

Using Hyperspectral Signatures for Predicting Foliar Nitrogen and Calcium Content of Tissue Cultured Little-leaf Mockorange (*Philadelphus microphyllus* A. Gray) Shoots

Razieh Khajehyar

rkhajehyar@yahoo.com

Germaines Seed Technology <https://orcid.org/0000-0002-5726-2605>

Milad Vahidi

Virginia Tech University: School of Plant and Environmental Sciences

Robert Tripepi

University of Idaho

Research Article

Keywords: Tissue culture, Axillary shoot culture, Hyperspectral signatures, Machine learning, Leaf mineral estimation, Spectral characteristics

Posted Date: September 7th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3284303/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Plant Cell, Tissue and Organ Culture (PCTOC) on May 27th, 2024. See the published version at <https://doi.org/10.1007/s11240-024-02765-x>.

**Using Hyperspectral Signatures for Predicting Foliar Nitrogen and Calcium Content of
Tissue Cultured Little-leaf Mockorange (*Philadelphus microphyllus* A. Gray) Shoots**

Razieh Khajehyar¹, Milad Vahidi², Robert Tripepi¹

¹ Department of Plant Sciences, University of Idaho, Moscow, ID 83843-3702, USA

² College of Agriculture and Life Sciences, Virginia Tech University, Blacksburg, VA 24061 – 0493

Corresponding author: Razieh Khajehyar, rkhajehyar@gmail.com; khaj7814@vandals.uidaho.edu

ORCID: [0000-0002-5726-2605](https://orcid.org/0000-0002-5726-2605)

8 **Abstract**

9 Determining foliar mineral status of tissue cultured shoots can be costly and time consuming, yet hyperspectral
 10 signatures might be useful for determining mineral contents of these shoots. In this study, hyperspectral signatures
 11 were acquired from tissue cultured little-leaf mockorange (*Philadelphus microphyllus*) shoots to determine the
 12 feasibility of using this technology to predict foliar nitrogen and calcium contents. After using a spectroradiometer to
 13 take hyperspectral images for determining foliar N and Ca contents, the correlation between the hyperspectral bands,
 14 vegetation indices, and hyperspectral features were calculated from the spectra. Features with high correlations were
 15 selected to develop the models via different regression methods including linear, random forest (RF), and support
 16 vector machines. The results showed that non-linear regression models developed through machine learning
 17 techniques, including RF methods and support vector machines provided satisfactory prediction models with high R^2
 18 values (%N by RF with $R^2 = 0.72$, and %Ca by RF with $R^2 = 0.99$), that can estimate nitrogen and calcium content of
 19 little-leaf mockorange shoots grown in vitro. Overall, the RF regression method provided the most accurate and
 20 satisfactory models for both foliar N and Ca estimation of little-leaf mockorange shoots grown in tissue culture.

21 Keywords: Tissue culture, Axillary shoot culture, Hyperspectral signatures, Machine learning, Leaf mineral
 22 estimation, Spectral characteristics.

23 **Key Message**

24 Hyperspectral signatures were used to estimate foliar %N and %Ca of micropropagated *Philadelphus microphyllus*
 25 (little-leaf mockorange) shoots. Non-linear regression models provided satisfactory prediction with high R^2 values.

26 **Introduction**

27 Hyperspectral sensing is the measurement of the spectral characteristics of materials by the using sensing systems
 28 with more than 60 spectral bands and with spectral resolutions less than 10 nm. This resolution can produce a
 29 continuous portion of the light spectrum defining the chemical composition of an object through its spectral signatures
 30 (Gomez, 2020). With substantial developments in recording spectral bands of electromagnetic waves, hyperspectral
 31 sensors can provide data with a large number of spectral bands due to their high resolution in the range of 350 to 2500
 32 nm, and spectral bands are acquired by passive optical sensors. Spectral data are detected from any surface that can
 33 reflect, absorb, and transmit electromagnetic radiation (Hruška et al., 2018).

Hyperspectral imaging provides the ability to complete reflectance or fluorescence spectroscopy on all single spatial pixels of a spectral image thereby discerning characteristics that cannot be seen by human eyes (Robila, 2004, Gomez, 2020). The basic shape of a curve over the spectral range is characteristic of the parent material of the object being analyzed by spectroscopy (Liang, 2004). In the visible (VIS) to near infrared (NIR) spectrum (approximately between 400 and 1100 nm), characteristics of water, soil, or plant canopy give rise to specific curvatures in the reflectance spectrum, which makes them recognizable (Liang, 2004, Robila, 2004).

Perhaps the biggest advantage of hyperspectral data over simpler red-green-blue (RGB) imagery and multispectral data is that hyperspectral data can detect more accurate information of the object due to more spectral bands being recorded. Hyperspectral acquisition devices, including sensor types, acquisition modes and unmanned aerial vehicle (UAV)-compatible sensors, provide information that is needed or used both for research and commercial purposes, (Adão et al., 2017). Hyperspectral sensors and UAV have been useful in many areas of study including material identification, precision agriculture (vegetative coverage, nutrition deficiencies, foliar water content, physiological disorders, etc.), environmental aspects (wetlands, hydrology, etc.), health care (medical diagnoses, food safety, food quality assessment, etc.), and many more applied fields (Adão et al., 2017, Gomez, 2020).

A vegetative index (VI) describes an equation that processes spectral data for the purpose of determining information about plant health. Detectable vegetation indices (VIs) from hyperspectral signatures can provide an estimation and analysis of several plant characteristics, such as biophysical, physiological, or even biochemical parameters in crops, including leaf chlorophyll content (LCC), leaf water content (LWC), leaf area index (LAI), fractional photosynthetically active radiation (FPAR) absorbed by a canopy, surface roughness, and phenology, which are some of the most important inputs to land surface process models (Liang, 2004, Adão et al., 2017, Morcillo-Pallarés et al., 2019). These VIs can be applied in the regression models to help estimating plant status, such as foliar mineral contents.

With the importance of nitrogen increasing yield efficiency and crop health, modern application of hyperspectral signatures in preventing nitrogen deficiencies in the field have become widespread. Hence, much research has been conducted using remote sensing and applying hyperspectral signatures to determine crop nitrogen deficiency, required rates of fertilizers to increase crop production, or even the amount of nitrogen uptake by plants to improve agricultural production and yield efficacy (Maes and Steppe, 2019). DeOliveira et al. (2017) applied selected vegetation indices to estimate foliar N concentration in three Eucalyptus tree clones grown in the field. Liu et al. (2016) applied multiple

linear regression and neural network analysis to find a relationship between the leaf nitrogen content of field grown winter wheat and vegetative indices in narrow bands. Other studies have used hyperspectral indices to check the nutrition status of sodium and potassium content in grass (Capolupo et al., 2015), potassium deficiency level in canola (Severtson et al., 2016), nitrogen concentration in field grown oat (Van Der Meij et al., 2017), corn (Gabriel et al., 2017), rice (Wen et al., 2018), and wheat (Zhu et al., 2018), and leaf N, P, K, Ca, Mg, and few micronutrients of corn and soybean plants (Pandey et al., 2017).

Putting a new plant species into tissue culture to establish *in vitro* shoot cultures may require adjusting the nutrient medium components to optimize desirable shoot growth of the new species. Finding the optimum concentration of each component is critical and requires time and money. Estimating an explant's foliar mineral status to check its health status is important to attain optimal *in vitro* growth. Usually, destructive methods are applied to estimate foliar mineral contents, especially for tissue cultured plants. Finding nondestructive methods, such as applying hyperspectral signatures can help growers to reduce their production cost and save time.

To date, reports on using hyperspectral devices and hyperspectral vegetation indices in tissue culture environments are lacking. To check the feasibility of using of this technology to evaluate the mineral content of tissue cultured little-leaf mockorange shoots, we used a spectroradiometer during the shoot proliferation stage of micropropagation to determine if hyperspectral imaging could help in estimating nutrition status of the explants during stage 2. If hyperspectral imaging shows success, it can help tissue culture plant producers save money by avoiding destructive sampling for foliar nutrient analysis and save time waiting for nutrient analyses to be completed.

Materials and Methods

Plant materials and tissue culture

Little-leaf mockorange (*Philadelphus microphyllus*) explants were grown on half-strength Murashige and Skoog (½ MS) medium supplemented with different cytokinins, such as zeatin, kinetin (Kin), benzylaminopurine (BA), *meta*-Topolin (MT), thidiazuron (TDZ), or dimethylallylamino purine (2iP) or different concentrations of minerals such as N (0, 15, 22.5, 30, 37.5, 45, or 60 mM), Fe (0, 0.5, 5, 25, 50, 75, 100, or 500 µM). Six stem explants (per jar) were placed on the culture medium in baby food jars (195 ml) containing 0.5 mg·L⁻¹ thiamine-HCl, 0.25 mg·L⁻¹ nicotinic acid, 0.25 mg·L⁻¹ pyridoxine-HCl, 1 mg·L⁻¹ glycine, and 0.05 g·L⁻¹ myo-inositol, with pH = 5.6. Explants were

incubated in a SG-30S germinator (Hoffman Manufacturing Inc., Albany, OR) at $25 \pm 1^\circ\text{C}$ under a 16-h photoperiod (cool-white fluorescent lamps), with $38 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ photosynthetic photon flux (PPF), for 8 weeks with one subculture onto the fresh media after the 4th week. At the end of week eight, explants were harvested for collection of growth data and measurement of hyperspectral signatures.

