyperspectral image analysis, but it works extremely well for variables that are strongly correlated [32]. PCA is very useful in data analysis using machine learning. Since the correlations are poor between the variables, according to the above results of simple regression analysis, PCA would not perform well to reduce the features in a dataset. Hence, the feature selection approach using the backward elimination method was tried in order to obtain a good correlation with multiple regression. Initially, the number of features was selected from all the features, based on the number of observations ($n$) and number of features ($k = n - 2$), from the best correlation results obtained and validated by XLMiner Analysis ToolPak. For the given observations, the number of features was selected accordingly, using the standard backward elimination method to obtain the best correlation and regression results. The importance of the features was checked sequentially with the help of ANOVA F/T tests for obtaining the best multiple regression model. An optimization was done, and the best multiple regression results for the different plant species with nitrogen concentrations are summarized in the following section.

With multiple regression analysis, employing the least square method, the residuals were obtained for multiple observations with selected features using Residual value = actual value – predicted value. A random pattern of residuals supports a linear model. The sum of the residuals is always zero, whether the dataset is linear or nonlinear. The residuals for different observations and the corresponding best regression line between actual versus predicted nitrogen concentrations for the four different plant species are presented in Figures 4 and 5, respectively.




(**a**) (**b**)




(**c**) (**d**)

**Figure 4.** Residuals in different number of observations for (**a**) canola; (**b**) wheat; (**c**) soybeans; and (**d**) corn with nitrogen concentrations.

The coefficient of determination is calculated using the following:

$$R^2 = \frac{Variance\ explained\ by\ the\ model}{Total\ Variance} = \frac{SSR}{SST} \quad (3)$$

The adjusted $R^2$ is calculated as

$$R^2_{adj} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} \quad (4)$$

and the root mean square error (RMSE) is calculated as follows:

$$RMSE = \sqrt{\frac{SSE}{n-k-1}} \quad (5)$$

Here,

$$\text{Sum of Square Regression, } SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \quad (6)$$

$$\text{Sum of Square Residual, } SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

$$\text{and Sum of Square Total, } SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (8)$$

where $y$ is the actual LNC obtained by the experiment of laboratory measurement, $\bar{y}$ is the mean value of the actual LNC, and $\hat{y}$ is the predicted LNC obtained by multiple regression analysis using the least square method [33–36].

