

Article

Classification and Prediction by Pigment Content in Lettuce (*Lactuca sativa* L.) Varieties Using Machine Learning and ATR-FTIR Spectroscopy

Renan Falcioni ^{1,*} , Thaise Moriwaki ¹, Mariana Sversut Gibin ² , Alessandra Vollmann ¹, Mariana Carmona Pattaro ¹, Marina Ellen Giacomelli ¹, Francielle Sato ², Marcos Rafael Nanni ¹  and Werner Camargos Antunes ¹ 

¹ Plant Ecophysiology Laboratory, Graduate Program in Agronomy, Department of Agronomy, State University of Maringá, Av. Colombo, 5790, Maringá 87020-900, Brazil

² Optical Spectroscopy and Thermophysical Properties Research Group, Graduate Program in Physics, Department of Physics, State University of Maringá, Av. Colombo, 5790, Maringá 87020-900, Brazil

* Correspondence: renanfalcioni@gmail.com; Tel.: +55-44-3011-8940

Abstract: Green or purple lettuce varieties produce many secondary metabolites, such as chlorophylls, carotenoids, anthocyanins, flavonoids, and phenolic compounds, which is an emergent search in the field of biomolecule research. The main objective of this study was to use multivariate and machine learning algorithms on Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR)-based spectra to classify, predict, and categorize chemometric attributes. The cluster heatmap showed the highest efficiency in grouping similar lettuce varieties based on pigment profiles. The relationship among pigments was more significant than the absolute contents. Other results allow classification based on ATR-FTIR fingerprints of inflections associated with structural and chemical components present in lettuce, obtaining high accuracy and precision (>97%) by using principal component analysis and discriminant analysis (PCA-LDA)-associated linear LDA and SVM machine learning algorithms. In addition, PLSR models were capable of predicting Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe contents, with R^2_P and RPD_P values considered very good (0.81–0.88) for Car, Anc, and Flv and excellent (0.91–0.93) for Phe. According to the RPD_P metric, the models were considered excellent (>2.10) for all variables estimated. Thus, this research shows the potential of machine learning solutions for ATR-FTIR spectroscopy analysis to classify, estimate, and characterize the biomolecules associated with secondary metabolites in lettuce.

Keywords: anthocyanins; carotenoids and chlorophylls; discriminant analysis; flavonoids; phenolic compounds; pigment analysis; PLSR analysis; phytochemistry; secondary metabolism



Citation: Falcioni, R.; Moriwaki, T.; Gibin, M.S.; Vollmann, A.; Pattaro, M.C.; Giacomelli, M.E.; Sato, F.; Nanni, M.R.; Antunes, W.C. Classification and Prediction by Pigment Content in Lettuce (*Lactuca sativa* L.) Varieties Using Machine Learning and ATR-FTIR Spectroscopy. *Plants* **2022**, *11*, 3413. <https://doi.org/10.3390/plants11243413>

Academic Editor: Ain Raal

Received: 21 October 2022

Accepted: 6 December 2022

Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plants produce many secondary metabolites as a function of species, genotype, and environmental and physiological conditions [1–4]. These secondary compounds may be responsible for important functions in plant biology [5–7]. The classification and prediction of secondary metabolite contents that are rich in biological activities is of great interest to researchers [2,3,8]. Thus, to elucidate the variation in many secondary compounds or metabolites based on secondary metabolism, high-throughput and large-scale data analyses have been used in recent years [9–12]. In this sense, one of the main methods of characterization and analysis with satisfactory precision and accuracy has been used by the combination of different FTIR-spectroscopic techniques and machine learning tools [11,13–17].

Many molecules and biocompounds, such as secondary metabolites, including flavonoids, alkaloids, phenylpropanoids, terpenes, carotenoids, amino acids, and other phenolic compounds, have been well explored because some of these compounds possess alimentary or pharmaceutical properties that are attractive to human health or affect plant

metabolism [2,18,19]. Flavonoids/phenolics represent one of the most varied groups of secondary metabolites in the plant kingdom [3,20,21] and include 6000 known chemical structures [18,22,23]. Phenolic compounds, flavonoids, anthocyanins, carotenoids, and chlorophylls have high economic and production values [2,7,24–26]. The structural diversity of these compounds contributes to their numerous physiological activities on plants, such as their advantage to plant adaptation to distinct environmental conditions, such as cold or heat, many biotic interactions and pathogen attacks, and light (high intensity) and UV stresses [18,23,27,28]. In addition, their unique vibrational bands can be spectroscopically detected. In this sense, there is much interest in the analysis of leaves, fruits, seeds, stems, tubers, roots, and flowers [18,29,30]. Furthermore, green, green-purplish, or full purple lettuce varieties represent one of the most economically important and popular vegetables consumed, for which global production is estimated to be 27 million tons worldwide (FAO, 2022).

ATR-FTIR equipment is a sensor based on spectroscopy measurements that can be used to characterize the properties of molecules. It can be utilized for crops and food [3,19,29–32]. The prediction of biomolecules from spectroscopy tools is based on the vibrational relationships of infrared with molecular bonds and, mainly, its vibrational modes, as well as C=O stretching, C–H bonds, aromatic C=C, –COOH, –NH₃, –C–H, and –O–H. The output is a spectrum of functional group vibrations at specific wavenumbers [3,4,7,22]. Thus, based on horticultural crops and postharvest, chemometric parameter analysis of many vibrational groups combined with machine learning algorithms makes it possible to define better strategies for vertical farms, industries, and consumers. For example, consider the digital revolution 4.0, in which an economy of US \$750 billion a year is estimated. Thus, machine learning in decision making and analysis by ATR-FTIR and other spectroscopy tools could be a promising approach by which to classify and predict biomolecules. In this sense, it could mitigate problems with prediction and training and avoid economic losses [31,32].

Many studies have explored multivariate statistical methods to directly model frequently used plant chemometrics as a function of machine learning algorithms [11,13,33–35]. Developments in the portability, accuracy, and sensitivity of ATR-FTIR coupled computational algorithms and multivariate tools for modelling and other statistical-machine learning methods [3,4,28] have allowed advances in classifying and predicting the variety of plants. Thus, this approach is used to evaluate a variety of commonly investigated plant secondary metabolite chemometrics processes based on vibrational groups and specific band properties [26,34,36–39]. In this sense, a calibration model is developed by combining leaf spectra collected on a uniform and calibrated spectral source in a consistent manner with independent and reference methods for measuring samples. Subsequently, model-based ATR-FTIR and machine learning algorithms were developed and validated by comparing relationships between observed and predicted values collected from other independent databases, samples, or environments of field experiments [4,7,11,36]. Furthermore, an adjustment model was then used to predict the variable of interest in uncharted samples based on their many spectral spectroscopy signatures. Currently, ATR-FTIR spectroscopy analysis uses the full or ranged (specific) spectrum for the analysis of many chemometric varieties in plants [7,22,33,36,40] to quantify a particular bond. From this perspective, infrared spectroscopy has been highly successful in both inorganic and organic chemistry, and one application is shown here to monitor biomolecules presenting different and principal pathways of secondary metabolism in lettuce (Figure 1).

Considering the needs described above and the importance of lettuce plants, the main objective was to assess the capacity to classify and predict seven functional molecules, such as chlorophylls, carotenoids, anthocyanins, flavonoids, and phenolic compounds, in eleven lettuce varieties based on a machine learning algorithm (PCA, LDA, and SVM) and ATR-FTIR spectroscopy (4000–400 cm^{−1}) to characterize and distinguish plant secondary metabolites. A full spectrum of lettuce variety plants was obtained (4000–400 cm^{−1}). In some cases, the range (1500–1150 cm^{−1}) was used as a rapid dataset to estimate seven principal chemometric attributes to classify and predict chlorophyll *a* (Chl*a*), chlorophyll *b*

(Chl b), total chlorophylls (Chl $a+b$), carotenoids (Car; total carotenes and xanthophylls), anthocyanins (AnC), flavonoids (Flv), and phenolic compounds (Phe). If successful, this approach could be directly used to classify important biomolecule-based chemometric and fingerprint parameters in lettuce varieties with higher accuracy and precision to predict model-based machine learning (Figures 1 and 2).

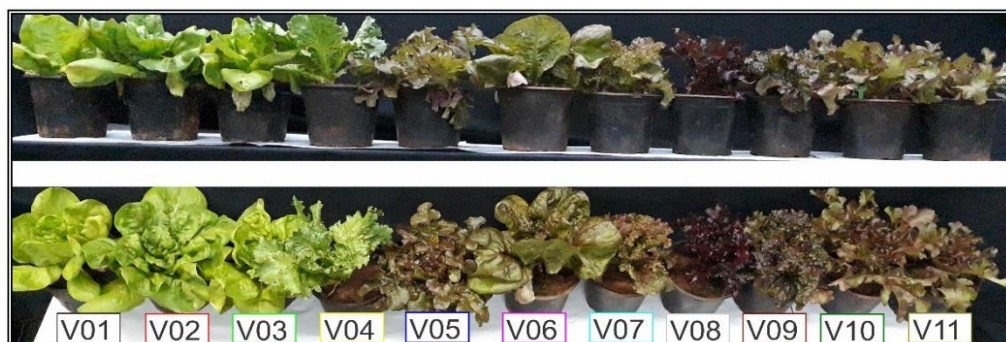


Figure 1. Images of lettuce (*Lactuca sativa* L.) plants evaluated in this study, displaying significant bioaccumulation of pigment variations (accumulation of the pigments based secondary metabolism). Upper panel, details of the lateral profile. Lower, detail of the plant leaves. The box displays eleven different genotypes of lettuce: V01—Rainha de Maio, V02—Vitória, V03—Maravilha de Inverno, V04—Grandes Lagos Americana, V05—Mimosa Prado, V06—Quatro Estações, V07—Batávia Joaquina, V08—Mimosa Vermelha, V09—Batávia Cacimba, V10—Pipa and V11—Mimosa Rubi.

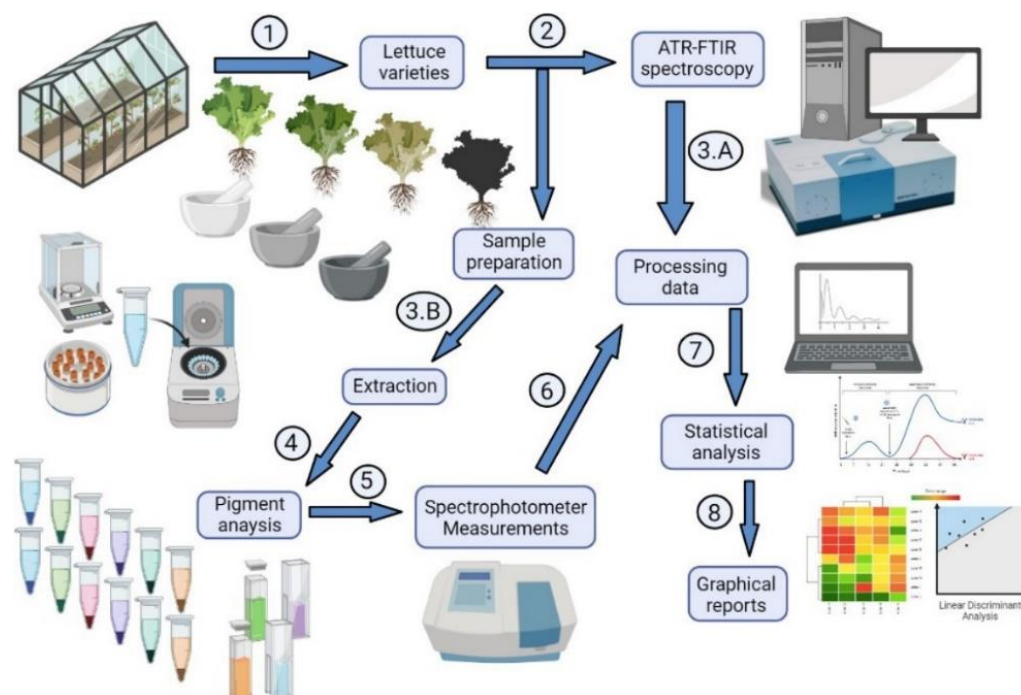


Figure 2. Flowchart describing classified and predicted pigments in *Lactuca sativa* L. (lettuce varieties) using machine learning and ATR-FTIR spectroscopy for analysis. (1) Greenhouse to grow the lettuce; (2) sample preparation and analysis; (3. A) ATR-FTIR spectroscopy direct analysis; (3. B) material for chemical quantification; (4) pigment extractions in different solvents; (5) analysis in a microreader or spectrophotometer; (6) data processing and multivariate statistical analysis based on machine learning; (7) modelling, calibration, classification, and prediction models; and (8) training, testing, and graphical reports by analysis.