Preparing the spectroradiometer and taking readings

For this research, we used either an Analytical Spectrum Devices FieldSpec 4 High-Resolution spectroradiometer (Malvern Panalytical Ltd., Westborough, MA, USA) or an Analytical Spectrum Devices FieldSpec HandHeld-2 spectroradiometer (Analytical Spectral Devices Company, Boulder, CO, USA). After 30 minutes of spectroradiometer warm up, the device was optimized and calibrated with a Spectralon[®] 99% white reference panel. During calibration, an average of 100 dark current measurements were calibrated together, and an average of 50 scans of the Spectralon[®] white reference were measured every two minutes (Labsphere Inc., North Sutton, NH, USA) (Beck, 2019). Target reference recordings displayed an average of 20 scans at an optimized integration time of approximately 1 second.

Reflectance readings of mockorange shoots were made immediately (within 2 minutes) after they were taken out of the jar and prior to completion of the reflectance spectra procedure. Measurements were completed in a dark-room and conducted on a black-colored bench to exclude external light and reduce outside lights. The probe was held about 5 to 10 cm over the explants to take the reflectance. Measurements were taken on all six shoots that were grown within each culture jar. Three duplicate readings were recorded for shoots grown in each jar in order to reduce error effects. After every 10 to 12 readings, a new calibration was completed to reduce the error from external white light. All measurements were acquired using RS3 software version 6.4 (Malvern Panalytical Ltd., Westborough, MA, USA).

Reflectance spectral data represented the full range of VIS, NIR, and short wave infrared (SWIR) light between 350 and 2500 nm, with a resolution of 1 nm. The spectral sampling interval was automatically interpolated from 1.4 nm to 1 nm at the time of each individual measurement by RS3 software, so a single value for each wavelength from 350 to 2500 nm was recorded (Beck, 2019). Data were exported by the ViewSpec Pro software version 6.2 (Malvern Panalytical Ltd., Westborough, MA, USA). The average of three readings of the reflectance from the group of six explants (per jar) was used to create a single treatment reflectance spectrum for each jar of shoots.

Tissue analysis for mineral content

After taking the hyperspectral reflectance, the shoots were separated from the agar medium, placed in an envelope and dried at 70°C for 72 hours. Dried shoots were ground using a pestle and mortar. Dried tissues were sent to the tissue analysis lab (Brookside Laboratories, Inc., New Bremen, OH) for foliar nutrient analysis. Tissue analysis was completed by using a combustion method applying a Carlo Erba 1500 C/N analyzer to estimate total N content (method B2.20, Miller et al., 2013). For Ca, lab procedures entailed use of nitric acid and hydrogen peroxide in a closed Teflon vessel and digested in a CEM Mars Microwave and analyzed on a Thermo 6500 Duo ICP (method B4.30, Miller et al., 2013). Results from foliar analyses were used for correlation model training with the hyperspectral signatures.

Hyperspectral data analysis

Preprocessing the spectral signatures was the first step in hyperspectral dataset analysis, particularly for spectra collected by the spectrometer. To further reduce noise, spectra were preprocessed with a Savitzky–Golay smooth filter (window size = 5 and polynomial order = 4) (Ge et al., 2019). The process of selecting an appropriate order and window size was done by trial and error, with the goal of smoothing out only large changes on a signature's surface.

The success of developing regression models is contingent upon the number of features assigned to the feature space (Zhao et al., 2019). Apart from the reflectance value at each wavelength, the hyperspectral dataset was used to extract spectral indices and geometric features from continuum removal regions. Thus, the number of features used for regression becomes even more critical when hyperspectral datasets are used; the large number of spectral bands makes determining whether spectral bands or spectral vegetation indices generated from spectral bands, or both, are associated with foliar chemical or physiological status, or in this case, leaf mineral content. To address this question, related features (explained in the following sections) were extrapolated from the spectra and then feature selection approaches were suggested for training the model with fewer but more informative features.

A. Spectral indices

Spectral indices defined by the mathematical operators between two or more spectral bands are also widely used for features extraction in remote sensing (Lu et al., 2020,). Many spectral indices used in agricultural applications are suitable for the specific purpose of plant monitoring. In this study, some commonly used spectral indices for mineral estimation were selected (Table 1).

B. Continuum removal

The absorption bands of the electromagnetic spectrum contain valuable information about the minerals or chemical compounds present in the target. This information has been used in various studies. Huang et al. (2004) and Gomez et al. (2008) used absorption features to estimate the amount of clay and calcium in the soil and the nitrogen concentration in a tree's canopy leaf surface, respectively.

Basically, the presence of organic components on the surfaces of plant leaves results in absorptions in the VIS and NIR wavelength ranges. These molecules include C-N, NH, and OH (Hunt, 1980), which indicate significant biochemical substances found on plant surfaces, such as lignin and starch, as well as nitrogen-containing components found in plants, such as protein and chlorophyll. These chemical and organic compounds may produce absorption in the spectral signature of plants due to the electron transfer phenomenon in the VIS region of the electromagnetic spectrum. On the other hand, specific absorptions in the SWIR region of the plant's spectral signature may be connected to the cellulose, glucose, and water content of the plant's leaf structure.

To demonstrate the geometrical differences between absorption regions, spectra need to be transformed into numerical features. To extract numerical information from the absorption region's surface, the spectrum's general concave shape must be ignored. This approach to normalization is referred to as “continuous removal” or “convex body”, and it enables the comparison of spectra acquired with various equipment or under varying lighting conditions (Sowmya and Giridhar, 2017).

The continuum removal, spectral signatures, and convex hulls of spectra can be shown graphically (Figure 1). Three characteristics are defined in this study by the geometry of the spectral signature following continuum removal. The depth, area, and asymmetry features in Figure 1 correspond to the continuum values at the lowest point of absorption, the area under the continuum curve in an absorption region, and the ratio of the left to right area. In this study, fifteen ranges in spectral signatures were selected (Table 2). To choose these spectral ranges, the spectral signature was carefully examined, and the absorption regions were selected based on a visual comparison between the absorption regions and the surrounding (left and right extremum) wavelengths.

The $Area_{Left}$ is the area of space between the continuum line and the continuum removed spectrum on the left, and $Area_{Right}$ is the area of space between the continuum line and the continuum removed spectrum on the right, the features are defined as follows (Aspinall et al., 2002):

- D = The absorption depth (the lowest point in continuum region)
- $Area = Area_{Left} + Area_{Right}$
- $Asymmetry = \frac{Area_{Left}}{Area_{Right}} = Asy$

For example, Asy 2 means the Asymmetry in the second wavelength range.

Model development

From the feature selection section, relevant features from spectral signatures were identified for tissue cultured shoots. The next steps were to 1) fit the regression model by using machine learning methods and 2) validate their significance using test data. Linear, Random Forest and Support Vector Machine were three regression models used in this research and are briefly explained below.

- **Linear Regression:** is a linear model that assumes a linear relationship between the input variable (x) and the single output variable (y). To select the relevant features, defined features (independent variables), such as reflectance values, continuum removal, and spectral indices for a linear model, a correlation test was used. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations (Freedman et al., 2007). Pearson correlation coefficient was used so that features with high correlation values were first recognized and selected from the list of defined features. In addition to a single variable linear model, a multi-variant linear model was also examined to determine the performance of different combinations of spectral features on the estimation results.
- **Random Forest Regression (RF):** This type of regression is a supervised learning algorithm that uses an ensemble learning method for regression and also is constructed by a set of decision trees. Group learning technique is combined with multiple decision trees to compare against a single regression model, enabling RF to obtain satisfactory and acceptable results for an R-square (R^2) value close to 1 or root mean square error (RMSE) close to zero, which shows ideal estimation. For this reason, RF has been widely used by researchers in regression and classification problems. The performance of a random forest model depends on

the number of trees and the input variables. Therefore, in this paper, different random forest regression models were trained to achieve the best model.

- Support-Vector Machines (SVM): This type of regression is a supervised learning model with associated learning algorithms that analyze data for classification and regression analysis. Various models can be produced based on changing the parameters in SVM, including the kernel type and the c-constant penalty term, which has the responsibility of balancing and maximizing the separator margin in features space (for example a two-dimensional space constructed by reflectance values in two wavelengths). In this study to reach an optimal model, parameter tuning was considered first by using RBF (Radial Basic Function), Linear and Polynomial (commonly used or built-in functions in SVM algorithm for transferring values of a variable to another space, these functions are known as kernels). Kernels and C values of 10, 100 and 1000 were used and then models with satisfactory results were selected.