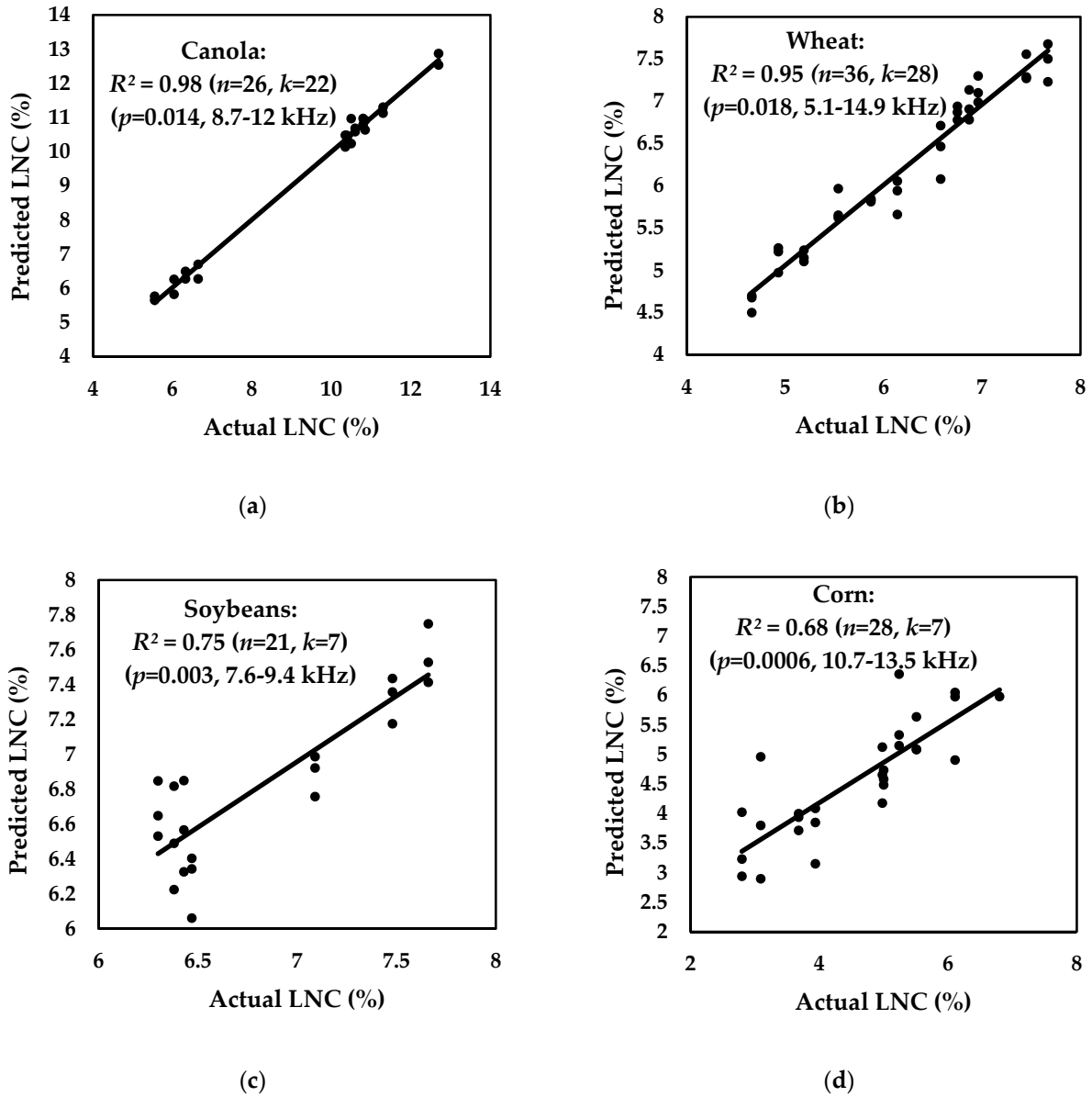



**Figure 5.** Multiple regression analysis for (**a**) canola; (**b**) wheat; (**c**) soybeans; and (**d**) corn. The extracted coefficient of determination ($R^2$) is for canola 0.98, for wheat 0.95, for soybeans 0.75, and for corn 0.68.

The predicted leaf nitrogen concentrations were validated by comparison with the actual leaf nitrogen concentrations. The overall multiple linear regression analysis results for different plant species are shown in Table 2. Overall, high correlation results were obtained using multiple regression analysis. The number of features was further reduced to avoid overfitting, and the regression results for different plant species are also included in Table 2. It was found that the correlation coefficient, the coefficient of determination, and adjusted $R^2$ decreased with the decrease of the features, and that the corresponding RMSE increased. The feature selection was done by positive ANOVA tests using *p*-value less than or equal to 0.05.

**Table 2.** Multiple linear regression analysis for four different plant species with nitrogen concentrations considering probability, $p \leq 0.05$ (probability of rejection of null hypothesis).

| Plant Species | | Overall Correlation Coefficient, $R$ | Overall Coefficient of determination, $R^2$ | Adjusted $R^2$ | RMSE (%) | ANOVA T-test/F-test |
|---|---|---|---|---|---|---|
| Canola ($n$ = 26) | $k$ = 22 (8.7–12 kHz) | 0.99 | 0.98 | 0.94 | 0.54 | Positive $p$ = 0.014 |
| | $k$ = 10 (8.8–11.5 kHz) | 0.85 | 0.72 | 0.54 | 1.56 | Positive $p$ = 0.008 |
| | $k$ = 9 (8.9–11.5 kHz) | 0.78 | 0.60 | 0.40 | 1.78 | Positive $p$ = 0.029 |
| Wheat ($n$ = 36) | $k$ = 28 (5.1–14.9 kHz) | 0.97 | 0.95 | 0.75 | 0.47 | Positive $p$ = 0.018 |
| | $k$ = 17 (5.2–14.9 kHz) | 0.86 | 0.74 | 0.50 | 0.67 | Positive $p$ = 0.011 |
| | $k$ = 11 (5.5–14.9 kHz) | 0.75 | 0.56 | 0.37 | 0.75 | Positive $p$ = 0.013 |
| Soybeans ($n$ = 21) | $k$ = 7 (7.6–9.4 kHz) | 0.86 | 0.75 | 0.62 | 0.33 | Positive $p$ = 0.003 |
| | $k$ = 5 (7.7–9.4 kHz) | 0.75 | 0.56 | 0.42 | 0.40 | Positive $p$ = 0.016 |
| | $k$ = 4 (7.7–9.4 kHz) | 0.70 | 0.49 | 0.37 | 0.42 | Positive $p$ = 0.018 |
| Corn ($n$ = 28) | $k$ = 7 (10.7–13.5 kHz) | 0.82 | 0.68 | 0.57 | 0.76 | Positive $p$ = 0.0006 |
| | $k$ = 4 (10.8–13.5 kHz) | 0.73 | 0.53 | 0.46 | 0.85 | Positive $p$ = 0.0008 |
| | $k$ = 3 (10.8–11.9 kHz) | 0.64 | 0.41 | 0.33 | 0.94 | Positive $p$ = 0.004 |