2. Results

2.1. Descriptive Analysis

The descriptive analysis of the seven classes of compounds (chlorophylls (*a*, *b*, and *a + b*), total carotenoids, anthocyanins, flavonoids, and phenolics) in different lettuce varieties is displayed in Table 1. The CVs (%) ranged from 16.79 to 126.53% (Table 1); 01 of the 07 parameters analyzed demonstrated CVs (%) labelled as medium to very high (Phe) and six very high parameters (Chla, Chlb, Chla+b, Car, AnC, and Flv) (see Abbreviation list), allowing, as broad as possible, to encompass natural variability, mainly for AnC contents.

Table 1. Descriptive analysis in lettuce plants. Parameters: Chlorophyll *a* (Chla; mg m⁻²); Chlorophyll *b* (Chlb; mg m⁻²); Chlorophyll *a + b* (Chla+b; mg m⁻²); Carotenoids (Car; mg m⁻²); Anthocyanins (AnC; nmol m⁻²); Flavonoids (Flv; nmol m⁻²); Phenolic compounds (Phe; mL L⁻¹).

Parameters	Count (<i>n</i>)	Mean	Median	Minimum	Maximum	CV (%)
Chla (mg m ⁻²)	132	128.89	118.95	63.15	306.40	39.18
Chlb (mg m ⁻²)	132	66.76	61.75	39.41	147.27	35.09
Chla+b (mg m ⁻²)	132	195.65	182.51	107.96	453.67	37.64
Car (mg m ⁻²)	132	45.70	42.41	25.56	101.92	35.47
AnC (nmol m ⁻²)	132	13.43	11.46	0.00	68.57	126.53
Flv (nmol m ⁻²)	132	59.55	49.75	11.77	178.28	60.66
Phe (mL L ⁻¹)	132	62.59	66.20	42.74	78.66	16.79

2.2. Cluster Heatmap Analysis

The cluster heatmap grouping lettuce varieties based on the chemometric parameters evaluated is shown in Figure 3. The heatmap separates the green, green-purplish, and purple varieties of lettuce, evidencing a rise from the green pigments (V01, V02, V03, and V04), intermediate contents (V05, V07, V10, and V11), and the highest purple pigments (V08 and V09), distinguishing the highest chlorophyll and anthocyanin accumulation (Figure 3), except for V06, which is naturally “centered” among distinct chemometrics data (Figures 1 and 3). The lowest contents of Phe, Flv, and AnC were observed in green lettuce, and the highest were observed in purple varieties. On the other hand, clusters formed with varieties that accumulated the lowest levels of Flv, AnC, and Phe (blue squares) were related to greater accumulation of Chla, Chlb, Chla+b, and Car (red squares) (Figure 3).

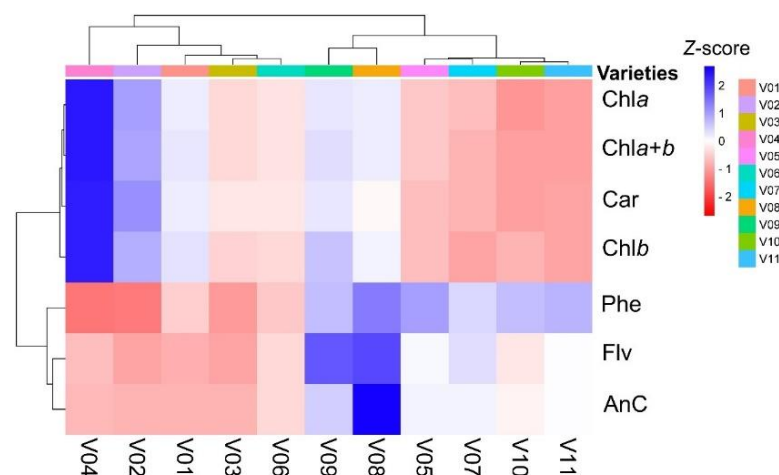


Figure 3. Cluster heatmap summarizing pigment contents from eleven varieties of *Lactuca sativa* L. (lettuces). Numerical differences within the data matrix are shown by the Z score of the color by scale red to blue indicate decreased to increased values, respectively. Parameters are clustered in the rows; sample groups are clustered in the columns by pigment content in leaves at independent factors. See varieties in Figure 1 or Abbreviation list. (*n* = 132 samples).

2.3. ATR-FTIR Spectroscopy Analysis

ATR-FTIR spectroscopy data for the eleven lettuce varieties (132 samples) are displayed in Figure 4A. PERMANOVA was shown to discriminate significant wavenumbers ($F: 6.73$; $p < 0.001$) from the spectra. Most functional groups (vibrational groups) detected in the full range (4000 to 400 cm^{-1}) showed differences between varieties in ATR-FTIR signals. A slight and significant variation in ATR-FTIR signal factor intensity was observed in a Gaussian fit based on leaf pigments, such as carotenoid and chlorophyll, and structural differences in the organic molecules of leaf mesophyll (Figure 4A).

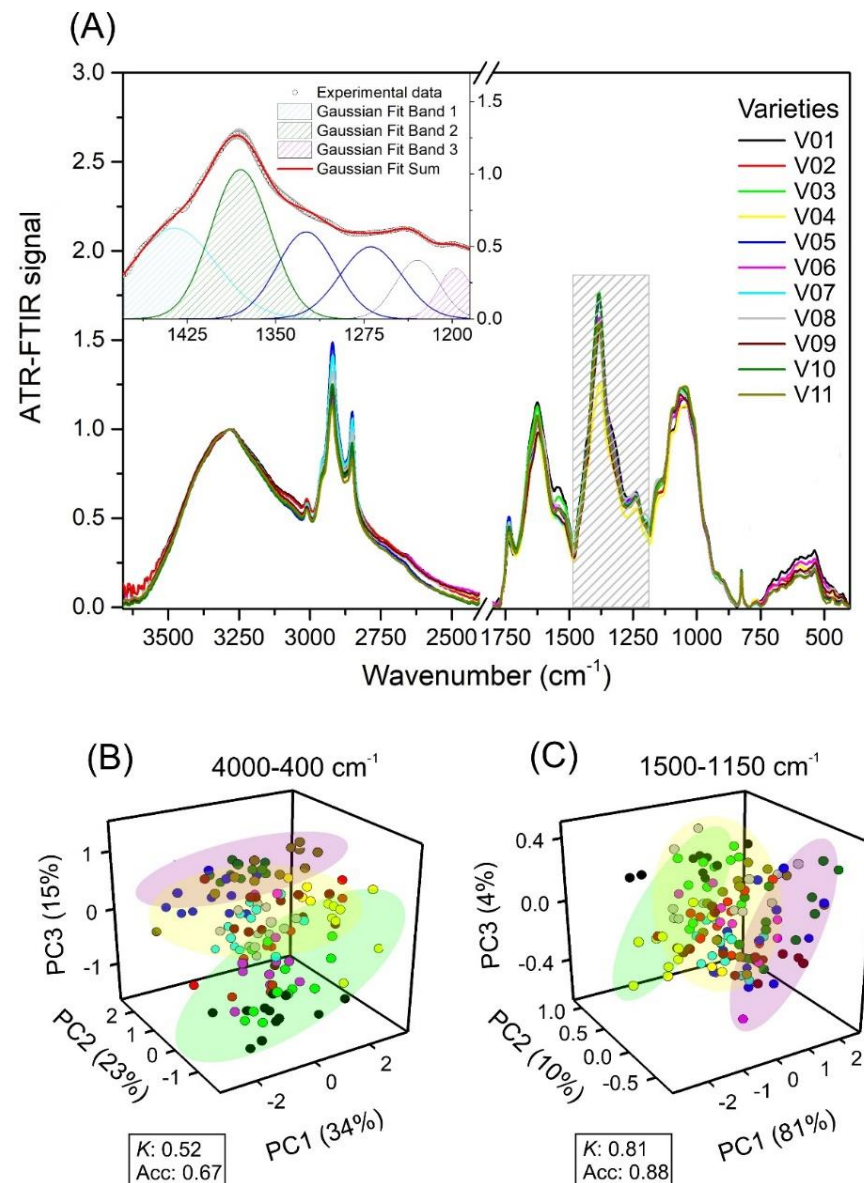


Figure 4. (A) Average ATR-FTIR signal profile of eleven varieties of lettuce over the full range (4000 – 400 cm^{-1}) of ATR-FTIR spectroscopy and ranged from 1500 to 1150 cm^{-1} . The inset displays the experimental data and Gaussian fit curves (1436 , 1380 , and 1197 cm^{-1}). (B) 3D plot of the principal component analysis score for the first three principal components (PC1, PC2, and PC3) of the spectroscopy data (4000 – 400 cm^{-1}). (C) 3D plot of the principal component analysis score for the first three principal components (PC1, PC2, and PC3) of the spectroscopy data (1500 – 1150 cm^{-1}). Data collected from lettuce plants. Green, yellow, and purple colors display correlation clusters of lettuce varieties that showed similar groups of pigments in leaves. K : $Kappa$ and Acc : accuracy. The ATR-FTIR signal intensity was normalized at 3270 cm^{-1} (OH stretching band).

The selected spectral range (1500 to 1150 cm^{-1}) is highlighted in the inset of Figure 4A. Gaussian and functional adjustment were accomplished to explore the main differences in vibrational modes of the secondary metabolites in these regions. Three main bands (1436, 1380, and 1197 cm^{-1} ; inset) showed significant differences. For example, there was a minor correlation at 1436 and 1380 cm^{-1} for flavonoids, anthocyanins, or phenolic compounds, and a higher correlation ($r = 0.960$; $p < 0.001$) at 1197 cm^{-1} with anthocyanins (Figure 4A; inset). The variability explained based on the first three PCA scores from 4000–400 cm^{-1} (PC1: 34%, PC2: 23%, and PC3: 15%; Figure 4B) was smaller than the ranged data of 1500 to 1150 cm^{-1} (PC1: 81%, PC2: 10%, and PC3: 4%; Figure 4C). *Kappa* (*K*) identified by PCA-PLSR was observed by calibration:cross-validation with a 60:40 ratio for calibration:validation data using the 7 components, which reported a *K* of 0.52 and 0.81 and accuracy (Acc) of 0.67 and 0.88, respectively (Figure 4B,C).

2.4. Correlation with Principal Component Analysis (PCA)

The correlation between wavenumbers and PCs can be observed in Figure 5. The regions close to the 3600–3200 cm^{-1} , 2800–2600 cm^{-1} , 1800–1600 cm^{-1} , 1500–1150 cm^{-1} , 1100–800 cm^{-1} , and 700–400 cm^{-1} vibrational bands presented the highest correlation values by functional pigment analyses. In this sense, for the first three principal components (PC1, PC2, and PC3), these regions may present properties and significant bands of differentiation among the varieties (Figure 5A). Consequently, the spectral bands of these regions can be selected to compose the linear discriminant analysis (LDA) and support vector machine classification (SVM)-based machine learning algorithms. In addition, the selected 1500–1150 cm^{-1} range showed a strong correlation for the discrimination of lettuce varieties based on vibration spectroscopy (Figure 5B).