To manage the results, the following procedures involved separately adding variables into the model and then calculating the coefficient of determination (R^2), RMSE and the correlation coefficient (Corr). Next, a combination of variables was added to the model (multiple-inputs) and then new calculations for R^2 , RMSE and Corr were made. The best model was chosen by comparing the results and using the best R^2 and Corr values and by using error bar plots and scatter plots. The error bar plots showed the error between observed and predicted values and the scatter plots showed the correlation between observed versus estimated values.

$$R^2 = \left[\frac{1}{N} \frac{\sum_{i=1}^N [(P_i - \bar{P})(O_i - \bar{O})]}{\sigma_p \sigma_o} \right]^2$$

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N [P_i - Q_i]^2 \right)^{1/2}$$

Where N is the number of observations, O_i is the observed values, P_i is the estimated values, $\bar{O}$ is the mean of the observational values, $\bar{P}$ is the mean of the estimated parameter and is the standard deviation of the observations and is the standard deviation of the estimated values.

Data partitioning

Data sets were divided into model training and model test groups for generating the optimum regression model. Data partitioning or splitting data sets (hyperspectral recorded samples) into training and sample (test) groups was one of the crucial steps in regression. In our case, 39 samples (reflectance spectra) out of 56 samples (70%) were used for model training and the rest of samples were used for model testing (17 samples out of 56 samples). The training data set was then used to develop a regression model with wavelengths in the spectral signature and vegetation indices calculated from those spectral signatures, as well as generated features obtained from those spectral signatures correlated to the foliar nutrient content from lab analysis. The developed model was validated and evaluated by using test datasets.

Model evaluation criteria (Statistical criteria for numerical evaluation of the developed model)

A schematic diagram of the methods used for developing a regression model from the hyperspectral bands and the mineral content in little-leaf mockorange shoots, is shown in Figure 2, and the evaluation criteria were calculated separately for foliar N or Ca contents.

The flowchart can be divided into the following steps:

- Step 1: Separately adding variables into model and calculation of R^2 , RMSE and Corr.
- Step 2: Adding a combination of variables (multiple inputs) into the model and then calculation of R^2 , RMSE and Correlation.
- Step 3: Comparing the results and choosing the best models given their performance in terms of evaluation criteria to be shown using error plot and scattering plots.
- Step 4: Plotting the best results in error plot (showing the error between observed and estimated values) and scatterplot (showing scattering of observed and predicted values to each other).

Results

The correlation between spectral features including spectral bands, spectral indices and continuum removal features were calculated. Spectral bands with higher correlation to leaf N content were used in regression model training (Figures 3 and 5). As shown, the wavelengths from 648 to 651 nm were shown to have a moderately high correlation with %N with correlation value of 0.30 (Figure 3). In general, leaf reflectance between 505 nm and 670 nm had the

highest correlation with N content of microshoots and was used for developing a linear model for N estimation (Figure 3).

Model development

Results showed that the reflectance values at the wavelength of 648 nm, asymmetric feature in range 1819 nm to 2150 nm (Asy 11) and the area from 559 to 772 nm (Area 3) had correlation values of 0.30, 0.31 and 0.37 with %N content (Figure 5). These spectral features provided information needed for predicting the %N to generate a linear model for N content measurement. The best single variable linear model was obtained by Asymmetric features in range 11 shown below.

$$\%Nitrogen = 4.47*(Asy\ 11) - 3.45$$

Based on these spectral data, N content acquired by a linear model was estimated by $R^2 = 0.21$, RMSE = 0.54 and Corr = -0.45 (Figure 4).

Random Forest regression was used in the next model. One of the main advantages of RF regression is that the number of input variables lack an effect on this model (Horning, 2010). The algorithm is able to apply the most effective variables given to entropy value, and then develop the regression model by using the most effective variables, meaning that RF algorithm could also be a feature selection. All the selected spectral bands from the correlation test and all the spectral features (indices and continuum removal) were added to the RF model. Based on the results, the RF regression model revealed that asymmetric point from 1819 to 2150 nm (Asy 11), asymmetric point from 559 to 772 nm (Asy 3), the reflectance values at the wavelength of 2480 nm, reflectance at wavelength of 525 nm, and the Double Peak Index (DPI) were the most effective features to generate a nonparametric (non-linear) model (Figure 5).

To develop a RF model, besides using optimal feature selection as effective inputs to the model, the number of trees in a RF model must be determined. By testing various models with different combinations of the mentioned features and/or indices, eventually the most accurate model was selected (Table 3). The fitted model with DPI index and reflectance at 525 nm and the tree number of 5 was a more accurate model fitted by RF regression, with $R^2 = 0.72$ and RMSE = 0.30, and correlation = 0.84 (Figure 6) compared to the other fitted models.

Support vector machine, one of the most commonly used regression methods, was used for the developing another regression model. In the SVM model, two main objectives were considered. First, the selected features from the

correlation test and those selected by RF methods were added to a SVM model. Second, the parameters of the SVM model, including the kernel type and the penalty term, were evaluated by trial and error such that the most accurate SVM model fitted by the optimal model had the lowest RMSE.

The model generated by SVM regression provided an estimation of foliar %N content that compared to the linear model, and the fitted SVM model including Double Peak Index (DPI) with asymmetric point from 1819 to 2150 nm (Asy 11) (Table 4). Another model including DPI with asymmetric point from 559 to 772 nm (Asy 3), provided an approximate accurate method to estimate foliar N content, respectively at $R^2 = 0.58$ and $RMSE = 0.32$, or $R^2 = 0.61$ and $RMSE = 0.33$ for little-leaf mockorange shoots produced in tissue culture (Figure 7).

Foliar calcium content

After analysis of the hyperspectral bands and checking for their correlation with the Ca content of the shoots received from the tissue analysis, the bands with higher correlations were selected, and those were 721 nm, 541 nm, 1293 nm, 1805 nm, and 2209 nm, with correlation values of 0.35, 0.33, 0.30, 0.28, and 0.26, respectively (Figure 8).

Examining the correlation values between %Ca with different features and VIs spectra showed that the minimum (depth) external of the wavelength between 1819 to 2150 nm (Min 11), and minimum (depth) external wavelength between 1287 to 1670 nm (Min 8) had the highest correlation values with Ca, respectively 0.59 and 0.45 (Figure 9).

Model development

Model development showed that Ca content determined by a linear model consisted of parameters of minimum (depth) external wavelengths between 1819 to 2150 nm (Min 11) and the area from 559 to 772 nm (Area 3) could be estimated by $R^2 = 0.83$ and $RMSE = 0.09$. Nevertheless, the coefficient of Area 3 was low enough to ignore it to draw the error bar graph (Figure 10).

$$\%Calcium = 1.13 * (Min11) + 0.08$$

The Random Forest algorithm provided a successful model to estimate the %Ca of little-leaf mockorange shoots. After examining several models with different feature combinations and tree numbers, the model including four features of minimum (depth) from 838 to 843 nm (Min 4), area from 2428 to 2490 nm (Area 15), asymmetric point from 1670 to 1714 nm (Asy 9), and Cellulose Absorption Index (CAI), with the tree number of 5 were the most

effective features to generate a nonparametric (non-linear) model (Figure 11, Table 5), yielding $R^2 = 0.99$ and RMSE = 0.03 and correlation value = 0.99 (Figure 12, right). The error bar plot in Figure 12 (left) reveals only slight differences between observed and estimated Ca among test samples proving the success of developed RF model for shoot Ca estimation.

Using the specific spectral features and a selected index (CAI) acquired from the RF algorithm as the best variables to use in model development. The specific spectral features and CAI index used for the RF algorithm were also used to develop a fitted model for SVM regression. After developing and running several models with different penalty terms (costs = 10, 50, or 100) and different kernels (linear, polynomial, or radial) (Table 6), eventually a model via linear kernel, including all four features of minimum (depth) reflectance from 838 to 843 nm (Min 4), area from 2428 to 2490 nm (Area 15), asymmetric point from 1670 to 1714 nm (Asy 9), and CAI was eventually developed. This model had a $R^2 = 0.59$ and RMSE = 0.16 and was determined to be the better model, regardless of the penalty term (cost value) (Table 6, Figure 13).

Discussion

Regression modeling plays an important role in estimating various plant characteristics, such as mineral content and water content. Accurate prediction of these parameters can assist in better understanding of plant growth and development, and improving agricultural practices. In this context, several regression models have been developed for hyperspectral data analysis, including the Random Forest (RF) and Support Vector Machine (SVM) models. In this study, we compared the performance of linear, RF, and SVM regression models in predicting the nitrogen (%N) and calcium (%Ca) content of tissue-cultured shoots. Additionally, we evaluated the importance of selecting the best features and wavelengths from the hyperspectral bands for accurate prediction. Our findings indicated that the RF model outperformed the SVM model in predicting %N, whereas %Ca was better predicted by the RF model with higher R^2 and lower RMSE values. These results demonstrated the importance of selecting the appropriate regression model and optimal features for hyperspectral data analysis in predicting plant characteristics.

This research demonstrated that hyperspectral imaging can be used to predict the percentages of N and Ca in little-leaf mockorange shoots produced in tissue culture. Linear, RF and SVM regression procedures were used to obtain an accurate model to estimate the %N and %Ca in little-leaf mockorange shoots produced in tissue culture. Among the three developed regression models used to estimate and predict the foliar nitrogen content, random forest

regressions and SVM, could estimate %N more accurately than the linear regression model. Nevertheless, the models developed to predict %N were slightly less accurate than those developed for predicting %Ca in the tissue cultured shoots.