A maximum correlation coefficient of 0.99 is obtained for canola, using multiple features ranging from 8.7 to 12 kHz. The maximum coefficient of determination for canola is 0.98, the adjusted $R^2$ is 0.94, RMSE is 0.54%, and the ANOVA tests are positive; here, $SSR$ = 133.58, $SSE$ = 0.87, $SST$ = 134.46, and $p$-value = 0.014 (F-test). From 101 features, only 22 were selected to obtain the best correlation and regression results using ANOVA tests. Overfitting was reduced by the backward elimination of features with $p$-values greater than the threshold. The chances of overfitting could also be reduced by minimizing the features to 10 or nine, which would also reduce the corresponding correlation coefficient to 0.85 or 0.78, respectively. On the basis of the maximum correlation, the proposed model for the predicted nitrogen concentrations in canola for multiple features is extracted as:

$$\begin{aligned}\hat{Y} = {} & 16.164 - 0.0087X_{f38} + 0.03614X_{f39} + 0.02706X_{f40} - 0.0794X_{f41} + 0.02173X_{f42} \\ & - 0.034X_{f54} + 0.05424X_{f56} - 0.0276X_{f57} - 0.036X_{f58} \\ & + 0.03908X_{f59} + 0.0122X_{f60} + 0.0339X_{f61} - 0.0216X_{f62} \\ & - 0.0128X_{f63} + 0.04425X_{f64} - 0.0908X_{f65} + 0.0483X_{f66} \\ & + 0.02437X_{f67} - 0.0398X_{f68} + 0.01023X_{f69} + 0.01485X_{f70} \\ & - 0.0162X_{f71}\end{aligned} \quad (9)$$

where the 11th feature of 10.9 kHz, 13th feature of 11.1 kHz, 14th feature of 11.2 kHz, and the 20th feature of 11.8 kHz with $p$-values of 0.014, 0.02, 0.02, and 0.027, respectively, in the T-test contributed less to the model than the other features.

A maximum correlation coefficient of 0.97 is obtained for wheat, using multiple features ranging from 5.1 to 14.9 kHz. The maximum coefficient of determination for wheat is 0.95, the adjusted $R^2$ is 0.75, RMSE is 0.47%, and the ANOVA tests are positive; here, $SSR$ = 30.67, $SSE$ = 1.56, $SST$ = 32.23, and $p$-value = 0.018 (F-test). From 101 features, only 28 were selected to obtain the best correlation and regression results using ANOVA tests. Overfitting was reduced by the backward elimination of features with $p$-values greater than the threshold. The chances of overfitting could also be reduced by minimizing the features to 17 or 11, which would also reduce the corresponding correlation coefficient to 0.86 or 0.75, respectively. On the basis of maximum correlation, the proposed model for the predicted nitrogen concentrations in wheat for multiple features is extracted as:

$$\begin{aligned}\hat{Y} = -2.5493 &- 0.0071X_{f2} - 0.0027X_{f3} + 0.01242X_{f4} + 0.0225X_{f6} - 0.0392X_{f7} \\ &+ 0.01188X_{f8} - 0.0108X_{f9} + 0.01941X_{f10} + 0.00297X_{f12} \\ &- 0.0258X_{f13} + 0.03884X_{f14} - 0.0218X_{f15} + 0.01109X_{f22} \\ &- 0.0049X_{f23} + 0.00981X_{f25} - 0.0299X_{f26} + 0.00611X_{f27} \\ &+ 0.02869X_{f28} - 0.0145X_{f29} - 0.0212X_{f30} + 0.01165X_{f31} \\ &+ 0.02851X_{f32} - 0.0352X_{f33} - 0.0178X_{f34} + 0.03033X_{f35} \\ &- 0.0018X_{f40} - 0.0072X_{f99} + 0.00675X_{f100}\end{aligned} \quad (10)$$

where the 2nd feature of 5.2 kHz, 7th feature of 5.8 kHz, 9th feature of 6.1 kHz, 14th feature of 7.2 kHz, 15th feature of 7.4 kHz, 17th feature of 7.6 kHz, 21st feature of 8 kHz, 24th feature of 8.3 kHz, and the 26th feature of 8.9 kHz with $p$-values of 0.021, 0.011, 0.024, 0.01, 0.015, 0.041, 0.028, 0.013, and 0.016, respectively, in the T-test contributed less to the model than the other features.

A maximum correlation coefficient of 0.86 is obtained for soybeans, using multiple features ranging from 7.6 to 9.4 kHz. The maximum coefficient of determination for soybeans is 0.75, the adjusted $R^2$ is 0.62, RMSE is 0.33%, and the ANOVA tests are positive; here, $SSR$ = 4.41, $SSE$ = 1.44, $SST$ = 5.85, and $p$-value = 0.003 (F-test). From 101 features, only seven were selected to obtain the best correlation and regression results using ANOVA tests. Overfitting was reduced by the backward elimination of features with $p$-values greater than the threshold. The chances of overfitting could also be reduced by minimizing the features to five or four, which would also reduce the corresponding correlation coefficient to 0.75 or 0.70, respectively. On the basis of the maximum correlation, the proposed model for the predicted nitrogen concentrations in soybeans for multiple features is extracted as:

$$\begin{aligned}\hat{Y} = 6.521914 &- 0.00164X_{f27} + 0.002375X_{f28} + 0.000735X_{f37} - 0.00245X_{f40} \\ &+ 0.002303X_{f42} + 0.00118X_{f44} - 0.0025X_{f45}\end{aligned} \quad (11)$$

where the 3rd feature of 8.6 kHz with a $p$-value of 0.033 in the T-test contributed less to the model than the other features.