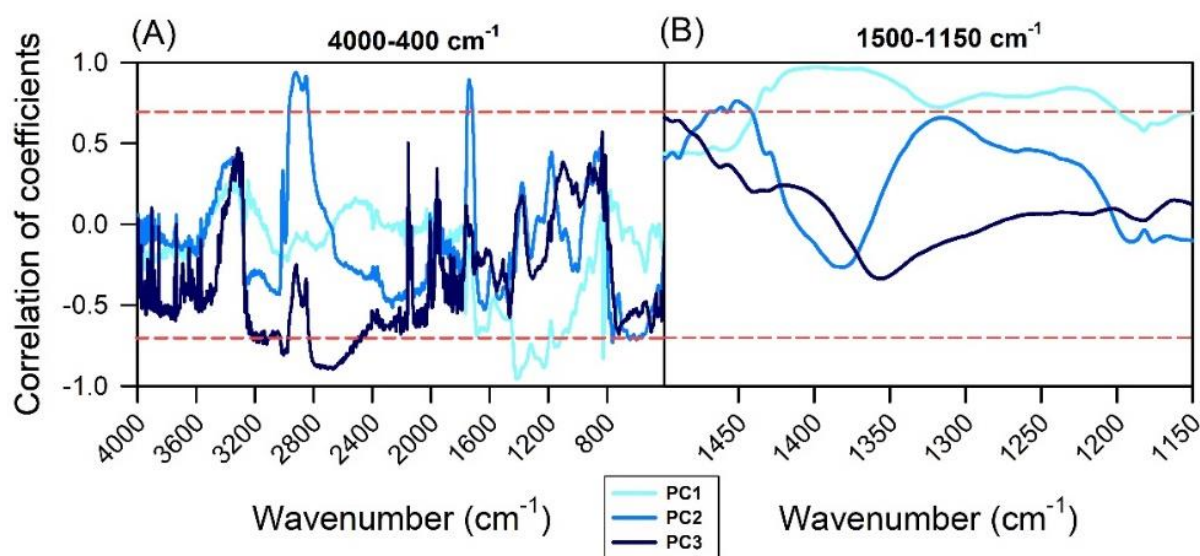


Figure 5. Correlation coefficients with the three principal components (light to dark blue lines, PC1, PC1, and PC3). (A) 4000–400 cm^{-1} by ATR-FTIR measurements. (B) 1500–1150 cm^{-1} by ATR-FTIR measurements. The red line indicates -0.70 and 0.70 correlation of coefficients.

2.5. Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) Based on Machine Learning Algorithms

Machine learning algorithms were useful for sample classification by using ATR-FTIR spectroscopy raw data based on LDA and SVM. The overall accuracy was slightly enhanced using LDA-linear in relation to quadratic or Mahalanobis models or first derivative models using transformed data (Figure 6A,B). The first derivative spectra data did not show significant accuracy (highest error values or lower accuracy) (Figure 6A,B). In this sense, an overall accuracy of 98.6% in the 4000–400 cm^{-1} spectra and 79.5% in the 1500–1150 cm^{-1} range was obtained by the LDA-PLS method by using the first three PCs, carrying 95%

data variance. The confusion matrix showed >97.9% accuracy to LDA-linear, contributing to lower misclassification between training and test data (i.e., 80 training samples and 52 test samples). Furthermore, 1500–1150 cm^{-1} ranged data showed higher accuracy in correctly classifying lettuce varieties, while full-spectrum machine learning algorithms did not classify lettuce varieties with similar precision. However, linear SVM showed 99.4% and 94.2% to quadratic-SVM in the training data, but having the lowest accuracy based on validation, 40.1% and 3.6%, respectively, of the data showed minor accuracy in classifying the lettuce varieties (Figure 6A,B; confusion matrix on the right).

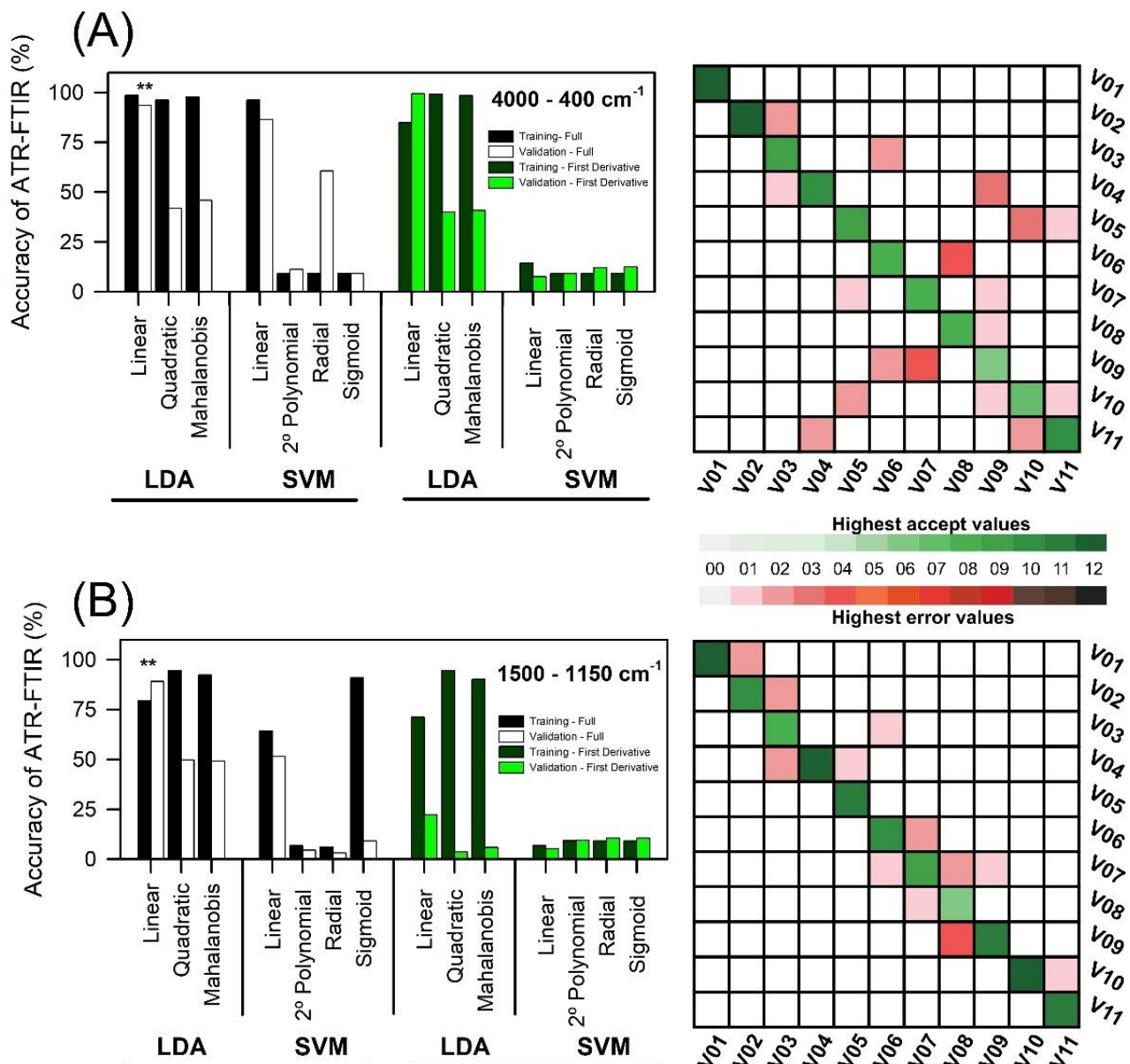


Figure 6. Accuracy of different machine learning algorithms based on linear discriminant analysis (LDA) and support vector machine classification (SVM) analyzed in full and transformed data (first derivative) on lettuce varieties ATR-FTIR spectroscopy data. (A) 4000 at 400 cm^{-1} . (B) 1500 at 1150 cm^{-1} . The first three PCA datasets were selected from the training and validation (testing) to create a confusion matrix for the machine learning algorithms, which showed higher overall accuracy. ** indicates the matrix confusion report on the right. The right boxes indicate the highest acceptance accuracy-precision, and the red boxes indicate the highest error accuracy-precision in the confusion matrix.

Validation by LDA-PLS was carried out with the independent dataset, and the percentage of discrimination varied between 49.8 and 89.7%. Classification success was observed for all varieties based on machine learning algorithms. When range data (1500–1150 cm^{-1}) were used, higher accept values with high accuracy and precision in the confusion matrix were found, although the full spectrum presented greater dispersion around the selection in the confusion matrix (Figure 6A,B; confusion matrix on the right).

The SVM analysis showed lower values (highest error values >37.69%) in relation to the LDA algorithms, even when using first derivative transformed data.

2.6. Prediction of Chemometric Parameters

PLSR models based on the calibration (Cal) and cross-validation (Cva) methods for biochemical and biomolecules related by pigment content parameters based on a vibrational band are shown in Table 2. Regarding ATR-FTIR spectroscopy data collected from 4000–400 cm^{-1} or 1500–1150 cm^{-1} , a distinct difference in behavior was observed between the statistical stages (Cal and Cva). The best results (bold in Table 2) of R^2_{CV} and RPD_{CV} , Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe to 4000–400 cm^{-1} and AnC, Flv, and Phe displayed very good and excellent prediction scores, respectively. Major values for R^2_{CV} and RPD_{CV} metrics were obtained using full spectroscopy data for Phe, AnC, and Flv. In this sense, bias showed similar results (close to zero) for the calibration and cross-validation evaluated.

Table 2. PLSR model in calibration and cross-validation at wavenumbers of 4000 at 400 cm^{-1} and 1500 at 1150 cm^{-1} . Model goodness-of-fit (R^2), offset, root mean square error (RMSE), and ratio of performance to deviation (RPD) for calibration (Cal) and cross-validation (Cva) data generated using 132 random permutations of the data with 60% used for Cal and 40% used for Val for the PLSR models predicting parameters biochemical pigments (chemometrics) contents from ATR-FTIR spectroscopy of lettuce leaves. Bias outputs are not shown, as they were always lower than 0.01 for both Cal and Cva. The bold represents significant regression models (R^2). The underline indicates a bad residual to the prediction deviation (RPD). Parameter abbreviations—see Table 1 or Abbreviation list.

Wavenumbers	Parameters	Calibration				Cross-Validation			
		R^2	Offset	RMSE	RPD	R^2	Offset	RMSE	RPD
4000–400 cm^{-1}	Chla (mg m^{-2})	0.77	31.70	24.91	2.09	0.76	33.63	26.60	2.04
	Chlb (mg m^{-2})	0.78	14.35	10.93	2.13	0.75	15.20	11.64	2.00
	Chla+b (mg m^{-2})	0.79	41.80	33.73	2.18	0.76	44.58	36.03	2.04
	Car (mg m^{-2})	0.81	8.55	7.05	2.29	0.78	9.09	7.52	2.13
	AnC (nmol m^{-2})	0.86	1.78	6.17	2.67	0.85	1.87	6.56	2.58
	Flv (nmol m^{-2})	0.80	11.75	15.60	2.24	0.76	12.60	16.85	2.04
	Phe (mL L^{-1})	0.91	5.57	3.15	3.33	0.90	5.94	3.38	3.16
1500–1150 cm^{-1}	Chla (mg m^{-2})	0.49	65.64	35.90	1.40	0.42	69.63	38.17	<u>1.31</u>
	Chlb (mg m^{-2})	0.53	30.97	14.45	1.46	0.49	32.99	15.47	1.40
	Chla+b (mg m^{-2})	0.48	99.95	50.15	<u>1.39</u>	0.45	106.06	53.45	<u>1.35</u>
	Car (mg m^{-2})	0.54	21.04	10.95	1.47	0.51	22.30	11.64	1.43
	AnC (nmol m^{-2})	0.91	1.25	4.91	3.33	0.89	1.35	5.26	3.02
	Flv (nmol m^{-2})	0.85	8.38	13.10	2.58	0.84	8.98	13.98	2.50
	Phe (mL L^{-1})	0.81	11.58	4.37	2.29	0.80	12.20	4.64	2.24

The prediction models for a pigment of the secondary metabolism present in lettuce parameters (i.e., biomolecules analyzed) were adjusted following the number of PLSR factors tested earlier by cross-validation (maximum of interaction but minor output noise). In this sense, the relationship between the predictor (wavenumbers) and predicted (lettuce chemometric parameters) indicated which variables were better explained using models containing the maximum of the seven factors for Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe (Table 3).

Table 3. PLSR model in the predicted phase. Model goodness-of-fit (R^2), offset, root mean square error (RMSEP), standard error of prediction (SEP), bias and ratio of performance to deviation (RPD) parameters from ATR-FTIR spectroscopy data of lettuce leaves. The bold represents statistically significant regression models (R^2). The underline indicates a bad Bias calculated. Parameter abbreviations—see Table 1 or Abbreviation list.