The RF (tree number = 5) could estimate %N better than SVM (no matter what the cost (parameter or penalty term) used for this regression model). For %Ca, the RF model had a higher R^2 (0.99), had a lower RMSE (0.03) and provided a better model than SVM with a lower R^2 (0.59) and a higher RMSE (0.16). Finding the best regression model and the best features or indices as well as the best wavelengths throughout the hyperspectral bands is highly important for predicting a specific mineral content or other plant characteristics, such as water content.

Although the linear regression model provided an acceptable R^2 value, the model failed to predict %Ca. Hence, RF and SVM regression models were alternately considered. Based on the results obtained from this research, foliar %Ca content could best be estimated using a non-linear regression model rather than a linear model. Although the features used in the model (including the Cellulose Absorption Index) worked for both RF regression model and SVM regression model, the RF regression had stronger R^2 and correlation, and therefore was a better model to estimate the %Ca of tissue cultured shoots of little-leaf mockorange. Cellulose is an important component in the structure of primary cell walls of green plants (Anonymous, Cellulose, 2021). Calcium interacts with cellulose as a cellular structural component. A high correlation between %Ca and CAI is likely due to this relationship, and in the future more detailed experiments can be conducted to determine any possible relationship between %Ca and CAI index.

To date, research using hyperspectral images to estimate shoot mineral contents of shoots or plantlets produced in tissue culture (*in vitro*) is lacking. Studies, however, have been conducted to estimate N content of agronomic field crops, such as estimating N in winter wheat at different growth stages, based on NIR wavelengths via multivariate linear regression and Back Propagation (BP) neural network using vegetation indices (Liu et al., 2016), estimating leaf N content of winter wheat via selected spectral indices and around NIR wavelengths (Zhu et al., 2018), estimating N content in potato plants in NIR (Clevers and Kooistra, 2012), N estimation in maize via VIs, such as NDVI, Renormalized difference vegetation index (RDVI) or Optimized Soil-Adjusted Vegetation Index (OSAVI) (Gabriel et al., 2017), N estimation in rice with Gaussian process regression (GPR) model (Wen et al., 2018), N estimation of eucalyptus using NDVI in red-edge and modified red-edge NDVI (DeOliveira et al., 2017), and estimation of macro-

and micronutrients such as N and Ca in soybean and maize via partial least squares regression (PLSR) models (Pandey et al., 2017).

Although some reports describe the use of NIR or lower short wave infrared (SWIR) wavelengths to provide effective estimates of N, almost all of these studies have used only vegetation indices such as NDVI or other VIs. The difference between this study and other hyperspectral studies was application of different geometric features generated from continuum removal, such as minimum reflectance (depth), area under the spectrum, and asymmetric point of the spectrum alongside the reflectance spectrum acquired from little-leaf mockorange shoots produced in tissue culture. Applying these geometric features for plants grown in an *in vitro* environment, nevertheless, resulted in satisfactory R^2 and RMSE values obtained from the regression models used to predict N and Ca contents in the shoots.

An interesting aspect of %N and %Ca estimation was that both were predictable in spectrum ranges from 1819 to 2150 nm (Range 11) and from 559 to 772 nm (Range 3). Using different features of these ranges provided information for each of these two minerals in little-leaf mockorange shoots. In addition, correlation plots of estimated and measured values for N and Ca concentrations, revealed a small gap between higher concentrations and lower concentrations of these two minerals, probably due to the limited number of samples (less than 100) used for predicting their concentrations. The other possibility for the gap was that hyperspectral images could estimate N or Ca only at higher concentrations, due to the tiny size of the leaves and stems on the shoot cultures, meaning less information was acquired from their reflectance.

A deeper look at the scatter plot of %Ca obtained from the RF algorithm (Figure 12, error bar plot) showed that samples with values higher than 0.15 of CAI features had much lower differences between the measured and estimated values compared to the differences between measured and estimated values of CAI less than 0.15. This result indicated that for a more accurate prediction, features with higher correlation values must be selected. On the other hand, except for two samples (error bars shown in Figure 12, left), the developed model either accurately estimated or slightly over-estimated %Ca.

Most of the earlier foliar nutrient content studies have used mostly the vegetation indices to estimate canopy minerals especially N. Unfamiliarity with hyperspectral features relative to prediction of foliar mineral status may be a limitation on using of this technique in comparison with vegetation indices. Recruitment of a team of plant scientists,

plant nutritionists, and hyperspectral scientists, may provide an opportunity to apply these features more effectively. This study illustrated the potential for success of such a team of a plant scientists and hyperspectral scientists.

All these results were obtained from a specific selected mockorange genotype. Application of hyperspectral imaging was successfully completed for shoots from this little-leaf mockorange grown *in vitro*, but the success of this method for other mockorange species as well as other plant species still needs to be tested.

This study showed that hyperspectral imaging could help to predict foliar nutrient contents (N and Ca particularly) of little-leaf mockorange shoots produced in tissue culture and could help to avoid destructive methods of foliar mineral analysis. This nondestructive method, can save tissue culture producers the time necessary for drying, grinding, sending the samples off to a tissue analysis lab, and then waiting for the analysis, and the money by avoiding paying for shipping and foliar tissue analyses, enabling producers to save money.

Conclusion

This study demonstrated that strong regression models could be developed to predict N and Ca contents of tissue cultured little-leaf mockorange shoots. The best features to estimate %N were reflectance values at the wavelength of 648 nm, asymmetric point from 1819 to 2150 nm (Asy 11) and the area from 559 to 772 nm (Area 3), and reflectance at wavelength of 1919 nm. These features were used in a nonparametric (non-linear) model, with RF regression to provide the best model for estimation of foliar %N content. Best features to estimate %Ca in the shoots were minimum reflectance from 838 to 843 nm (Min 4), area from 2428 to 2490 nm (Area 15), asymmetric point from 1670 to 1714 nm (Asy 9), and Cellulose Absorption Index (CAI). Random forest regression provided a more accurate model to estimate %Ca than the other regression models. The best RF regression model for %N in little-leaf mockorange shoots resulted in an $R^2 = 0.72$ and correlation = 0.84. Likewise, the best RF model for %Ca estimation resulted in an $R^2 = 0.99$ and correlation = 0.99. These strong statistical values clearly demonstrated that hyperspectral imaging can be used to predict accurately %N and %Ca in tissue cultured shoots from one selected little-leaf mockorange genotype. Other mockorange species as well as other plant species produced in tissue culture would need to be tested to validate using hyperspectral imaging to predict N and Ca contents of their shoots.

392 **Acknowledgments**

393 This research was supported by the USDA National Institute of Food and Agriculture, Hatch/Evans project IDA01545.
394 We thank the Idaho Agricultural Experiment Station for supporting this research. We also thank Stephen Love for
395 providing the mockorange plant used in this study. We also thank William Price and Stephen Love for their editorial
396 advice.

397 **Disclosures and declarations**

398 This research was supported by the USDA National Institute of Food and Agriculture, Hatch/Evans project IDA01545.
399 Also, the Idaho Agricultural Experiment Station partially supported this research.

400 **Competing Interests**

401 The authors have no relevant financial or non-financial interests to disclose.

402 **Author contributions** To complete this research project, Razieh Khajehyar brought up the idea of applying remote
403 sensing and hyperspectral signature in the tissue culture scale, designed and completed the research, analyzed the
404 hyperspectral and statistical data, developed the models, and wrote the manuscript as part of her PhD dissertation.
405 Milad Vahidi a graduate student from Virginia Tech, trained Razieh to accomplish the hyperspectral analysis, and
406 helped with the machine learning part, data analysis and data interpretation. Robert Tripepi directed the research as
407 Razieh's PhD supervisor, wrote small parts of the manuscript, and edited the manuscript. All authors revised the final
408 manuscript.

409 **Data Availability Statements** The datasets generated and/or analyzed during the current study are not publicly
410 available but are available from the corresponding author on reasonable request.

411

412

413

414

References

- Adão T, Hruška J, Pádua L, Bessa J, Peres E, Morais R, Sousa JJ (2017) Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens*, 9:1110; doi:10.3390/rs9111110.
- Anonymous (2021) Cellulose. Accessed in December 2021. <https://en.wikipedia.org/wiki/Cellulose>.
- Anonymous, 2021. Index Database. Accessed in October 2021. <https://www.indexdatabase.de/>.
- Aspinall RJ, Marcus WA, Boardman JW (2002) Considerations in collecting, processing, and analyzing high spatial resolution hyperspectral data for environmental investigations. *J Geograph Syst*, 4:15–29. DOI: 10.1007/s101090100071.
- Beck KD (2019) Evaluating the use of hyperspectral remote sensing and narrowband spectral vegetation indices to diagnose onion pink root at the leaf and canopy level. M.Sc. Thesis. University of Idaho.
- Capolupo A, Kooistra L, Berendonk C, Boccia L, Suomalainen J (2015) Estimating plant traits of grasslands from UAV-acquired hyperspectral images: A comparison of statistical approaches. *ISPRS Int. J. Geo-Inf*, 4: 2792-2820.
- Clevers JGPW, Kooistra L (2012) Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 5:574–583. DOI: 10.1109/JSTARS.2011.2176468.
- DeOliveira LFR, DeOliveira MLR, Gomes FS, Santana RC (2017) Estimating foliar nitrogen in *Eucalyptus* using vegetation indexes. *Scientia Agricola*, 74:142-147. <http://dx.doi.org/10.1590/1678-992X-2015-0477>.
- Freedman D, Pisani R, Purves R (2007) Statistics (international student edition). Pisani, R. Purves, 4th Edn. WW Norton & Company, New York.
- Gabriel JL, Zarco-Tejada PJ, López-Herrera PJ, Pérez-Martín E, Alonso-Ayuso M, Quemada M (2017) Airborne and ground level sensors for monitoring nitrogen status in a maize crop. *Biosyst. Eng.*, 160:124–133.