A maximum correlation coefficient of 0.82 is obtained for corn, using multiple features ranging from 10.7 to 13.5 kHz. The maximum coefficient of determination for corn is 0.68, the adjusted $R^2$ is 0.57, RMSE is 0.76%, and the ANOVA tests are positive; here, $SSR$ = 25.04, $SSE$ = 11.63, $SST$ = 36.68, and $p$-value = 0.0006 (F-test). From 101 features, only seven were selected to obtain the best correlation and regression results using ANOVA tests. Overfitting was reduced by the backward elimination of features with $p$-values greater than the threshold. The chances of overfitting could also be reduced by minimizing the features to four or three, which would also reduce the corresponding correlation coefficient to 0.73 or 0.64, respectively. On the basis of the maximum correlation, the proposed model for the predicted nitrogen concentrations in corn for multiple features is extracted as:

$$\begin{aligned}\hat{Y} = 6.521914 &- 0.00164X_{f27} + 0.002375X_{f28} + 0.000735X_{f37} - 0.00245X_{f40} \\ &+ 0.002303X_{f42} + 0.00118X_{f44} - 0.0025X_{f45}\end{aligned} \quad (12)$$

where the 1st feature of 10.7 kHz and the 6th feature of 13.1 kHz with $p$-values of 0.037 and 0.027, respectively, in the T-test contributed less to the model than the other features.

The proposed models using EIS are accurate and global for the individual plant species only. For each crop, different features were selected in the model based on the positive ANOVA tests because of different physiological properties. For canola and wheat, the individual features contribute less to the correlation, and thus a higher number of features are required as compared with soybeans and corn. The computation is complex in EIS and it is model dependent. For different plant species, datasets are different, and different models are required for the estimation. Appropriate fitting of the models shows the accuracy of the measurement using an EIS board.

Next, all the observations of different plant species were combined, and the multiple regression analysis was done using PrimaXL Analysis ToolPak. For multiple regression, a linear regression line was found between actual leaf nitrogen concentration and the predicted leaf nitrogen concentration for canola + wheat + soybeans + corn, using EIS, as shown in Figure 6. The coefficient of determination is 0.89 and the overall correlation coefficient is 0.94.

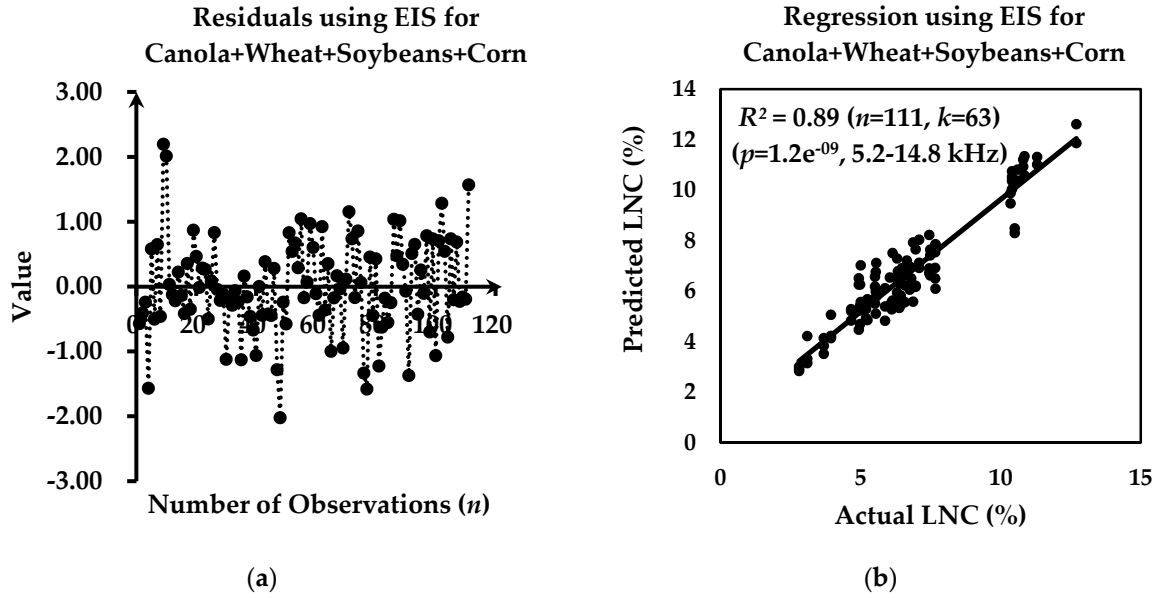



**Figure 6.** Plots of (**a**) number of observations versus value of residuals; and (**b**) actual versus predicted LNC for canola + wheat + soybeans + corn using EIS. The extracted coefficient of determination ($R^2$) is 0.89 and the corresponding overall correlation coefficient ($R$) is 0.94.