Wavenumbers	Parameters	Predicted					
		R^2	Offset	RMSEP	SEP	Bias	RPD
4000–400 cm^{-1}	Chla (mg m^{-2})	0.71	43.20	30.45	30.09	5.87	1.86
	Chlb (mg m^{-2})	0.88	11.64	9.30	9.27	<u>−1.49</u>	2.89
	Chla+b (mg m^{-2})	0.87	39.58	29.80	29.25	<u>−7.03</u>	2.77
	Car (mg m^{-2})	0.81	6.10	5.69	5.60	<u>−1.27</u>	2.29
	AnC (nmol m^{-2})	0.90	0.73	5.27	5.33	0.03	3.16
	Flv (nmol m^{-2})	0.87	6.68	13.67	13.66	2.01	2.77
	Phe (mL L^{-1})	0.93	7.34	2.92	2.95	0.04	3.78
1500–1150 cm^{-1}	Chla (mg m^{-2})	0.61	54.29	26.65	26.53	<u>−4.47</u>	1.60
	Chlb (mg m^{-2})	0.58	30.98	70.99	31.36	<u>−63.85</u>	1.54
	Chla+b (mg m^{-2})	0.61	88.39	38.23	38.35	<u>−4.81</u>	1.60
	Car (mg m^{-2})	0.66	19.17	7.92	7.84	<u>−1.59</u>	1.71
	AnC (nmol m^{-2})	0.91	1.43	4.98	5.02	0.21	3.33
	Flv (nmol m^{-2})	0.89	8.45	12.85	12.90	1.27	3.02
	Phe (mL L^{-1})	0.79	9.68	4.71	4.76	0.03	2.18

The PLSR models, according to the R^2_{CV} , were considered very good (0.81–0.88) for Car, AnC, and Flv, and excellent (0.91–0.93) for Phe in the 4000–400 cm^{-1} data. Conforming to the RPD_p metric, six parameters were considered excellent (>2.10) in the models evaluated, with the exception of Chla (Table 3).

The RMSE_{CV} of each variable was close to the RMSE_p ($\text{RMSE}_{\text{CV}} \approx \text{RMSE}_p$); for example, in some cases, it was slightly smaller, and slightly higher in others (Tables 2 and 3). The values of bias were close to zero (AnC, and Phe, full spectra; AnC, and Phe, 1500–1150 cm^{-1}) or negative for many (Chlb, Chla+b, Car, 4000–400 cm^{-1} ; Chla, Chlb, Chla+b, Car; 1500–1150 cm^{-1}) variables in the cross-validation and prediction analyses.

All seven evaluated parameters were adjusted with an independent dataset (ATR-FTIR spectroscopy data) from those used in the cross-validation phase to evaluate the ability of PLSR models to predict these chemometric parameters. In this sense, the reference vs. predicted data, as well as the results of the multivariate statistical results and machine learning algorithms, are shown in Tables 2 and 3.

The models of the predicted lettuce variety parameters were adjusted according to PLSR of the number based on factors before being tested by cross-validation phases. Therefore, the relationship between the predictor (wavenumbers; vibrational bands) and predicted (chemometric; lettuce parameters) variables was better explained using models containing five or seven factors for 1500–1150 cm^{-1} and 4000–400 cm^{-1} , respectively.

2.7. Regression Coefficients (RCs) and Variable Importance in Projection (VIP)

The β -coefficients (RCs) of the regression and variable importance in projection (VIP) metrics of the PLSR model are reported in Figure 7. Many regions of valleys and peaks where the RC and VIP applied a substantial effect on the construction of the prediction model were generally well distributed among all spectra or ranged wavenumbers (medium IR-bands).

The RC and VIP values used for PLSR models contrast between 10 and 16 selected wavenumbers (peaks and valleys). In some cases, the same wavenumbers were selected using 4000–400 cm^{-1} or 1500–1150 cm^{-1} data (Table 4; Figure 7). In this sense, the parameters that obtained RPDs higher than 3.00 (Table 3) (AnC, Flv, and Phe) found excellent predictions for VIP-selected wavenumbers to improve models (Table 4).

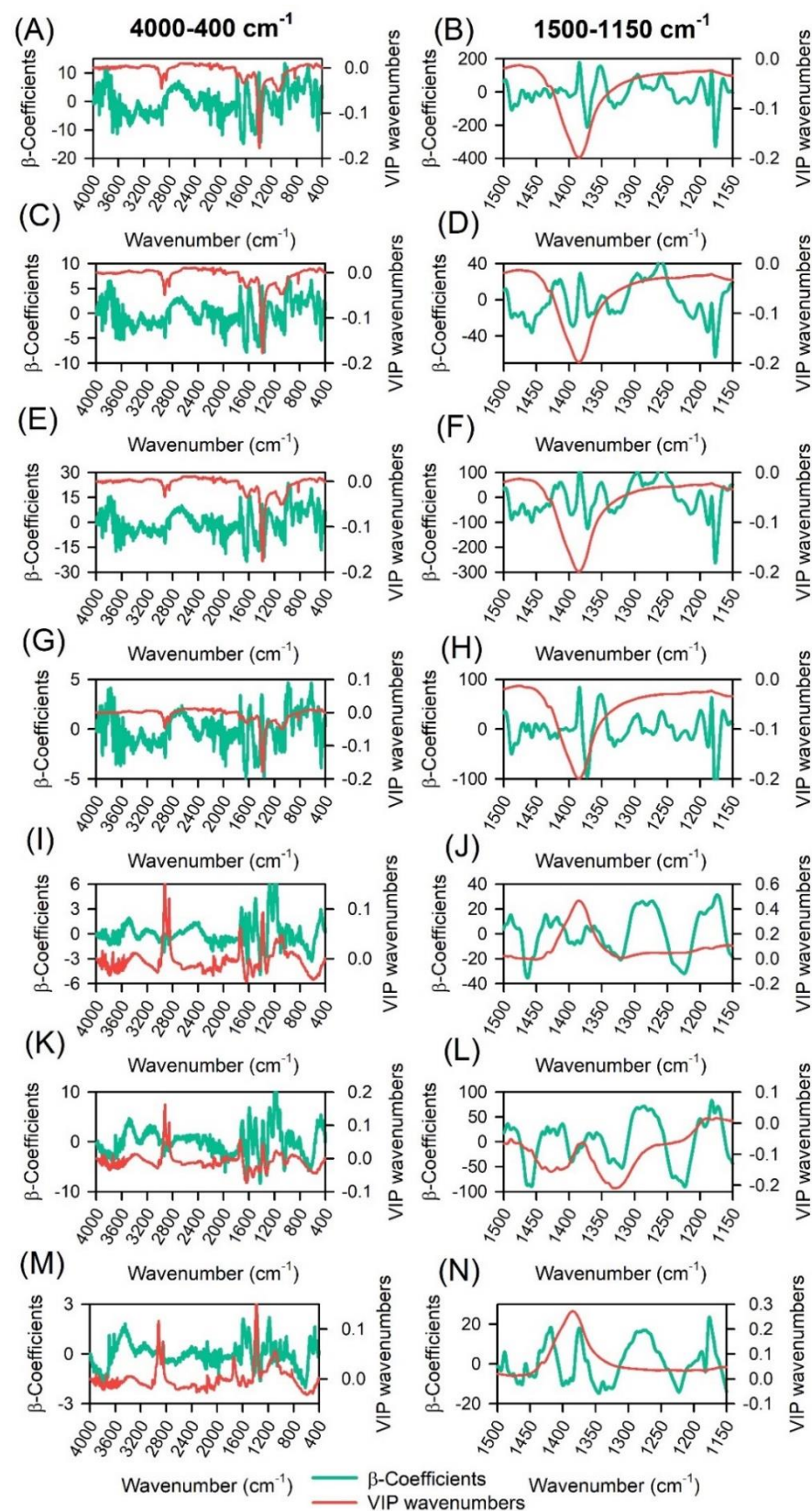


Figure 7. β -Coefficients (RCs) and variable importance to the projection (VIP) method-based ATR-FTIR spectroscopy. (A,C,E,G,I,K,M) to 4000 at 400 cm^{-1} and (B,D,F,H,J,L,N) to 1500 at 1150 cm^{-1} by ATR-FTIR spectroscopy. (A,B) Chlorophyll *a* (mg m^{-2}); (C,D) Chlorophyll *b* (mg m^{-2}); (E,F) Chlorophyll *a* + *b* (mg m^{-2}); (G,H) Carotenoids (mg m^{-2}); (I,J) Anthocyanins (nmol m^{-2}); (K,L) Flavonoids (nmol m^{-2}); (M,N) Phenolic compounds (mL L^{-1}).

Table 4. The most responsive variable importance for projection (VIP) by wavenumbers selected between 4000 at 400 cm^{−1} and 1500 at 1150 cm^{−1} by ATR-FTIR spectroscopy according to the regression β -coefficient for quality parameter for the PLSR model to predict. Parameter abbreviations—see Abbreviation list or Table 1.

Parameters	4000-400 cm ^{−1}	1500-1150 cm ^{−1}	Number of Shared VIP	Most Responsive VIP by Wavenumbers (cm ^{−1})	
				4000-400 cm ^{−1}	1500-1150 cm ^{−1}
Chla (mg m ^{−2})	13	10	3	2624; 2613; 2590; 2576; 2509; 2453; 2435; 2165; 1410; 1392; 1384 ; 1369; 1361;	1486; 1477; 1461; 1450; 1415; 1396; 1382 ; 1359; 1182; 1177;
Chlb (mg m ^{−2})	16	12	3	2622; 2613; 2590; 2580; 2511; 2460; 2453; 2433; 2322; 2163; 2050; 1405; 1400; 1384 ; 1375; 1359 ;	1498; 1486; 1479; 1463; 1452; 1415; 1402; 1384 ; 1375; 1363; 1359 ; 1182
Chla+b (mg m ^{−2})	14	11	4	2624; 2613; 2601; 2590; 2578; 2511; 2457; 2447; 2435; 1415 ; 1404; 1388; 1378; 1359 ;	1496; 1485; 1479; 1461; 1452; 1415 ; 1398; 1386; 1377 ; 1361 ; 1184;
Car (mg m ^{−2})	13	11	4	2624; 2613; 2590; 2460; 2449; 2439; 2323; 2163; 2050; 1415 ; 1384; 1367; 1359 ;	1492; 1486; 1475; 1463; 1448; 1415 ; 1392; 1386; 1373 ; 1361 ; 1182;
AnC (nmol m ^{−2})	15	12	1	2935; 2923; 2914; 2902; 2852; 1630; 1648; 1637; 1629; 1384; 1378 ; 607; 599; 584; 578;	1485; 1477; 1463; 1459; 1452; 1415; 1407; 1392; 1377 ; 1359; 1323; 1313;
Flv (nmol m ^{−2})	13	13	0	2939; 2929; 2916; 2898; 2858; 2848; 1666; 1652; 1647; 1639; 1621; 1537; 1533;	1431; 1427; 1355; 1346; 1334; 1326; 1307; 1205; 1193; 1180; 1159; 1153; 1149;
Phe (mL L ^{−1})	13	12	5	2929; 2921; 2914; 1404; 1396 ; 1384; 1375; 1367 ; 613; 599; 588; 572; 561;	1500; 1496; 1481; 1479; 1459; 1446; 1415; 1409; 1402; 1394 ; 1373; 1361 ;

The fingerprints of ATR-FTIR spectra in Table 5 could be simplified by using the most VIP values of wavenumbers observed in the majority 1500-1150 cm^{−1}. In addition, many values between 1800-1500 cm^{−1} were related to phenolic compounds. Thus, values in reference bands vs. experimental sample spectra showed a high correlation between the pigments analyzed. For example, anthocyanin and flavonoid fingerprints were marked by ring stretching of C=C or aromatic rings to vibrational bands. VIP showed many bands related to aromatic stretching ν (C=C), stretching ν (C–C), or ring stretching ν (C=O), which could be Phe, Flv, as well as –CH stretching by AnC, Chls, and Car compounds (Tables 4 and 5).