- 439 Ge Y, Atefi A, Zhang H, Miao C, Ramamurthy RK, Sigmon B, Yang J, Schnable CJ (2019) High-throughput analysis
 440 of leaf physiological and chemical traits with VIS-NIR-SWIR spectroscopy: a case study with a maize
 441 diversity panel. *Plant Methods*, 15:66. <https://doi.org/10.1186/s13007-019-0450-8>.
- 442 Gomez R (2020) Professional short course on hyperspectral and multispectral imaging. *Appl Sci Technol*,
 443 <https://aticourses.com/>.
- 444 Gomeza C, Lagacherie P, Coulouma G (2008) Continuum removal versus PLSR method for clay and calcium
 445 carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*,
 446 148:141-148.
- 447 Horning N (2010) Random Forests: an algorithm for image classification and generation of continuous fields data
 448 sets. *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development*
 449 *in Earth and Allied Sciences, Osaka*. 911:1-6.
- 450 Hruška J, Adão T, Pádua L, Marques P, Cunha A, Peres E, Sousa A, Morais R, Sousa JJ (2018) Machine learning
 451 classification methods in hyperspectral data processing for agricultural applications. *J ACM*, 137–141.
 452 <https://doi.org/10.1145/3220228.3220242>.
- 453 Huang Z, Turnera BJ, Durya SJ, Wallis IR, Foley WJ (2004) Estimating foliage nitrogen concentration from HYMAP
 454 data using continuum removal analysis. *Remote Sens Environ*, 93:18-29.
- 455 Hunt GR (1980) Electromagnetic radiation: the communication link in remote sensing. In *Remote Sensing in Geology*;
 456 Siegal, B.S., Gillespie, A.R., Eds.; John Wiley & Sons, New York, NY, USA, pp 5-45.
- 457 Liang S (2004) *Quantitative remote sensing of land surfaces*, Wiley, Print ISBN: 9780471281665. 534 p. DOI:
 458 10.1002/047172372X.
- 459 Liu H, Zhu H, Wang P (2016) Quantitative modelling for leaf nitrogen content of winter wheat using UAV-based
 460 hyperspectral data, *Intl J of Remote Sens*, 38:2117–2134. DOI: 10.1080/01431161.2016.1253899.
- 461 Lu B, Dao P, Liu J, He Y, Shang J (2020) Recent advances of hyperspectral imaging technology and applications in
 462 agriculture. *Remote Sens*, 12:2659. DOI: 10.3390/rs12162659.

- 463 Maes WH, Steppe K (2019) Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture.
464 Trends Plant Sci, Vol. 24, <https://doi.org/10.1016/j.tplants.2018.11.007>.
- 465 Miller RO, Gavlak R, Horneck D (2013) Soil, plant and water reference methods for the Western region. 4th Ed.
466 WREP 125. 155 pp.
- 467 Morcillo-Pallarés P, Rivera-Caicedo JP, Belda S, De Grave C, Burriel H, Moreno J, Verrelst J (2019) Quantifying the
468 robustness of vegetation indices through global sensitivity analysis of homogeneous and forest leaf-canopy
469 radiative transfer models. Remote Sens, 11, 2418:1-23. DOI:10.3390/rs11202418.
- 470 Pandey P, Ge1 Y, Stoerger V, Schnable JC (2017) High throughput *in vivo* analysis of plant leaf chemical properties
471 using hyperspectral imaging. Front. Plant Sci, 8:1348. DOI: 10.3389/fpls.2017.01348.
- 472 Robila SA (2004) An analysis of spectral metrics for hyperspectral image processing, 2004 IEEE International
473 Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004. IEEE, 5:3233-
474 3236.
- 475 Severtson D, Callow N, Flower K, Neuhaus A, Olejnik M, Nansen C (2016) Unmanned aerial vehicle canopy
476 reflectance data detects potassium deficiency and green peach aphid susceptibility in canola. Precis Ag,
477 17:659–677.
- 478 Sowmya P, Giridhar MVSS (2017) Analysis of continuum removed hyperspectral reflectance data of *Capsicum*
479 *annuum* of ground truth data. ACST, 10:2233-2241.
- 480 Van Der Meij B, Kooistra L, Suomalainen J, Barel JM, De Deyn GB (2017) Remote sensing of plant trait responses
481 to field-based plant–soil feedback using UAV-based optical sensors. Biogeosciences, 14:733–749.
- 482 Wen D, Tongyu X, Fenghua Y, Chunling C (2018) Measurement of nitrogen content in rice by inversion of
483 hyperspectral reflectance data from an unmanned aerial vehicle. *Ciência Rural, Santa Maria*, 48(06)
484 e20180008. <http://dx.doi.org/10.1590/0103-8478cr20180>.
- 485 Xue J, Su B (2017) Significant remote sensing vegetation indices: A review of developments and applications. J Sens,
486 2017. <https://doi.org/10.1155/2017/1353691>.

- 487 Zhang J, Huang Y, Pu R, Gonzalez-Moreno P, Yuan L, Wu K, Huang W (2019) Monitoring plant diseases and pests
488 through remote sensing technology: A review. *Comput Electron Agric*, 165, 104943.
489 <https://doi.org/10.1016/j.compag.2019.104943>.
- 490 Zhao J, Karimzadeh M, Masjedi A, Wang T, Zhang X, Crawford MM, Ebert DS (2019) Feature explorer: Interactive
491 feature selection and exploration of regression models for hyperspectral images. 2019 IEEE Visualization
492 Conference (VIS). DOI: 10.1109/visual.2019.8933619.
- 493 Zhu H, Liu H, Xu Y, Guijun Y (2018) UAV-based hyperspectral analysis and spectral indices constructing for
494 quantitatively monitoring leaf nitrogen content of winter wheat. *Appl Opt*, 57:7722–7732.

Tables

Table 1 The highest correlated vegetation indices determined by using hyperspectral imaging in this study. Formula calculations were obtained from Anonymous, Index Database, 2021

Index Name	Abbreviation	Formula
Cellulose Absorption Index	CAI	$0.5 (2000nm + 2200nm) - 2100nm$
Normalized Difference Nitrogen Index	NDNI	$\frac{\log\left(\frac{1}{1754nm}\right) - \log\left(\frac{1}{1680nm}\right)}{\log\left(\frac{1}{1754nm}\right) + \log\left(\frac{1}{1680nm}\right)}$
Chlorophyll Absorption Ratio Index	CARI	$\left(\frac{700nm}{670nm}\right) \sqrt{(a * 670 + 670nm + b)^2} / \sqrt{(a^2 + 1)}$
Normalized difference vegetation index	NDVI	$\frac{(NIR - RED)}{(NIR + RED)}$
Leaf Area Index	LAI	$\frac{NIR}{RED}$
Double Peak Index	DPI	$\frac{(688nm + 710nm)}{697nm^2}$
Normalized Difference Water Index	NDWI	$(NIR - MNIR) / (NIR + MNIR)$
Normalized Difference Lignin Index	NDLI	$\frac{\log\left(\frac{1}{\rho_{1754}}\right) - \log\left(\frac{1}{\rho_{1600}}\right)}{\log\left(\frac{1}{\rho_{1754}}\right) + \log\left(\frac{1}{\rho_{1600}}\right)}$

507 **Table 2** Fifteen wavelength ranges taken from spectra extracted from little-leaf mockorange shoot cultures by using a
 508 spectroradiometer

Range number	From (Lower wavelength (nm))	To (Higher wavelength (nm))
1	364	369
2	378	559
3	559	772
4	838	843
5	898	903
6	928	1057
7	1121	1258
8	1287	1670
9	1670	1714
10	1714	1819
11	1819	2150
12	2253	2332
13	2341	2389
14	2389	2419
15	2428	2490

509

510

511 **Table 3** Various models developed for %N estimation in little-leaf mockorange shoots produced in tissue culture. Each model has
 512 different feature combinations and a different number of trees via the Random Forest algorithm

Features (N)	ntree = 5			ntree = 50			ntree = 100		
	R ²	RMSE	Correlation	R ²	RMSE	Correlation	R ²	RMSE	Correlation
2 features (DPI + Reflectance 525)	0.72	0.30	0.84	0.37	0.61	0.43	0.22	0.59	.
2 features (DPI + Asy11)	0.78	0.72	.	0.60	0.64	.	0.58	0.64	.
2 features (DPI + Asy3)	0.73	0.86	.	0.49	0.72	.	0.39	0.74	.