The SPAD leaf chlorophyll meter is a handheld, self-calibrating, and convenient device for a rapid and nondestructive assessment of leaf chlorophyll content in different crops. The leaf chlorophyll content is correlated to the leaf nitrogen concentration depending on the variety of plant species, locations, and growth stages [9,10,37]. For this, the measurement of leaf nitrogen concentration is possible using SPAD, and the relationship is curvilinear. SPAD measures the transmittance of red (650 nm) and infrared (940 nm) radiation through the leaf using two silicon photodiode detectors and has gained in popularity for its ease of use, although it is not as accurate as the destructive method. It utilizes the light attenuation difference between these two wavelengths to determine leaf greenness. Green color intensity of a crop leaf is directly related to the leaf nitrogen concentration and depending on the position of measurement on a leaf surface and area of the leaf, SPAD can have utility in predicting leaf nitrogen concentration [37]. For the combined observations, the performance of EIS as compared with SPAD readings is presented in Figure 7, and the summarized results are shown in Table 3.

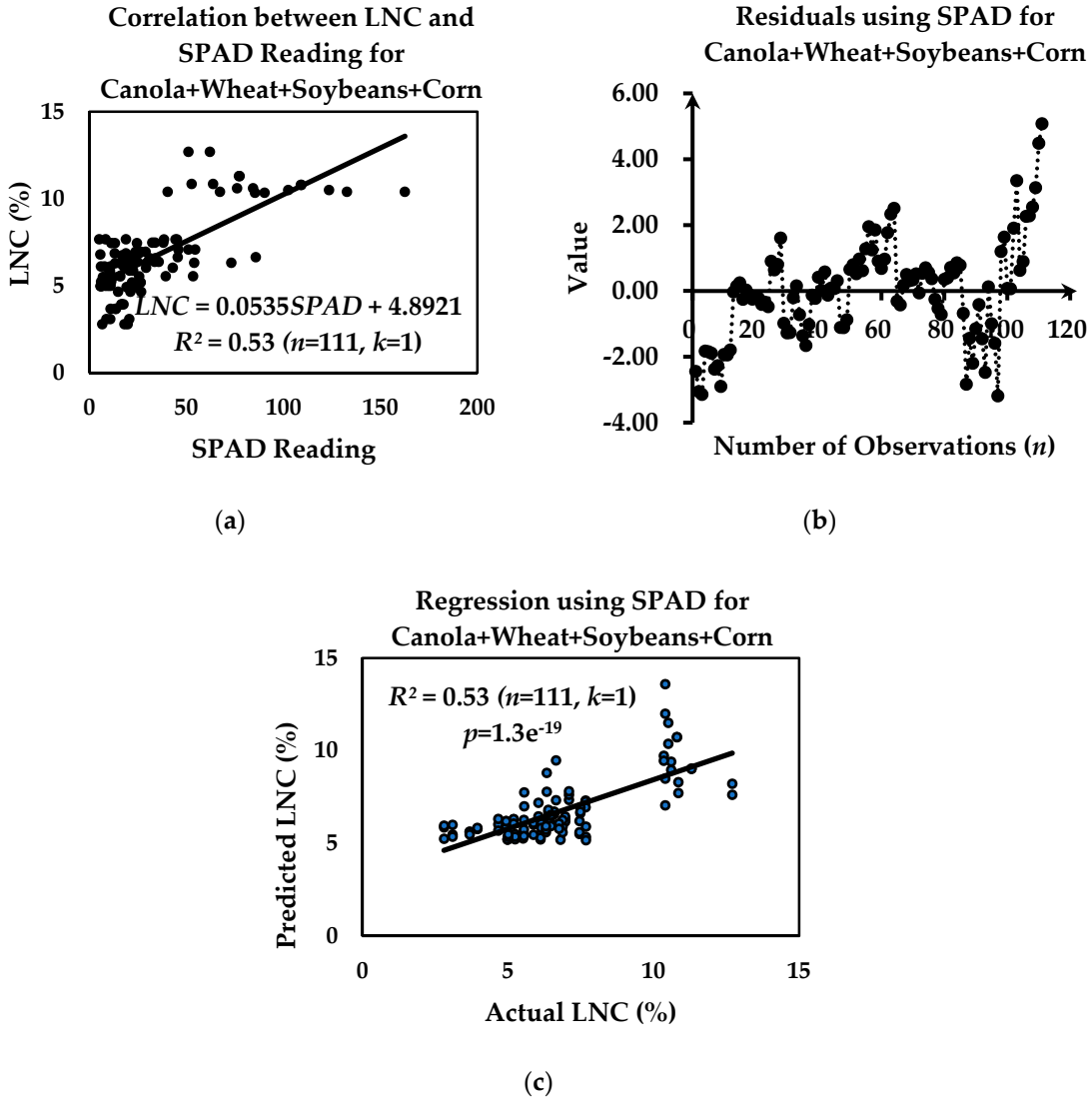



**Figure 7.** Plots of (**a**) soil-plant analyses development (SPAD) reading versus LNC; (**b**) number of observations versus value of residuals; and (**c**) actual LNC versus predicted LNC for canola+wheat+soybeans+corn using SPAD. The extracted coefficient of determination ($R^2$) is 0.53 and the corresponding overall correlation coefficient ($R$) is 0.72.