Table 5. Band assignments of ATR-FTIR spectra characteristic of the phenolic compounds and flavonoids in different spectra obtained by ATR-FTIR and the corresponding references.

Band (cm ^{−1})	Origin	Assignment	Reference
3436	chlorophyll and carotenoids	<i>trans</i> –CH=CH– bond	[41,42]
2863	chlorophylls and carotenoids	–CH stretching	[11,14,43]
1735	lignin-phenolic compounds	stretching ν (C=O)	[44]
1650-1550	phenolic compounds and flavonoids	stretching ν (C=O)	[44,45]
1650-1400	flavonoids, phenolic compounds, and other phenolics	ring stretching ν (C=O), ν (C=C) phenolic acids, ν (C–C) aromatic	[24,44]
1545	amide II	stretching (C=N) and (N–H)	[44]
1525	chlorophylls and carotenoids	(N–H bending)	[46]
1515	phenolic compounds and flavones phenyl ring	ring stretching C=C, aromatic conjugated C=C, N–H	[44,47]
1464-1429	presence of flavonoids and triterpenoids	–O–H	[47]
1455-1440	polysaccharides, lipids, proteins	asymmetric bending of –CH ₂ and –CH ₃	[44]

Table 5. Cont.

Band (cm ⁻¹)	Origin	Assignment	Reference
1440-1436	phenolic components	aromatic stretching $\nu(\text{C}-\text{C})$ stretching $\nu(\text{C}=\text{C})$	[24,25]
1430-1420	polysaccharides, alcohol, carboxylic acid	bending $-\text{OH}$	[44]
1425-1380	aromatic compounds	stretching $\nu(\text{C}-\text{C})$	[7,14]
1300-1260	polyphenols and flavonoids	stretching $\nu(\text{C}-\text{C})$	[7,14]
1381	polysaccharides	symmetric bending $-\text{CH}_2$ and $-\text{CH}_3$	[48]
1293	wavenumber of the corresponds flavonoids	O-H	[47]
1260-1180	phenolic components	stretching $-\text{C}-\text{C}-\text{O}$	[48]
825-810	polyphenols	C-H out of plane bending vibrations of phenyl rings	[7]
790-760	phenolic components and flavonoids	aromatic ring vibration	[7,14,48]

3. Discussion

3.1. Descriptive Analysis and Cluster Heatmap

The variability observed in lettuce varieties with distinct pigment contents (Figures 1 and 3) was explored to estimate parameters such as Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe based on the ATR-FTIR spectra between 4000-400 cm⁻¹ (Figures 2 and 3). Some approaches used (Tables 3 and 4; Figures 2–6) aid a rapid determination with high-throughput measurements that produce data-rich results. In particular, ATR-FTIR fingerprint-based biomolecule analysis techniques should play a key role in the development of fast, efficient, and simple prediction crop phenotyping in response to monitoring the dynamic bioaccumulation of pigments [7,9,28,35]. However, some statistical approaches, combined with machine learning algorithms (LDA and SVM), were labeled as having low and moderate prediction accuracy when the first derivative or 2° polynomial, radial, or sigmoid was applied with algorithms (Tables 2 and 3; Figures 4 and 6).

In general, comparison of the full-green, green-purplish, and full-purple groups showed that plants were associated with high plastic development and high pigment content. Natural variability interferes with machine learning algorithms or classification based on cluster analyses (Figures 1 and 3). In this sense, the cluster heatmap showed the highest efficiency in grouping similar lettuce varieties based on pigment profile analysis [48–50]. Heatmaps are a better tool to differentiate varieties or distinguish possible interactions of molecules present in leaves [51,52]. For example, a frequent collection of data from different leaves accumulates distinct levels of pigments. Under these conditions, it was possible to separate large differences between the minimum and maximum values observed with high accuracy from ATR-FTIR [14,22,42]. We observed that the relationship among pigments was more significant than the absolute contents. Additionally, it is possible to associate many tools to classify plants based on ATR-FTIR to understand the clustering of lettuce varieties [53,54] and other crops [55,56].

3.2. ATR-FTIR Spectroscopy and Tracking of Fingerprints

The differences among varieties were detected in the all-spectroscopy curve [42]. All spectral curves exhibit bands associated with vibrational modes from pigments, such as carotenoids and chlorophylls, which are more or less intensely detected, and those variations are related to pigments in each lettuce variety. Many compounds and proteins present in leaves contribute to the total vibration bands. However, the absorbance of a chemical band with characteristic vibrations defines functional groups, such as Flv or AnC, that decrease in green lettuce varieties [25,57]. Those outside chloroplast pigments from minor vibrational bands could be related to pigments present in vacuoles. Flv and AnC

were often present at higher levels in the leaf epidermis [9,48] (Table 5). At 1800–1500 cm^{-1} , the highest values were related to the vibrational wavenumber related to the cell wall components but not the chloroplast pigments.

Concerning the chlorophyll variable (Figure 4), the high RC and VIP values at 3467–3353 and 1732–1151 cm^{-1} , which are related vibrational bands, possibly obtained a higher precision and were more stable in the characteristics of each lettuce variety. The value obtained at 1764–1658 or 1550–1443 cm^{-1} is associated with the maximal and minimum reflectance by chlorophyll. Notably, the absorption of 1550–1150 cm^{-1} was influenced by a few bands by the chlorophyll content [58] to distinguish between green-purple varieties. In addition, the chlorophyll content and these vibrational bands promote biomass accumulation in plants and promote possible predictions of these pigments [3,22,36]. The bands at 2800–2500 and 1200–1100 cm^{-1} , in turn, are possibly due to the constituent by amino acids and proteins (for example, N–H in plane bending/C=O stretch + N–H bend/C–N stretch combination bands) in plant constituents and metabolites, which could likewise be associated with many important dissipative energies (carotenoids) and antioxidant molecules that protect lettuce [3,17,58–61].

At 1750–1700 cm^{-1} , substantial differences were detected in the spectral analysis. Spectral data in this region support obtaining fingerprints related to vibrational modes of phenol compounds and aromatic groups, such as in 1750–1735 cm^{-1} associated with saturated aliphatic esters (C=O stretching), 1740–1720 cm^{-1} associated with saturated α,β -unsaturated esters and aliphatic aldehydes (C–C) or 1715 cm^{-1} , followed by saturated aliphatic ketones [14,16,42]. In addition, the peak at 1735 cm^{-1} was characterized by double bond stretching ($\nu\text{C}=\text{O}$) from ester bonds, which may be linked to lignin in cell walls or other phenolic compounds deposited on cells and soluble phenolics [62]. Other research suggests that the peak at 1700 cm^{-1} was similar to the keto C=O band of chlorophylls [46,63]. The peak at 1670 cm^{-1} by FTIR showed good correlation with the stretching vibration of carbonyl (C=O) and amide I (C–N). Furthermore, the peak at 1640–1630 cm^{-1} can be correlated to the stretching of aromatic C=C bonds in anthocyanins and other phenolic compounds with chlorogenic acids, coumarins, catechins, flavonoids, and antioxidant compounds [7,42,64].

Moreover, bands at 1660, 1650, 1606, 1545, and 1515 cm^{-1} at a stretching of C=O referred to the amide I and II groups, and aromatic elongation C=O was linked to flavonoids, anthocyanins, and other soluble phenolics. Additionally, it was reported [7,25] that an elongation of C=N and NH, referring to amide II, and a stretching of aromatic C=C, mainly indicated phenolic compounds. In addition, the 1800–1501 cm^{-1} spectral range included many wavenumbers corresponding to spectral double bond (C=C) stretching (Tables 4 and 5; Figure 4). It is possible to link it to anthocyanins and flavonoids [11,14,41]. The peak at 1640–1630 cm^{-1} can be correlated to the stretching of aromatic C=C bonds in anthocyanin molecules [25]. Specifically, 1640 cm^{-1} , following the report in [48], was related to the vibrational modes of C=C groups in aromatic rings. The 1440 to 1436 cm^{-1} peak corresponds to vibration aromatic stretching ($\nu\text{C}=\text{O}$) and double bond stretching between carbon ($\nu\text{C}=\text{C}$). Furthermore, the region between 1260 and 1180 cm^{-1} is known to be related to the phenol compound vibrational mode due to stretching out of the C–C–O phase. For example, peaks between 1293 and 790–760 cm^{-1} were associated with flavonoids, such as O–H and aromatic ring vibrations. Following the report in [45], a wavenumber of 1640 cm^{-1} was correlated, in which these bands were mainly occupied by vibrational motions of C=C groups in the aromatic rings. In this sense, other research has similarly reported evidence for flavonoids and phenolic compounds when using hyperspectroscopy, ^1H -NMR/MRS spectroscopy, and analysis by multivariate tools [53]. Additionally, peaks before 1410 cm^{-1} were reported for phenolic compounds, even at 1663 cm^{-1} for possible terpenes as C–H bonds attached to the benzene ring of phenol and aromatic molecules. All peaks contribute to the elucidation of lettuce ATR-FTIR fingerprints [9,11,65–67].

The range of 1455–1440 cm^{-1} is associated with $-\text{CH}_2$ and $-\text{CH}_3$ asymmetric bending related to polysaccharides, lipids, and proteins present in plant samples. In addition,

1430 to 1420 cm^{-1} is associated with $-\text{OH}$ bending with polysaccharides, alcohol, and carboxylic acid in the cell wall. The bands at 1440 to 1436 cm^{-1} were associated with phenolic compounds due to the vibrational mode of aromatic stretching between carbons ($\nu\text{C}-\text{C}$) conjugated with a double bond elongation between carbons ($\nu\text{C}=\text{C}$) [24,25]. The region between 1260 and 1180 cm^{-1} is known for the vibrational mode of phenols due to out-of-phase $\text{C}-\text{C}-\text{O}$ stretching, which can be associated with flavonoids since they belong to the group of phenolic compounds [9,48]. In this sense, the 1197 cm^{-1} peak likely corresponds to the band/peak of anthocyanin molecules ($r = 0.960$; $p < 0.001$), and similarly, the following results are shown in [25]. Here, ATR-FTIR data also showed a high and positive correlation for anthocyanins and the 1197 cm^{-1} peak (Tables 4 and 5; Figures 3 and 7) [53,62,68], followed by analyses by Gaussians or LDA algorithms (Figures 3 and 4). ATR-FTIR spectroscopy displayed stretching $\text{C}-\text{O}$ modes that are mainly assigned to the region from 1198 to 1025 cm^{-1} for symmetric elongation of the $\text{C}-\text{O}-\text{C}$ group in the pyran ring of condensed tannins and other flavonoids and phenolic compounds [10,22], similar to the report in Figures 4, 6 and 7. For example, for anthocyanins, flavonoids, or phenolic compounds, the corresponding asymmetric stretching occurs approximately in the 1285 to 1260 cm^{-1} spectral range [69,70]. Analogously, the 1202 to 1200 cm^{-1} peak, attributed to $-\text{C}-\text{OH}$ aliphatic stretching, is reported to be a typical feature for red wine or purple carrots correlated with condensed tannins [22] or increased anthocyanin levels (Figure 3). Thus, the data obtained by PCA and LDA proved to be good tools by which to discriminate the most significant spectral bands following [14,15,22,25,41].

When all spectroscopy curves were applied to distinguish a particular variable, only some FTIR-specific bands were used, and the effects of several background errors on misclassification were reduced, as described in [17,36,71]. In this sense, when applied, the full spectrum might improve the accuracy and reliability of the analysis to better generate a prediction model. Similarly, [72] reported that the effects of various external-factor interference sources can be reduced or eliminated when using the total curve in relation to using individual or only a few peaks [59,73]. In general, the full spectrum, in contrast to a specific range, is more appropriate, following [17,36,71], which demonstrated a valuable method to analyze ATR-FTIR by LDA and SVM, with greater accuracy and precision to classify and predict functional pigment content in lettuce varieties.