513

514

Table 4 Various models developed for %N estimation of little-leaf mockorange shoots produced in tissue culture. The models had different feature combinations and different penalty terms via SVM algorithm. RMSE, R^2 and correlation values resulted from the SVM model tuned by the linear kernel and three different values of penalty term

Features (N)	cost = 10			cost = 50			cost = 100		
	R^2	RMSE	Correlation	R^2	RMSE	Correlation	R^2	RMSE	Correlation
2 features (DPI+ Asy11)	0.58	0.32	0.76	0.58	0.32	0.76	0.58	0.32	0.76
2 features (DPI+ Asy3)	0.61	0.33	0.78	0.61	0.33	0.78	0.61	0.33	0.78
2 features (DPI+ Area4)	0.56	0.34	0.75	0.56	0.34	0.75	0.56	0.34	0.75

Table 5 Various models developed for %Ca estimation in little-leaf mockorange shoots produced in tissue culture. The models contain different features combinations and different number of trees via the Random Forest algorithm

Features	ntree = 5			ntree = 50			ntree = 100		
	R^2	RMSE	Correlation	R^2	RMSE	Correlation	R^2	RMSE	Correlation
2 features (Min4+Area15)	0.99	0.06	0.96	0.95	0.09	0.95	0.95	0.1	0.96
3 features (Min4+Area15+CAI)	0.98	0.04	0.93	0.89	0.09	0.93	0.87	0.09	0.98
4 features (Min4+Area15+Asy9+CAI)	0.99	0.03	0.99	0.92	0.08	0.96	0.85	0.1	0.92

526 **Table 6** Various models developed for %Ca estimation of little-leaf mockorange shoots produced with tissue culture. The models
 527 used different feature combinations and different penalty terms via the SVM algorithm

Features	cost = 10			cost = 50			cost = 100		
	R ²	RMSE	Correlation	R ²	RMSE	Correlation	R ²	RMSE	Correlation
2 features									
(Min4+CAI)	0.57	0.17	0.75	0.57	0.17	0.75	0.57	0.17	0.75
3 features									
(Min4+Area15+Asy9)	0.18	0.22	0.42	0.18	0.22	0.42	0.18	0.22	0.42
3 features									
(Min4+Area15+CAI)	0.51	0.18	0.71	0.51	0.18	0.71	0.51	0.18	0.71
4 features									
(Min4+Area15+Asy9+CAI)	0.59	0.16	0.76	0.58	0.16	0.76	0.58	0.16	0.76

528

Legends to Figres

Fig. 1 Spectral signature derived of certain wavelength range (X axis shows wavelength and Y axis shows reflectance values) and its continuum removal and geometric features extracted from continuum removal regions which include the area in left side (Area_L), the area in right side (Area_R) and depth (D)

Fig. 2 Schematic diagram of model development from hyperspectral bands and foliar mineral analysis for little-leaf mockorange shoots grown in tissue culture

Fig. 3 Correlation between leaf %N and the hyperspectral signatures acquired by a spectroradiometer from tissue cultured little-leaf mockorange shoots. The boxes show the wavelength of the peak in the spectrum and the correlation value

Fig. 4 Error bar plot (Left) and scatter plot (Right) of the correlation between observed and estimated %N of tissue-cultured little-leaf mockorange shoots by a linear regression model

Fig. 5 Correlation value between features and VIs with leaf N content of little-leaf mockorange shoots produced in tissue culture. The ovals show the features with highest correlation with the leaf N content

Fig. 6 Error bar plot and scatter plot of leaf %N estimated and measured test samples (DPI + Ref 525) for tissue cultured little-leaf mockorange shoots via Random Forest regression

Fig. 7 Error bar plot and scatter plot of leaf %N estimated and measured test samples (DPI + Asy11) for little-leaf mockorange shoots produced in tissue culture as determined by Support Vector Machines regression

Fig. 8 Correlation between leaf %Ca and the hyperspectral signatures acquired by the spectroradiometer from tissue cultured little-leaf mockorange shoots. The numbers in the rectangles represent wavelength (in nm) and correlation values, respectively

Fig. 9 Correlation value between features and VIs with leaf calcium content of little-leaf mockorange shoots produced in tissue culture. The abbreviations and numbers in the ovals represent the features and their correlation with the leaf Ca content

Fig. 10 Error bar plot and scatter plot of the correlation between observed and estimated tests (Min 11+ Area 3) for %Ca of little-leaf mockorange shoots produced in tissue culture as determined by a linear regression model

Fig. 11 The importance value of generated features and selected VIs regarding leaf %Ca in little-leaf mockorange shoots produced in tissue culture as determined via Random Forest algorithm. The abbreviations and numbers in the rectangles represent the features and their correlation with the leaf Ca content via RF regression model

- 554 **Fig. 12** Error bar plot (left) and scatter plot (right) of leaf %Ca estimated and measured test samples (Min4+Area15+Asy9+CAI)
555 in little-leaf mockorange shoots produced in tissue culture determined via the Random Forest algorithm
- 556 **Fig. 13** Error bar plot and scatter plot of leaf %Ca estimated and measured test samples (Min4+Area15+Asy9+CAI) for little-leaf
557 mockorange shoots produced in tissue culture as determined via Support Vector Machines regression