Using EIS, an overall maximum correlation coefficient of 0.94 and a coefficient of determination of 0.89 are obtained for the combined 111 observations of the plant species with nitrogen concentrations, where 63 features from 5.2 to 14.8 kHz were selected, and the RMSE is 1.12%. Overfitting is reduced by the backward elimination of features with $p$-values greater than the threshold. The chances of overfitting could also be reduced by minimizing the features from 63 to 33, which would also reduce the maximum correlation coefficient to 0.81. However, for the same number of observations using SPAD, a maximum correlation coefficient of 0.72 is obtained, where the coefficient of determination is 0.53 and the RMSE is 1.52%. Thus, EIS performs well as a good alternative to optical spectroscopy and to other nondestructive methods, such as SPAD, for the determination of leaf nitrogen concentrations.

**Table 3.** Comparative regression results between EIS and SPAD for the plant species with nitrogen concentrations using combined observations.

| | Canola + Wheat + Soybeans + Corn | Overall Correlation Coefficient (*R*) | Overall Coefficient of determination ($R^2$) | Adjusted $R^2$ | RMSE (%) | ANOVA T-test/ F-test |
|---|---|---|---|---|---|---|
| **EIS** | Observations, *n*: 111 Features, *k*: 63 (5.2–14.8 kHz) | 0.94 | 0.89 | 0.74 | 1.12 | Positive $p = 1.2e^{-09}$ |
| | Observations, *n*: 111 Features, *k*: 33 (5.3–14.8 kHz) | 0.81 | 0.66 | 0.52 | 1.53 | Positive $p = 1.7e^{-08}$ |
| **SPAD** | Observations, *n*: 111 Features, *k*: 1 | 0.72 | 0.53 | 0.52 | 1.52 | Positive $p = 1.3e^{-19}$ |

## 4. Summary and Discussion

In EIS measurement, it was determined that impedance varies with the variation of frequency for the four different crop leaves. The leaf impedance decreases with an increase of frequency and also with an increase of the nitrogen fertilization level. The actual nitrogen concentrations in the leaves were measured by a nitrogen analyzer, the samples were trained, and the nitrogen concentrations for all of them were predicted by regression analysis.

The correlation between actual nitrogen concentrations and the measured impedances of the leaves was found with simple regression analysis, but the obtained correlation with a single feature was not satisfactory. Therefore, multiple linear regression was also utilized to obtain better correlation results with the help of PrimaXL Toolpak. The selection of the number of features was challenging, but, along with the observations, played an important role in regression analysis. Removal of features using backward elimination had to be done very carefully, otherwise, the wrong selection may have affected the correlation and regression results. The overall correlation coefficient ($R$), coefficient of determination ($R^2$) and its adjusted value, and RMSE were calculated for the four different crops using Equations (3) to (8). After various experiments, it was found that the correlation coefficient and the coefficient of determination increased with an increase in the number of features for a given number of observations, and the corresponding RMSE decreased. The optimized selected features create a suitable model for good predictions.

The residuals were obtained from the difference between actual and predicted nitrogen concentrations. The lower residuals helped to achieve a good regression model for different observations. Multiple linear regression results, presented in Table 2, show that the highest correlation coefficient of 0.99 is obtained for canola, while 0.97 is obtained for wheat, 0.86 for soybeans, and 0.82 for corn. The corresponding RMSE values are 0.54%, 0.47%, 0.33%, and 0.76%, respectively. After training, the predicted results were tested and validated with the resulting positive ANOVA F-test and T-test, using $p$-values less than or equal to 0.05. The obtained results are satisfactory in comparison with previously published works [16–19]. The proposed models show that a large number of features are required for canola and wheat because the individual feature correlation is not very strong, a few features in the models contributed less to the correlation. The proposed models for soybeans and corn required a lower number of features because of strong individual feature correlation. However, a few features still contributed less to the correlation.

The performance of EIS was also compared with the SPAD measurement. A maximum correlation coefficient of 0.94 is obtained with a minimum RMSE of 1.12%, using EIS measurements for 111 observations, whereas, for the same observations and using SPAD, a maximum correlation coefficient of 0.72 is obtained, and the RMSE is 1.52%. Overall, satisfactory results are presented in

this work in comparison with previously published works on optical spectroscopy [4], [7], and on electrical impedance spectroscopy [21,22]. A strong correlation was found between nitrogen concentrations and impedances of the crop leaves measured by EIS with multiple features, and the nitrogen concentrations of the leaves were also determined accurately with the best multiple regression results.