3.3. Efficiency of Machine Learning Algorithms to Classify Lettuce Plants

Machine learning algorithms were applied to sample group classification by using ATR-FTIR spectroscopy raw data and PCA, LDA, and SVM machine learning algorithms. As a whole, the accuracy was slightly improved by using PCA and LDA data. Using ranged data (1500–1150 cm^{-1}) produced superior accuracy in relation to the full spectra (4000–400 cm^{-1}), which showed higher differences in the overall accuracy between datasets. Using the first three PCs improved computational processing because they contained a reduced dimension. An overall accuracy of $>92.2\%$ for linear LDA and SVM was obtained by using three PCs with $\approx 95\%$ of variance. Similarly, but with slightly higher accuracy at 1500–1150 cm^{-1} in the ATR-FTIR spectroscopy method, LDA-linear carried 95.6% of the data variance. Following [11], data mining and machine learning can be improved to classify with higher precision in relation to quadratic, Mahalanobis, or 2° polynomial, radial or sigmoid validation methods (testing), and SVM or LDA can be considered the most promising algorithm, since it reached an accuracy with the smallest number of PCs (Figure 6). However, this is true for all machine learning experiment analyses of experimental data following [11,14,74]. The influence of data distribution was more significant for machine learning (based on a linear interaction-pigment contents vs. lettuce varieties), which possibly contributes to the high accuracy and separation of the analyzed lettuce groups [3,53,54].

Many studies use PCA to extract the most useful FTIR or ^1H -NMR spectroscopy information [7,14,75], which is a useful method to classify and discriminate the most responsive wavenumbers [14,43,67]. This method shows good results in situations with

fewer impurities (high signal-to-noise ratio) or simple samples [14,59]. For lettuce, there are uncountable types of molecules present in leaf powder that affect the results, and which produce a more complex spectrum. There are many interferences in a particular band. Choosing appropriate bands from all available bands is a challenge to obtain useful information. It often requires expert knowledge and rich practical expertise, and it is not easy to achieve automated and rapid detection. SVM and LDA are often used for the quantitative analysis of spectra, but LDA adjusts many features when processing spectroscopy data, which leads to the occurrence of overfitting (Figures 2 and 6).

3.4. Partial Least Squares Regression (PLSR) for Classification-Prediction Pigments

The results of R^2_{CV} , R^2_P , RPD_{CV} , RPD_P , $RMSE_{CV}$, and $RMSE_P$ for the full (4000–400 cm^{-1}) and selected spectra (1500–1500 cm^{-1}) were modified according to the parameters estimated (Table 2; Figure 4). As a whole, the approach used (Figure 6) showed good results. Other techniques and multivariate statistical analysis might achieve similar predictive results using variable importance to the projection (VIP), STEPDisc tools, regression coefficients (RCs), STEPWISE methods, or processes with noise reduction efficiency. In this sense, standard normal variate (SNV), Savitzky–Golay (SG), multiplicative scatter correction (MSC), and other correction methods vary according to the attribute measured but have been alternatively used to obtain good to excellent results based on PLSR methods [5,13,24,53]. It is interesting and very important because pathways of secondary metabolism produce many biomolecules in the plant cell. For example, the inclusion of vibrational modes induced by spectroscopy, such as $-\text{NH}_3$, $-\text{N}-\text{H}$, and $-\text{C}-\text{H}$ stretching vibrations from aromatic rings associated with specific wavenumbers and associated with compounds along with the structure of the pigments, proteins, and fatty acids, allowed us to obtain better PLSR models with statistical metrics, such as R^2 and RPD, ranging from good to excellent, respectively, for all assessed chemometric vibrations [4,35,42,54].

Cross-validation showed that the results for the statistical metrics were slightly higher than those obtained in the prediction phase, as projected, since the number of samples used to obtain the model was smaller in the calibration phase [76]. Moreover, [71,76] used ATR-FTIR spectroscopy methods to predict distinguishing pigments in lettuce varieties and similarly obtained an increase in RMSE in the prediction phase, which connected many metabolites produced in secondary metabolism [17].

However, in the prediction phase, the R^2_P values calculated for Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe were similar to other variables from red lettuces [4], green tea [46], *Camellia sinensis* [77], *Enhalus acoroides* [16], tobacco [17], and peatlands [78] employing SWIR and/or MWIR and/or LWIR spectroscopy or hyperspectroscopy [14,79–82]. In this research, the difficulty of establishing metrics with better results for the leaf pigments (chloroplast and outside-chloroplast pigments, particularly in the phase of the prediction) could be due to the limited number of samples ($n = 52$). In addition, the authors of a previous study [36] also used specific range bands and were able to define R^2_P with better metrics [36,71], but RPD, offset, or bias obtained the lowest precision. However, when they set specific spectral peaks or valleys, the prediction models showed high precision, but with higher noise by the backgrounds and lower accuracy [4,5,55].

Our data showed the high accuracy of the model to estimate attributes with spectroscopy techniques confirmed to be useful (or excellent according to the Offset and RPD_P) for many of the variables (i.e., biomolecules) tested. In this sense, the main advantage of the proposed method is the ability to simply form by predicting simultaneous inferences from seven chemometrics (Chla, Chlb, Chla+b, Car, AnC, Flv, and Phe). In this sense, it could be used for monitoring many molecules and metabolites synthesized by secondary metabolism. In addition, the crop production status in a single collection without the need for preparation using reagents or expensive preparation of samples (i.e., Si and KBr) and high-cost equipment for acquisition (^1H -NMR, LC-MS/MS, SAXS, DRX, or hyperspectroradiometer) [2,17,71].

3.5. Regression Coefficients

Regression coefficients (RCs) and variable importance in projection (VIP) were demonstrated with essential tools to classify and avoid possible errors to estimate by correlations. Thus, this work aimed to better understand how each variable (wavenumbers) contributed to the significant variation in the model used to estimate the parameters. In most cases, high RC (Figures 4 and 7) and VIP wavenumbers (Table 4; Figure 7) were well distributed across the entire portion of the spectrum for all groups of molecules analyzed [14,83].

Many studies have investigated the potential of physical tools as a technique for estimating radish, cauliflower, lettuce, native species, and other important crop and cultivate species, which, by pigments, primarily consider the region with wavenumbers less than $1500\text{--}1150\text{ cm}^{-1}$ [34,35,53]. In general, they were biased and obtained low-accuracy outputs. In the present study, higher regression coefficients were obtained using full range values. We highlight that, on average, 23.4% of VIP values were shared with ranged data; in other words, $\approx 75\%$ of VIP values were selected outside of the selected values.

Acquisition of full spectra using spectroscopy sensors and analysis by curve deconvolution associated with PLSR [84] and many other multivariate tools [85–87] show supplementary, robust, and reliable models, as evidenced by the RMSE values that are related to the use of range bands. The use of full bands takes advantage of specific wavenumbers [2,12]. In this sense, the flexibility of choosing between different spectral bands by performing a discriminant analysis linked with the high precision and accuracy of high-resolution spectral data enables metabolites to be identified in secondary metabolism, variety discrimination, predicted pigment content and concentrations, and applied machine learning in ATR-FTIR spectroscopy tools [3,12,36,53].

3.6. Benefits and Limitations of Using ATR-FTIR Spectroscopy for Classified and Predicted Pigments in Lettuce

The method used here with available spectroscopy allows us to rapidly and precisely categorize based on pigment and vibrational group profiles. In addition, the capacity to specifically discriminate between variety-mediated regulation of development and dynamics is applicable to monitoring crops and dynamics of the metabolite states in lettuce plants [4,7,22,36,58]. By combining ATR-FTIR spectroscopy data and robust multivariate statistical modeling-based machine learning, this research described a prospective method with the possibility of concomitantly classifying and predicting the spectroscopy pigments in leaves.

Furthermore, the knowledge of wavenumbers for FTIR spectroscopy phenolic/anthocyanin compound analysis variability among lettuce varieties could be used at a commercial level as increasingly important criteria for the identification of new high-performing genotypes and cultivars for particular environments and cultivation conditions [2,86], which are rich in phytochemicals and have potential health benefits. In this sense, agronomic protection, food security, and molecular characterization showed the highest accuracy of ATR-FTIR tools, together with machine learning algorithms, in making better decisions regarding increased productivity and selecting the most responsive varieties in biomolecules plant and crop sciences.

Many metabolites (molecules, biocompounds, or pathways of secondary metabolism) differentiate varieties of crop plants. However, their differentiation is challenging (i.e., depending on the plant genotype or species), and currently, still depends on human expertise. In this sense, they are associated with many molecules, including chloroplast or outside-chloroplast molecules [11,13,35].

Finally, techniques involving ATR-FTIR spectroscopy, together with machine learning algorithm approaches, are highly valuable and promising in regard to meeting this need, as they are fast and do not require previous sample preparation with chemical reagents (Si and KBr), or expensive and classic equipment [2,71]. Thus, in this study, we showed that the use of ATR-FTIR spectra is a promising strategy for understanding, classifying,

and predicting molecular content relationships derived from eleven lettuce varieties, and could be applicable to crop plants with satisfactory accuracy and precision.

4. Material and Methods

4.1. Plant Material, Growth Conditions, and Experimental Design

Lettuce seeds (*Lactuca sativa* L.) were germinated on Germitest[®] paper immersed in 4 mL of Hoagland's solution (pH 5.4) in a dish. After 15 days of growth, seedlings were transplanted to MecPlant[®] (MecPrec Ind., Telêmaco Borba, Paraná, Brazil), a commercial substrate, and then transported to grow in a greenhouse.

Experiments were conducted in the greenhouse at the State University of Maringá, Maringá, Paraná, Brazil. Eleven lettuce varieties (V_X) were grown (V01—Rainha de Maio, V02—Vitória, V03—Maravilha de Inverno, V04—Grandes Lagos Americana, V05—Mimosa Prado, V06—Quatro Estações, V07—Batávia Joaquina, V08—Mimosa Vermelha, V09—Batávia Cacimba, V10—Pipa and V11—Mimosa Rubi) in pots with a capacity of 3 L containing substrate (soil:sand:organic compounds, in a proportion of 3:2:2). Fertilization was applied with N-P-K (10-10-10) as recommended for lettuce plants. Light was monitored and reached $1200 \mu\text{mol m}^{-2} \text{s}^{-1}$ irradiance, measured by a LI-190R quantum sensor (Li-Cor Inc., Lincoln, NE, USA) under a natural photoperiod (approx. 12/12; light/dark) at 30 °C (± 5 °C) with 50–80% relative humidity. The varieties were visually chosen by color (green to purple colors) due to the classes and content of different pigments (Figure 1). The cultivation was conducted through a completely randomized design with 11 treatments and 12 biological repetitions.

4.2. Leaf Analysis

Leaf area was measured using an LI-3100C leaf area meter (Li-Cor Inc., Lincoln, NE, USA) and was used to estimate the leaf pigment content expressed in area.

4.3. Leaf Pigment Profiling

Simultaneous quantification of chlorophyll *a* (Chl*a*), chlorophyll *b* (Chl*b*), total chlorophyll (Chl*a+b*), carotenoids (Car), anthocyanins (AnC), and flavonoids (Flv) in methanolic extracts of the leaves was conducted, as described in [21,88]. Spectral scans were performed on a Lambda 1050 UV/VIS/NIR Spectrophotometer (PerkinElmer, Inc., Waltham, MA, USA). The Chls, Car, AnC, and Flv concentrations were reported based on area unity.

4.4. Total Soluble Phenolic Compound Quantification

Total soluble phenol (PhC) quantification was carried out exactly as described in [2,3] and analyzed on a Lambda 1050 UV/VIS/NIR Spectrophotometer (PerkinElmer, Inc., Waltham, MA, USA) at 725 nm. The equivalent PhC concentration was determined using gallic acid as a reference, $Y = 87.65x + 1.651$; $r^2 = 0.994$.

4.5. Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR) Analysis

Fourier transform infrared spectroscopy was performed on oven-dried (70 °C) leaf samples with a Bruker Vertex 70v FTIR spectrometer (Bruker Optik GmbH, Ettlingen, Germany, DEU) equipped with A225/Q Platinum attenuated total reflectance (ATR) with a single reflection diamond crystal (Bruker Optik GmbH, Rosenheim, Germany, DEU). The spectra were obtained from 4000 to 400 cm^{-1} wavenumbers with a spectral resolution of 4 cm^{-1} . Two replications were performed for each spectrum, which were collected from an average of 300 scans sample^{-1} at room temperature (25 °C). All spectra were equally corrected by baseline using OPUS software (Bruker Optik GmbH, Rosenheim, DEU), and the intensity was normalized at 3270 cm^{-1} (OH stretching band).