Figures

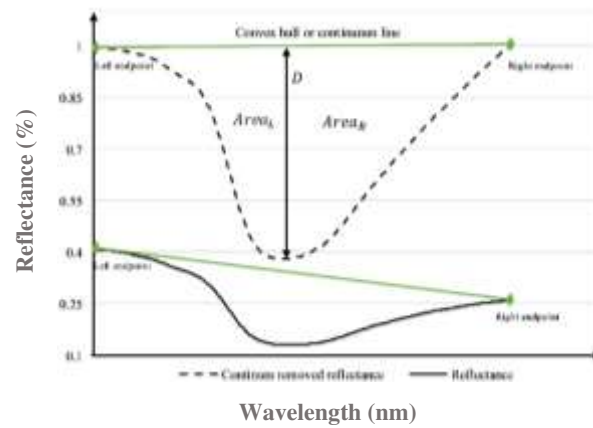


Figure 1

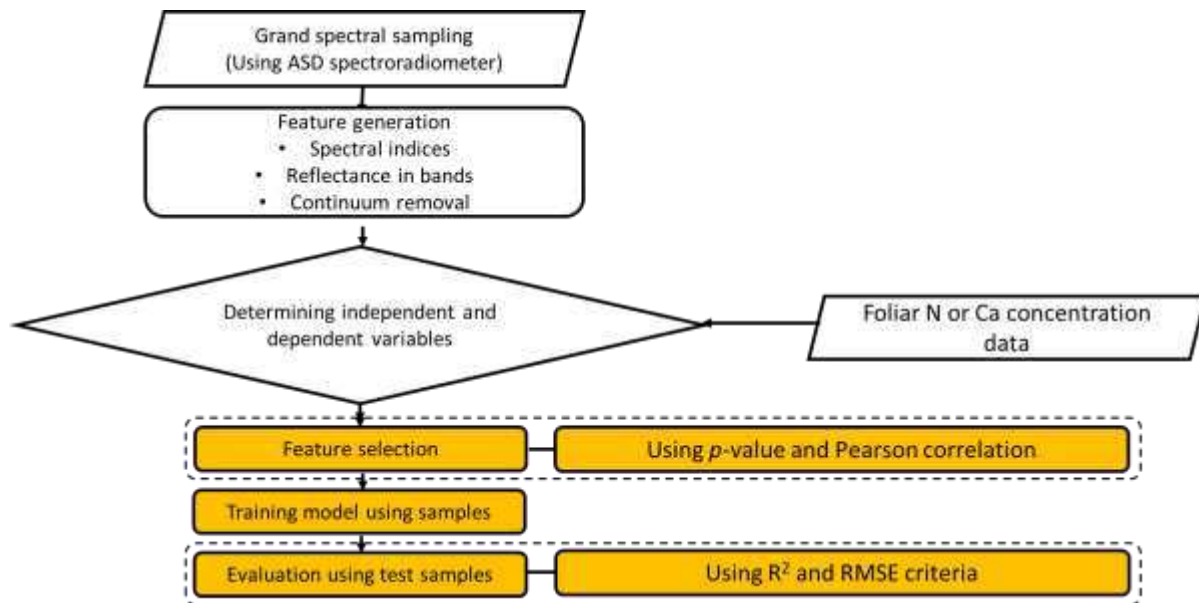


Figure 2

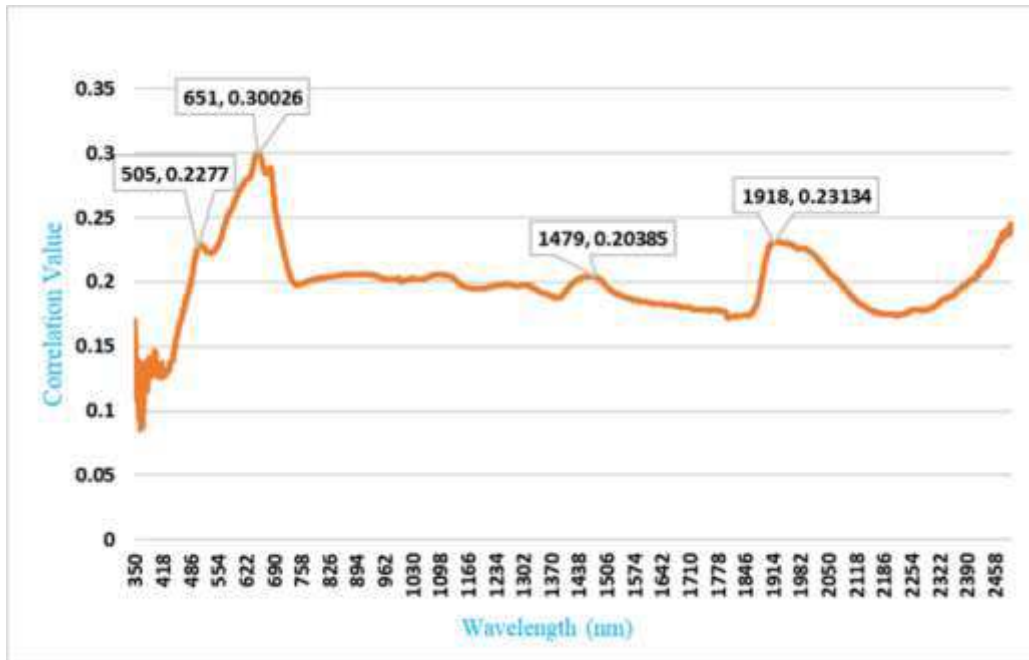


Figure 3

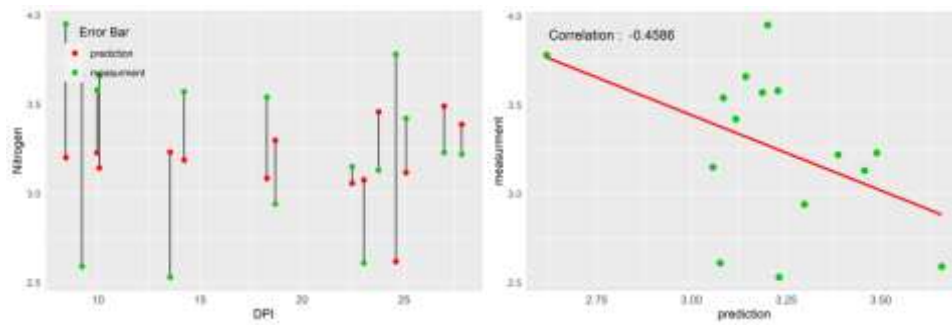


Figure 4

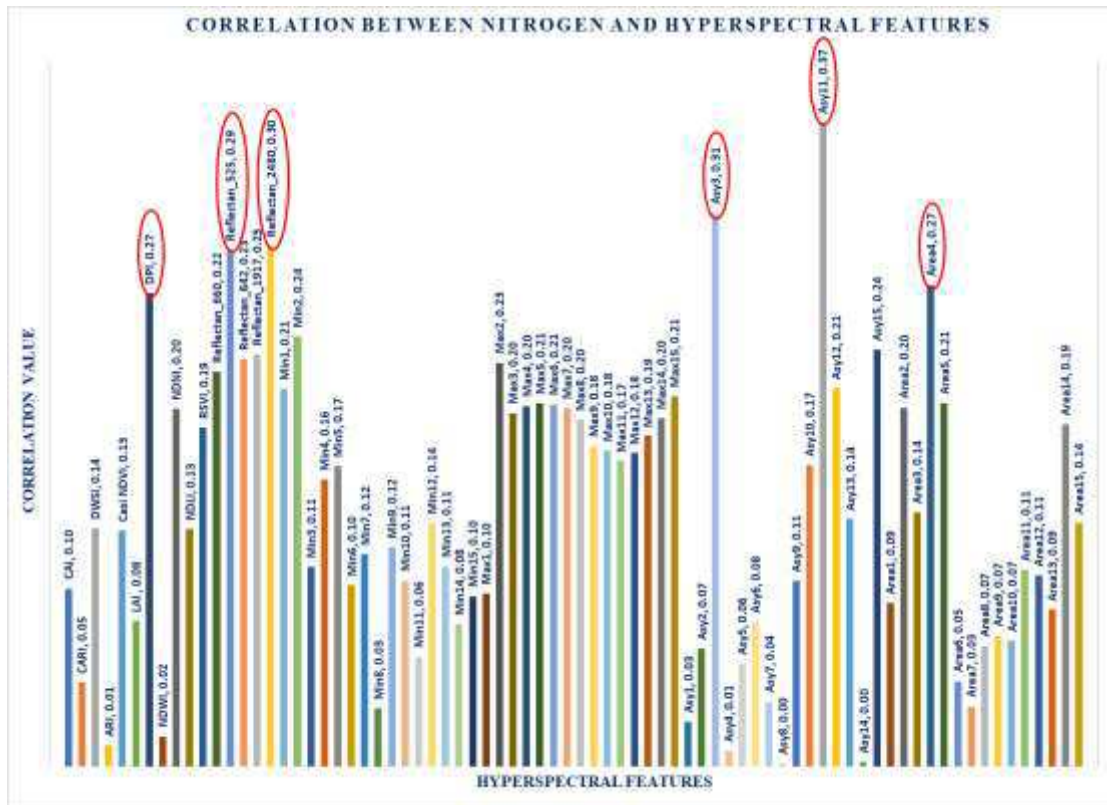


Figure 5

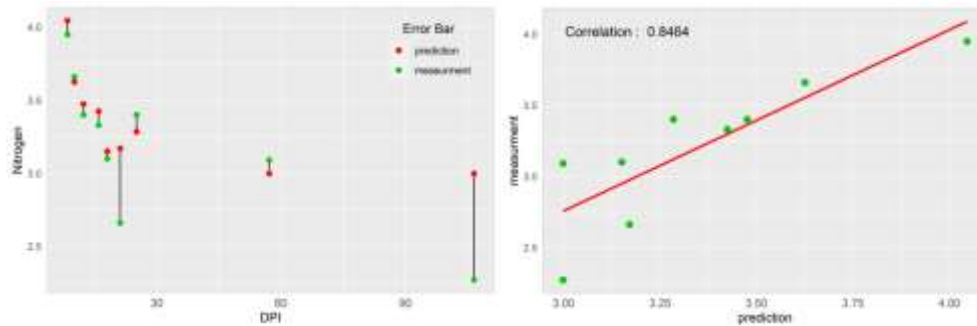


Figure 6

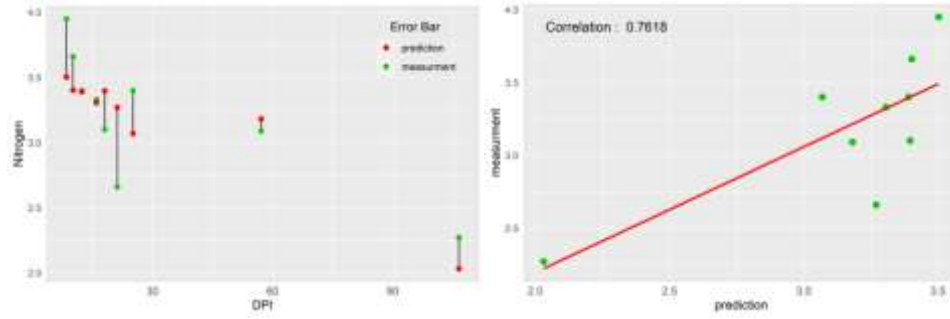