## 5. Conclusions

In this work, a simple, inexpensive, rapid, and noninvasive EIS method is proposed for the prediction of the leaf nitrogen concentrations of canola, wheat, soybeans, and corn. The impedances of the plant leaves were measured with a variation of frequency from 5 to 15 kHz, using the EVAL-AD5933EBZ evaluation board. A correlation between leaf nitrogen concentrations and leaf impedances was found with the help of simple and multiple linear regression using the least square method. A comparative analysis among the four different crops was done, and the optimized regression results were obtained using ANOVA tests. With the use of multiple linear regression, of all the crops, the highest correlation of 0.99 is obtained for canola between 8.7 and 12 kHz, the corresponding coefficient of determination is 0.98, and the RMSE is 0.54%. New multiple linear regression models are proposed in this work for the determination of leaf nitrogen concentrations of the four crops. The overall prediction of nitrogen concentrations based on all the observations of the crops is better using EIS as compared with SPAD. The obtained correlation coefficient using EIS is 0.94 and the corresponding RMSE is 1.12%, whereas a maximum correlation coefficient of 0.72 and corresponding RMSE of 1.52% are obtained using SPAD for the same number of combined observations. The proposed models based on the portable EIS measurements are significantly sensitive to the nitrogen concentrations of the plant leaves.

**Author Contributions:** R.B. performed the experiments, analyzed the data, and wrote the draft of the manuscript; K.W. and A.D. suggested the experiments and data analysis, helped edit the draft, and provided critical comments to improve the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Canada First Research Excellence Fund (CFREF) (Grant no. 418406) through the Global Institute for Food Security (GIFS), University of Saskatchewan, Canada.

**Acknowledgments:** The authors would like to extend their appreciation to the staff of the Agriculture and Agri-Food Canada (AAFC), Research and Development Centre, Saskatoon, Saskatchewan, Canada, along with Raju Soolanayakanahally, Agroforestry Researcher and Shankar Pahari, for their support and access to facilities during the experimental period.

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

## References

1. 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. *Sens.* **2013**, *13*, 10823–10843.
2. Zhao, H.; Song, X.; Yang, G.; Li, Z.; Zhang, D.; Feng, H. Monitoring of Nitrogen and Grain Protein Content in Winter Wheat Based on Sentinel-2A Data. *Remote Sen.* **2019**, *11*, 1–25.
3. Evans, J.R. Improving Photosynthesis. *Plant Physiol.* **2013**, *162*, 1780–1793.
4. Liang, L.; Di, L.; Huang, T.; Wang, J.; Lin, L.; Wang, L.; Yang, M. Estimation of Leaf Nitrogen Content in Wheat Using New Hyperspectral Indices and a Random Forest Regression Algorithm. *Remote Sen.* **2018**, *10*, 1–16.
5. Riccardi, M.; Mele, G.; Pulvento, C.; Lavini, A.; d'Andria, R.; Jacobsen, S.-E. Non-destructive evaluation of chlorophyll content in quinoa and amaranth leaves by simple and multiple regression analysis of RGB image components. *Photosynth Res* **2014**, *120*, 263–272.
6. Odabas, M.S.; Kayhan, G.; Ergun, E.; Senyer, N. Using Artificial Neural Network and Multiple Linear Regression for Predicting the Chlorophyll Concentration Index of Saint John's Wort Leaves. *Commun. Soil Sci. Plant Anal.* **2016**, *47*, 237–245.