4.6. Statistical Analyses

4.6.1. Descriptive Analysis

Statistical analyses for the classification and prediction of pigment contents were performed using The Unscrambler x10.4[®] (Camo Software, Oslo, Norway, NOK), SigmaPlot[®] 10.0 software (Systat Software Inc., GmbH, Ettlingen, Germany, DEU), Statistica[®] 12.0 software (Statsoft Inc., Tulsa, OK, USA), Excel 2021[®] (Microsoft Office Inc., Sunnyvale, CA, USA), and the R package R-Core-Team (2020) (<https://www.R-project.org> (accessed on 20 September 2022)) [87].

4.6.2. Statistical Analyses of the ATR-FTIR Spectral Signature

The effects of ATR-FTIR spectra on the untransformed wavenumber profiles (averaged per sample). Data were evaluated using PERMANOVA by using Euclidian measurement algorithms of dissimilarity using the Euclidean distance with the “vegan” package in R-Core-Team (2020) (<https://www.R-project.org> (accessed on 20 September 2022)).

4.6.3. Principal Component Analysis (PCA)

Principal component analysis (PCA) was performed using “The Unscrambler X” software, version x10.4 (CAMO AS, Oslo, Norway, NOK), as an indicator of whether the variance in the ATR-FTIR spectroscopy between the lettuces could be explained or not ($p < 0.05$) and how effectively biomolecules based on secondary metabolism in plant varieties could be clustered. The numbers of principal components (PCs) were based on the highest average *Kappa* (*K*) and accuracy (Acc) values obtained for the validation models from partial least squares regression (PLSR). PLSR analysis was performed exactly as described in [2,3,17].

4.6.4. Linear Discriminant Analysis (LDA) for Machine Learning

Linear discriminant analysis (LDA) was used to obtain models to classify each ATR-FTIR spectroscopy-based machine learning algorithm of lettuce variety plants. Before obtaining the discriminant models, a routine PCA-LDA (three components) was performed to select wavenumbers that best explained the differences in lettuce varieties [84,87]. The routine PCA-LDA selects the variables to compose the model one by one, in accordance with the partial F value input in each stage, the Wilks’ lambda value, and partial variables, such as R^2 . The proc LDA technique is carried out until no variable can be arrived at or removed from the model, as described in [2,3,17].

The ability and quality to acceptably classify the obtained models of the LDA were evaluated based on external data for the validation method by random dataset division (60:40), as described in [3]. Linear, quadratic, and Mahalanobis models of analysis were applied to classify lettuce-based machine learning algorithms [50].

4.6.5. Support Vector Machine (SVM) for Machine Learning

The support vector machine (SVM) is a supervised, kernel-based learning method [11,13,33]. Using kernel functions and a first derivative, SVM applied to a spectroscopy dataset allows for higher-dimensional feature space data than the original ATR-FTIR spectroscopy, as described in [11,13,35,89]. In addition, the k-kernel reports should be viewed as a mapping of nonlinear data to a higher dimensional feature space while providing a computation short-cut by allowing linear algorithms to work with a higher dimensional feature space [11,13,35,89]. The last decision-based purpose of the SVM is determined by a few support vectors (machine learning). The function based on a complexity algorithm depends on the number of support vectors based on 4000 to 400 cm^{-1} and 1500 to 1150 cm^{-1} . The training and validation (testing) of the SVM algorithms were performed to obtain the accuracy of ATR-FTIR data analysis to predict contents similar to chemical analysis.

4.6.6. Partial Least Squares Regression (PLSR) Analysis of ATR-FTIR Spectroscopy

The data were subjected to Shapiro–Wilk and Bartlett’s tests to evaluate the normal distribution, homogeneity, and variance before obtaining the prediction models of quality (chlorophylls, carotenoids, anthocyanins, flavonoids, and phenolic compounds). For all variables, 4000 to 400 cm^{-1} or 1500 to 1150 cm^{-1} were recorded to improve the accuracy of the data. Subsequently, the data (ATR-FTIR spectroscopy of lettuce varieties) were centered on the mean and subjected to PLSR analysis. The algorithm for model inputs was NIPALS, and output outlier limits were defined by Leverage’s type and analyzed using Hotelling’s T^2 test (limit of 5%). Spectral data of the 132 samples of different secondary metabolites collected were divided into two groups (based on biological replicates). The first group consisted of 60% (80) of the samples with the aim of generating the model (training), while the second group was represented by 40% (52) of the remaining samples with the aim of testing (prediction) the PLSR model, exactly as described in [2].

The calibration (Cal) and leave-one-out cross-validation (Cva) methods were used to classify and predict the quality parameters based on chemometrics [biochemical compounds, i.e., chlorophyll *a* (mg m^{-2}), chlorophyll *b* (mg m^{-2}), chlorophyll *a+b* (mg m^{-2}), carotenoids (carotenes and xanthophyll; mg m^{-2}), anthocyanins (nmol m^{-2}), flavonoids (nmol m^{-2}), and phenolic compounds (mL L^{-1})].

The predictive capacity of the calibration models was estimated by calculating all metrics, such as R^2 (coefficient of determination), offset, RMSE (root mean square error), and RPD (ratio of performance to deviation), and bias was determined to assess the quality and accuracy of the model, exactly as described in [3,8,17,71]. The simplified flowchart is displayed in Figure 2.

5. Conclusions

As shown in the present study, each lettuce variety has unique spectral signatures of secondary metabolites based on ATR-FTIR spectroscopy. Their relationship, based on a heatmap, is a promising tool for clustering varieties. A typical spectral signature, even as attributed to vibrational excitation of C-H bonds, covalently bonded (vC=O), (C=O stretching), aromatic C=C , C-H , N-H , NH_3 , COOH , O-H stretching from aromatic rings, is associated with the main compounds present in lettuce varieties. All pigment-based chemometric analyses, such as Chl*a*, Chl*b*, Chl*a+b*, Car, AnC, Flv, and Phe, were efficiently estimated and categorized using ATR-FTIR spectroscopy together with linear discriminant analysis (LDA) and principal component analysis (PCA)-based machine learning algorithms.

The estimated PLSR models presented R^2_P and RPD_P values > 0.80 and > 2.10 , respectively, for many variables predicted based on the full range (4000–400 cm^{-1}) or selected range (1500–1150 cm^{-1}) from ATR-FTIR spectra. The most important wavenumbers were well distributed within the three operating ranges, generally close to 3000–2700 cm^{-1} , 2600–2200 cm^{-1} , 1700–1500 cm^{-1} , 1400–1150 cm^{-1} , and 650–550 cm^{-1} , as highlighted by the regression coefficients.

Thus, our main objective was achieved in the present study, which shows the potential of ATR-FTIR spectroscopy coupled to multivariate analysis to classify and estimate specific metabolites in lettuce plants. Therefore, this procedure is a promising alternative, as it offers the advantages of being rapid and fast, and does not require previous sample preparation (i.e., Si and KBr by chemical reagents) and accurate data acquisition over a wide infrared range. ATR-FTIR spectra analyzed by machine learning algorithms could facilitate data processing. Their potential can be applied to field or other metabolites based on their vibrational bands, with an excellent capacity for monitoring and machine learning applied in lettuce and possibly to other crop-field plants.

Author Contributions: Conceptualization, R.F. and T.M.; Data curation, R.F., T.M. and M.S.G.; Formal analysis, R.F., T.M. and M.S.G.; Funding acquisition, R.F., W.C.A. and M.R.N.; Investigation, R.F., T.M., M.S.G., A.V., M.C.P., M.E.G. and F.S.; Methodology, R.F., T.M., M.S.G., A.V., M.C.P. and M.E.G.; Project administration, R.F.; Resources, R.F., T.M. and M.S.G.; Software, R.F. and M.S.G.; Supervision, R.F.; Validation, R.F. and M.S.G.; Visualization, R.F.; Writing—original draft, R.F., T.M., M.S.G., A.V., F.S., W.C.A. and M.R.N.; Writing—review and editing, R.F., T.M., M.S.G., A.V., M.C.P., M.E.G., F.S., W.C.A. and M.R.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank the editors and reviewers for their comments and suggestions to improve our work. Thank you are due to Programa de Pós-Graduação em Agronomia (PGA-UEM) and Programa de Pós-Graduação em Física (PFI-UEM) at State University of Maringá for encouragement and supporting communication.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. Abbreviations list: Acc: Average accuracy; AnC: Anthocyanins; ANOVA: Analysis of variance; ATR-FTIR: Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy; Cal: calibration; Car: Carotenoids; Chla: Chlorophyll *a*; Chla+b: Chlorophyll *a*+*b*; Chlb: Chlorophyll *b*; Cva: Leave-one-out cross-validation; Flv: Flavonoids; *K*: *Kappa* coefficient; LDA: Linear discriminant analysis; LWIR: Longwave infrared; MWIR: Mid-wavenumber infrared; PC: Principal component; PCA: Principal component analysis; PERMANOVA: Permutational analysis of variance; Phe: Phenolic compounds; PLSR: Partial least squares regression; *R*²: Coefficient of determination; RMSE: Root mean square error; RPD (ratio of performance to deviation); STEPWISE: Stepwise regression is a method of fitting regression models in which the choice of predictive variables is carried out by an automatic procedure; SVM: Support vector machine; SWIR: Shortwave infrared spectral region; V01: Rainha de Maio; V02: Vitória; V03: Maravilha de Inverno; V04: Grandes Lagos Americana; V05—Mimosa Prado; V06—Quatro Estações; V07—Batávia Joaquina; V08—Mimosa Vermelha; V09—Batávia Cacimba; V0_χ: Lettuce varieties; V10—Pipa; V11—Mimosa Rubi; VIP: Variable importance to the projection

References

1. Park, C.H.; Yeo, H.J.; Kim, Y.J.; Van Nguyen, B.; Park, Y.E.; Sathasivam, R.; Kim, J.K.; Park, S.U. Profiles of Secondary Metabolites (Phenolic Acids, Carotenoids, Anthocyanins, and Galantamine) and Primary Metabolites (Carbohydrates, Amino Acids, and Organic Acids) during Flower Development in *Lycoris radiata*. *Biomolecules* **2021**, *11*, 248. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Falcioni, R.; Moriwaki, T.; Pattaro, M.; Furlanetto, R.H.; Nanni, M.R.; Antunes, W.C. High Resolution Leaf Spectral Signature as a Tool for Foliar Pigment Estimation Displaying Potential for Species Differentiation. *J. Plant Physiol.* **2020**, *249*, 153161. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Furlanetto, R.H.; Moriwaki, T.; Falcioni, R.; Pattaro, M.; Vollmann, A.; Sturion Junior, A.C.; Antunes, W.C.; Nanni, M.R. Hyperspectral Reflectance Imaging to Classify Lettuce Varieties by Optimum Selected Wavelengths and Linear Discriminant Analysis. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100400. [\[CrossRef\]](#)
4. Clemente, A.A.; Maciel, G.M.; Siquieroli, A.C.S.; de Araujo Gallis, R.B.; Pereira, L.M.; Duarte, J.G. High-Throughput Phenotyping to Detect Anthocyanins, Chlorophylls, and Carotenoids in Red Lettuce Germplasm. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102533. [\[CrossRef\]](#)
5. Gould, K.S.; Jay-Allemand, C.; Logan, B.A.; Baissac, Y.; Bidet, L.P.R. When Are Foliar Anthocyanins Useful to Plants? Re-Evaluation of the Photoprotection Hypothesis Using *Arabidopsis thaliana* Mutants That Differ in Anthocyanin Accumulation. *Environ. Exp. Bot.* **2018**, *154*, 11–22. [\[CrossRef\]](#)
6. Fine, P.V.A.; Salazar, D.; Martin, R.E.; Metz, M.R.; Misiewicz, T.M.; Asner, G.P. Exploring the Links between Secondary Metabolites and Leaf Spectral Reflectance in a Diverse Genus of Amazonian Trees. *Ecosphere* **2021**, *12*, e03362. [\[CrossRef\]](#)
7. Sinanoglou, V.J.; Zoumpoulakis, P.; Fotakis, C.; Kalogeropoulos, N.; Sakellari, A.; Karavoltsos, S.; Strati, I.F. On the Characterization and Correlation of Compositional, Antioxidant and Colour Profile of Common and Balsamic Vinegars. *Antioxidants* **2018**, *7*, 139. [\[CrossRef\]](#) [\[PubMed\]](#)