Figure 7

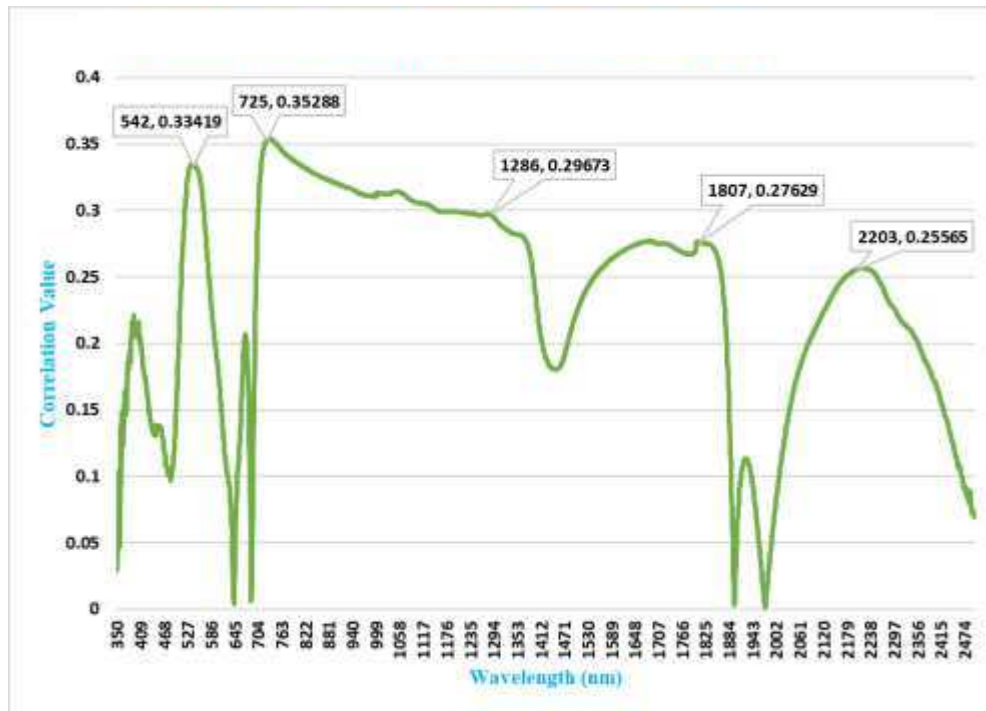


Figure 8

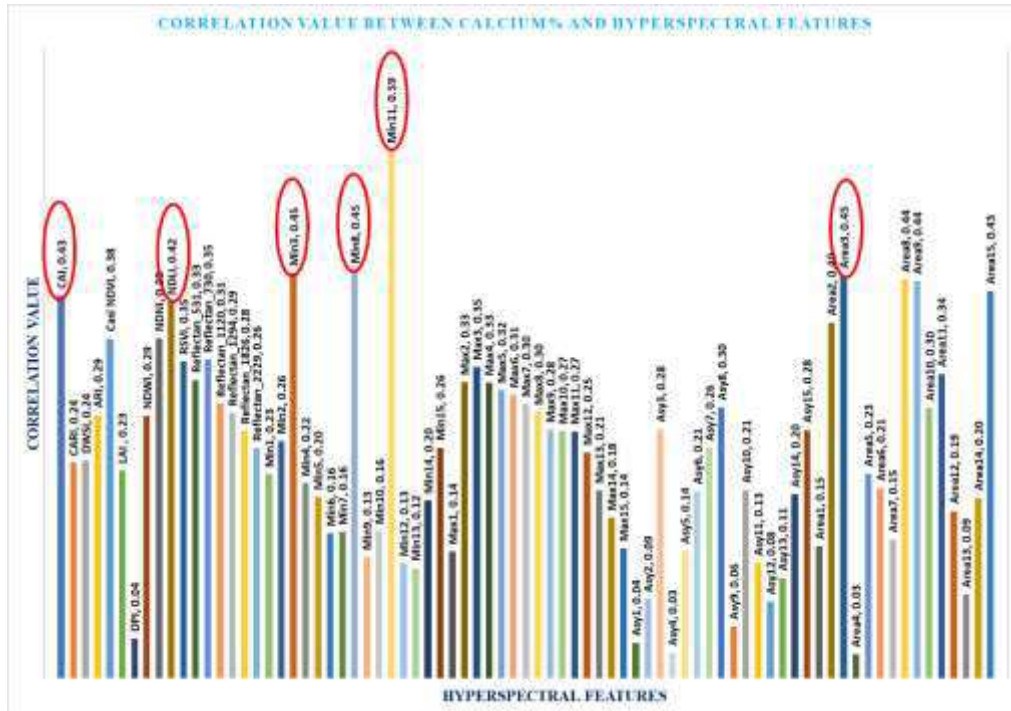


Figure 9

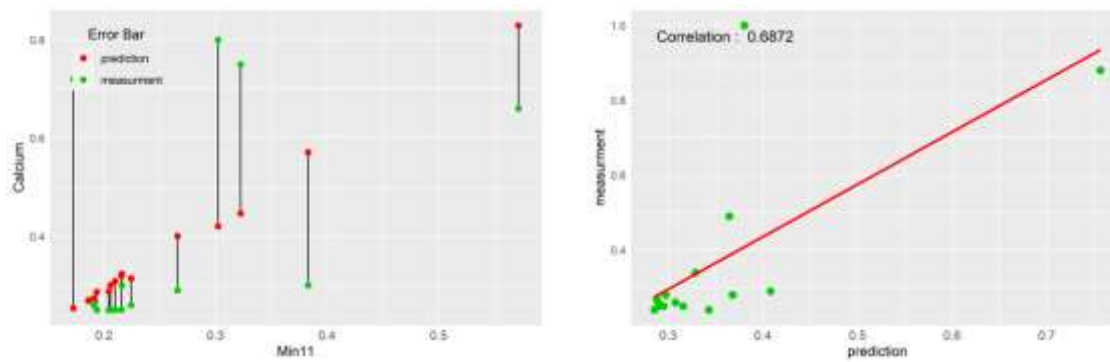


Figure 10

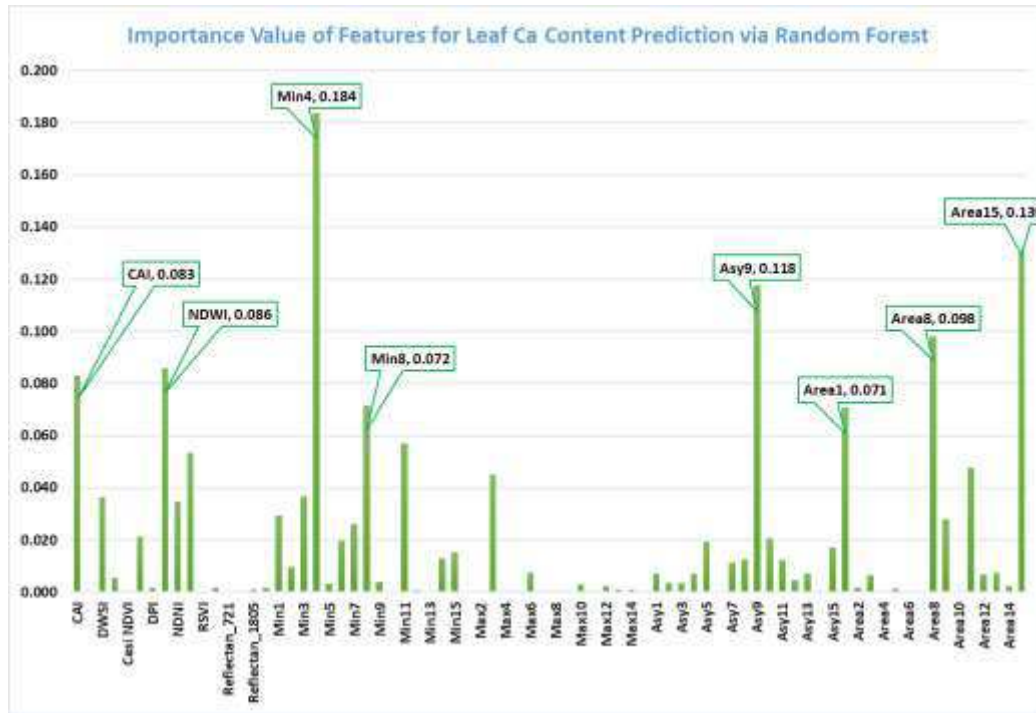


Figure 11

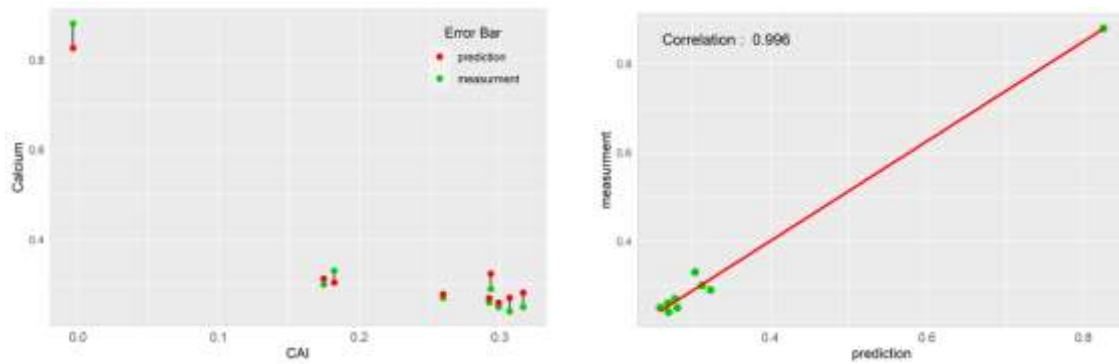


Figure 12

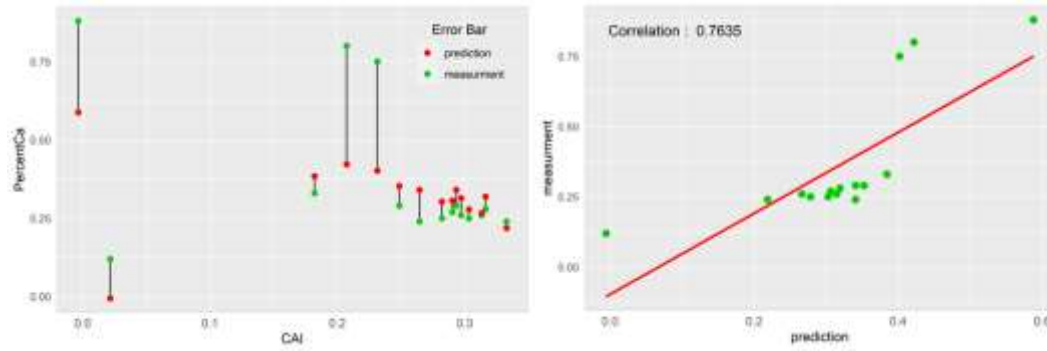


Figure 13