7. 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*, 1–15.
8. Shah, S.H.; Angel, Y.; Houborg, R.; Ali, S.; McCabe, M.F. A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. *Remote Sens.* **2019**, *11*, 1–26.
9. Mascarello1, A.C.; Benetoli da Silva1, T.R.; Gouveia1, B.T.; Bernardi, D.; Secco, D.; Santos, R.F.; Alves, C.Z. Chlorophyll meter reading and total nitrogen content applied as top dressing in parts of the crambe plant. *Afr. J. Biotechnol.* **2016**, *15*, 45–49.
10. Gholizadeh, A.; Saberioon, M.; Boruvka, L.; Wayayok, A.; Soom, M.A.M. Leaf chlorophyll and nitrogen dynamics and their relationship to lowland rice yield for site-specific paddy management. *Inf. Process. Agric.* **2017**, *4*, 259–268.
11. 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.
12. Bojović, B.; Marković, A. Correlation between Nitrogen and Chlorophyll Content in Wheat (Triticum aestivum L.). *Kragujev. J. Sci.* **2009**, *31*, 69–74.
13. Li, Z.; Jin, X.; Yang, G.; Drummond, J.; Yang, H.; Clark, B.; Li, Z.; Zhao, C. Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (*Triticumaestivum, L.*) Based on N-PROSAIL Model. *Remote Sens.* **2018**, *10*, 1–18.
14. Luo, J.; Ma, R.; Feng, H.; Li, X. Estimating the Total Nitrogen Concentration of Reed Canopy with Hyperspectral Measurements Considering a Non-Uniform Vertical Nitrogen Distribution. *Remote Sens.* **2016**, *8*, 1–13.
15. Ramoelo, A.; Cho, M.A. Explaining Leaf Nitrogen Distribution in a Semi-Arid Environment Predicted on Sentinel-2 Imagery Using a Field Spectroscopy Derived Model. *Remote Sens.* **2018**, *10*, 1–15.
16. Meskini-Vishkaee, F.; Mohammadim, M.H.; Neyshabouri, M.R.; Shekari, F. Evaluation of canola chlorophyll index and leaf nitrogen under wide range of soil moisture. *Int. Agrophys.* **2015**, *29*, 83–90.
17. Li, D.; Wang, X.; Zheng, H.; Zhou, K.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cheng, T. Estimation of area- and mass-based leaf nitrogen contents of wheat and rice crops from water-removed spectra using continuous wavelet analysis. *Plant Methods* **2018**, *14*, 1–20.
18. Fritschi, F.B.; Ray, J.D. Soybean leaf nitrogen, chlorophyll content, and chlorophyll a/b ratio. *Photosynth.* **2007**, *45*, 92–98.
19. Rorie, R.L.; Purcell, L.C.; Karcher, D.E.; King, C.A. The Assessment of Leaf Nitrogen in Corn from Digital Images. *Crop Sci.* **2011**, *51*, 2174–2180.
20. 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.
21. Meiqing, L.; Jinyang, L.; Xinhua, W.; Wenjing, Z. Early diagnosis and monitoring of nitrogen nutrition stress in tomato leaves using electrical impedance spectroscopy. *Int J Agric & Biol Eng* **2017**, *10*, 194–205.
22. Muñoz-Huerta, R.F.; Ortiz-Melendez, A.J.; Guevara-Gonzalez, R.G.; Torres-Pacheco, I.; Herrera-Ruiz, G.; Contreras-Medina, L.M.; Prado-Olivarez, J.; Ocampo-Velazquez, R.V. An Analysis of Electrical Impedance Measurements Applied for Plant N Status Estimation in Lettuce (*Lactuca sativa*). *Sens.* **2014**, *14*, 11492–11503.
23. Juansah, J.; Budiastra, I.W.; Dahlan, K.; Seminar, K.B. The Prospect of Electrical Impedance Spectroscopy as Non-destructive Evaluation of Citrus Fruits Acidity. *IJETAE*, **2012**, *2*, 58–64.
24. Mizukami, Y.; Yamada, K.; Sawai, Y.; Yamaguchi, Y. Measurement of Fresh Tea Leaf Growth using Electrical Impedance Spectroscopy. *Agric.* **2007**, *2*, 134–139.
25. Grossi, M.; Riccò, B. Electrical impedance spectroscopy (EIS) for biological analysis and food characterization: A review. *J. Sens. Sens. Syst.* **2017**, *6*, 303–325.
26. Greenham, C.G.; Randall, P.J.; Müller, W.J. Studies of phosphorus and potassium deficiencies in Trifolium subterraneum based on electrical measurements. *Can. J. Bot.* **1982**, *60*, 634–644.
27. Tomkiewicz, D.; Piskier, T. A plant based sensing method for nutrition stress monitoring. *Precis. Agric.* **2012**, *13*, 370–383.
28. Margo, C.; Katrib, J.; Nadi, M.; Rouane, A. A four-electrode low frequency impedance spectroscopy measurement system using the AD5933 measurement chip. *Physiol. Meas.* **2013**, *34*, 391–405.
29. Hamed, K.B.; Zorrig, W.; Hamzaoui, A.H. Electrical impedance spectroscopy: A tool to investigate the responses of one halophyte to different growth and stress conditions. *Comput Electron. Agric.* **2016**, *123*, 376–383.

30. Zhao, X.; Zhuang, H.; Yoon, S.C.; Dong, Y.; Wang, W.; Zhao, W. Electrical Impedance Spectroscopy for Quality Assessment of Meat and Fish: A Review on Basic Principles, Measurement Methods, and Recent Advances. *Food Qual.* **2017**, *2*, 1–16.
31. Jamaludin, D.; Aziz, S.A.; Ahmad, D.; Jaafar, H.Z.E. Impedance analysis of Labisia pumila plant water status. *Information Process. Agric.* **2015**, *2*, 161–168.
32. Liu, L.; Song, B.; Zhang, S.; Liu, X. A Novel Principal Component Analysis Method for the Reconstruction of Leaf Reflectance Spectra and Retrieval of Leaf Biochemical Contents. *Remote Sens.* **2017**, *9*, 1–24.
33. Zhang, Q.; Li, Q.; and Zhang, G. Rapid Determination of Leaf Water Content Using VIS/NIR Spectroscopy Analysis with Wavelength Selection. *Spectroscopy: An Int. J.* **2012**, *27*, 93–105.
34. Jin, X.; Shi, C.; Yu, C.Y.; Yamada, T.; Sacks, E.J. Determination of Leaf Water Content by Visible and Near-Infrared Spectrometry and Multivariate Calibration in Miscanthus. *Front. Plant Sci.* **2017**, *8*, 1–8.
35. El-Mohsen, A.A.A. Correlation and regression analysis in barley. *World Essays J.* **2013**, *1*, 88–100.
36. Mokarram, M.; Bijanzadeh, E. Prediction of biological and grain yield of barley using multiple regression and artificial neural network models. *Aust. J. Crop Sci. (AJCS)* **2016**, *10*, 895–903.
37. Wood, C.W.; Reeves, D.W.; Himelrick, D.G. Relationships between chlorophyll meter readings and leaf chlorophyll concentration, N status, and crop yield: A review. *Agron. Soc. N.Z.* **1993**, *23*, 1–9.