8. Falcioni, R.; Moriwaki, T.; Furlanetto, R.H.; Nanni, M.R.; Antunes, W.C. Simple, Fast and Efficient Methods for Analysing the Structural, Ultrastructural and Cellular Components of the Cell Wall. *Plants* **2022**, *11*, 995. [\[CrossRef\]](#)
9. Baranović, G.; Šegota, S. Infrared Spectroscopy of Flavones and Flavonols. Reexamination of the Hydroxyl and Carbonyl Vibrations in Relation to the Interactions of Flavonoids with Membrane Lipids. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *192*, 473–486. [\[CrossRef\]](#)
10. Ge, Y.; Atefi, A.; Zhang, H.; Miao, C.; Ramamurthy, R.K. High-Throughput Analysis of Leaf Physiological and Chemical Traits with VIS–NIR–SWIR Spectroscopy: A Case Study with a Maize Diversity Panel. *Plant Methods* **2019**, *15*, 66. [\[CrossRef\]](#)
11. Franca, T.; Goncalves, D.; Cena, C. ATR-FTIR Spectroscopy Combined with Machine Learning for Classification of PVA/PVP Blends in Low Concentration. *Vib. Spectrosc.* **2022**, *120*, 103378. [\[CrossRef\]](#)
12. Falcioni, R.; Moriwaki, T.; Bonato, C.M.; Souza, L.A.; Nanni, M.R.; Antunes, W.C. Distinct Growth Light and Gibberellin Regimes Alter Leaf Anatomy and Reveal Their Influence on Leaf Optical Properties. *Environ. Exp. Bot.* **2017**, *140*, 86–95. [\[CrossRef\]](#)
13. Cui, X.; Zhao, Z.; Zhang, G.; Chen, S.; Zhao, Y.; Lu, J. Analysis and Classification of Kidney Stones Based on Raman Spectroscopy. *Biomed. Opt. Express* **2018**, *9*, 4175–4183. [\[CrossRef\]](#)
14. Ertani, A.; Francioso, O.; Ferrari, E.; Schiavon, M.; Nardi, S. Spectroscopic-Chemical Fingerprint and Biostimulant Activity of a Protein-Based Product in Solid Form. *Molecules* **2018**, *23*, 1031. [\[CrossRef\]](#)
15. Rana, R.; Langenfeld-Heyser, R.; Finkeldey, R.; Polle, A. FTIR Spectroscopy, Chemical and Histochemical Characterisation of Wood and Lignin of Five Tropical Timber Wood Species of the Family of Dipterocarpaceae. *Wood Sci. Technol.* **2010**, *44*, 225–242. [\[CrossRef\]](#)
16. Pharmawati, M.; Wrasati, L.P. Phytochemical Screening and FTIR Spectroscopy on Crude Extract from *Enhalus acoroides* Leaves. *Malays. J. Anal. Sci.* **2020**, *24*, 70–77.
17. Falcioni, R.; Moriwaki, T.; Antunes, W.C.; Nanni, M.R. Rapid Quantification Method for Yield, Calorimetric Energy and Chlorophyll a Fluorescence Parameters in *Nicotiana tabacum* L. Using Vis-NIR-SWIR Hyperspectroscopy. *Plants* **2022**, *11*, 2406. [\[CrossRef\]](#)
18. Hichri, I.; Barrieu, F.; Bogs, J.; Kappel, C.; Delrot, S.; Lauvergeat, V. Recent Advances in the Transcriptional Regulation of the Flavonoid Biosynthetic Pathway. *J. Exp. Bot.* **2011**, *62*, 2465–2483. [\[CrossRef\]](#)
19. Kalt, W.; Cassidy, A.; Howard, L.R.; Krikorian, R.; Stull, A.J.; Tremblay, F.; Zamora-Ros, R. Recent Research on the Health Benefits of Blueberries and Their Anthocyanins. *Adv. Nutr.* **2020**, *11*, 224–236. [\[CrossRef\]](#)
20. Aoki, T.; Akashi, T.; Ayabe, S.I. Flavonoids of Leguminous Plants: Structure, Biological Activity, and Biosynthesis. *J. Plant Res.* **2000**, *113*, 475–488. [\[CrossRef\]](#)
21. Gitelson, A.; Solovchenko, A. Non-Invasive Quantification of Foliar Pigments: Possibilities and Limitations of Reflectance- and Absorbance-Based Approaches. *J. Photochem. Photobiol. B Biol.* **2018**, *178*, 537–544. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Ricci, A.; Olejar, K.J.; Parpinello, G.P.; Kilmartin, P.A.; Versari, A. Application of Fourier Transform Infrared (FTIR) Spectroscopy in the Characterization of Tannins. *Appl. Spectrosc. Rev.* **2015**, *50*, 407–442. [\[CrossRef\]](#)
23. Njaci, I.; Ngugi-Dawit, A.; Oduor, R.O.; Kago, L.; Williams, B.; Hoang, L.T.M.; Mundree, S.G.; Ghimire, S.R. Comparative Analysis Delineates the Transcriptional Resistance Mechanisms for Pod Borer Resistance in the Pigeonpea Wild Relative *Cajanus scarabaeoides* (L.) Thouars. *Int. J. Mol. Sci.* **2021**, *22*, 309. [\[CrossRef\]](#)
24. Heredia-Guerrero, J.A.; Benítez, J.J.; Domínguez, E.; Bayer, I.S.; Cingolani, R.; Athanassiou, A.; Heredia, A. Infrared and Raman Spectroscopic Features of Plant Cuticles: A Review. *Front. Plant Sci.* **2014**, *5*, 305. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Nogales-Bueno, J.; Baca-Bocanegra, B.; Rooney, A.; Miguel Hernández-Hierro, J.; José Heredia, F.; Byrne, H.J. Linking ATR-FTIR and Raman Features to Phenolic Extractability and Other Attributes in Grape Skin. *Talanta* **2017**, *167*, 44–50. [\[CrossRef\]](#)
26. Harscoat-Schiavo, C.; Khoualdia, B.; Savoie, R.; Hobloss, S.; Buré, C.; Samia, B.A.; Subra-Paternault, P. Extraction of Phenolics from Pomegranate Residues: Selectivity Induced by the Methods. *J. Supercrit. Fluids* **2021**, *176*, 105300. [\[CrossRef\]](#)
27. Ramakrishna, A.; Ravishankar, G.A. Influence of Abiotic Stress Signals on Secondary Metabolites in Plants. *Plant Signal. Behav.* **2011**, *6*, 1720–1731.
28. Roeber, V.M.; Bajaj, I.; Rohde, M.; Schmülling, T.; Cortleven, A. Light Acts as a Stressor and Influences Abiotic and Biotic Stress Responses in Plants. *Plant. Cell Environ.* **2021**, *44*, 645–664. [\[CrossRef\]](#)
29. Landi, M.; Tattini, M.; Gould, K.S. Multiple Functional Roles of Anthocyanins in Plant-Environment Interactions. *Environ. Exp. Bot.* **2015**, *119*, 4–17. [\[CrossRef\]](#)
30. Zivcak, M.; Brückova, K.; Sytar, O.; Brestic, M.; Olsovska, K.; Allakhverdiev, S.I. Lettuce Flavonoids Screening and Phenotyping by Chlorophyll Fluorescence Excitation Ratio. *Planta* **2017**, *245*, 1215–1229. [\[CrossRef\]](#)
31. Sawicka, B. Post-Harvest Losses of Agricultural Produce. In *Zero Hunger*; Leal Filho, W., Azul, A.M., Brandli, L., Özuyar, P.G., Wall, T., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–16. ISBN 978-3-319-69626-3.
32. Kitinoja, L.; AlHassan, H.Y. Identification of Appropriate Postharvest Technologies for Small Scale Horticultural Farmers and Marketers in Sub-Saharan Africa and South Asia -Part 1. Postharvest Losses and Quality Assessments. *Acta Hortic.* **2012**, *934*, 31–40. [\[CrossRef\]](#)
33. Zhu, H.; Chu, B.; Zhang, C.; Liu, F.; Jiang, L.; He, Y. Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-Learning Classifiers. *Sci. Rep.* **2017**, *7*, 4125. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Sobejano-Paz, V.; Mikkelsen, T.N.; Baum, A.; Mo, X.; Liu, S.; Köppl, C.J.; Johnson, M.S.; Gulyas, L.; García, M. Hyperspectral and Thermal Sensing of Stomatal Conductance, Transpiration, and Photosynthesis for Soybean and Maize under Drought. *Remote Sens.* **2020**, *12*, 3182. [\[CrossRef\]](#)
35. Fernandez-Pierna, J.A.; Baeten, V.; Michotte-Renier, A.; Cogdill, R.P.; Dardenne, P. Combination of Support Vector Machines (SVM) and near-Infrared (NIR) Imaging Spectroscopy for the Detection of Meat and Bone Meal (MBM) in Compound Feeds. *J. Chemom.* **2004**, *18*, 341–349. [\[CrossRef\]](#)
36. Cotrozzi, L.; Lorenzini, G.; Nali, C.; Pellegrini, E.; Saponaro, V.; Hoshika, Y.; Arab, L.; Rennenberg, H.; Paoletti, E. Hyperspectral Reflectance of Light-Adapted Leaves Can Predict Both Dark- and Light-Adapted Chl Fluorescence Parameters, and the Effects of Chronic Ozone Exposure on Date Palm (*Phoenix dactylifera*). *Int. J. Mol. Sci.* **2020**, *21*, 6441. [\[CrossRef\]](#)
37. Kume, A. Importance of the Green Color, Absorption Gradient, and Spectral Absorption of Chloroplasts for the Radiative Energy Balance of Leaves. *J. Plant Res.* **2018**, *131*, 501–514. [\[CrossRef\]](#)
38. Murchie, E.H.; Lawson, T. Chlorophyll Fluorescence Analysis: A Guide to Good Practice and Understanding Some New Applications. *J. Exp. Bot.* **2013**, *64*, 3983–3998. [\[CrossRef\]](#)
39. Natali, A.; Gruber, J.M.; Dietzel, L.; Stuart, M.C.A.; Van Grondelle, R.; Croce, R. Light-Harvesting Complexes (LHCs) Cluster Spontaneously in Membrane Environment Leading to Shortening of Their Excited State Lifetimes. *J. Biol. Chem.* **2016**, *291*, 16730–16739. [\[CrossRef\]](#)
40. Jia, M.; Li, D.; Colombo, R.; Wang, Y.; Wang, X.; Cheng, T.; Zhu, Y.; Yao, X.; Xu, C.; Ouer, G.; et al. Quantifying Chlorophyll Fluorescence Parameters from Hyperspectral Reflectance at the Leaf Scale under Various Nitrogen Treatment Regimes in Winter Wheat. *Remote Sens.* **2019**, *11*, 2838. [\[CrossRef\]](#)
41. Luz, R.B. Attenuated Total Reflectance Spectroscopy of Plant Leaves: A Tool for Ecological and Botanical Studies. *New Phytol.* **2006**, *172*, 305–318.
42. Tahir, H.E.; Xiaobo, Z.; Zhihua, L.; Jiyong, S.; Zhai, X.; Wang, S.; Mariod, A.A. Rapid Prediction of Phenolic Compounds and Antioxidant Activity of Sudanese Honey Using Raman and Fourier Transform Infrared (FT-IR) Spectroscopy. *Food Chem.* **2017**, *226*, 202–211. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Mérida, H.; Largo-Gosens, A.; Novo-Uzal, E.; Santiago, R.; Pomar, F.; García, P.; García-Angulo, P.; Acebes, J.L.; Álvarez, J.; Encina, A. Ectopic Lignification in Primary Cellulose-Deficient Cell Walls of Maize Cell Suspension Cultures. *J. Integr. Plant Biol.* **2015**, *57*, 357–372. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Türker-Kaya, S.; Huck, C.W. A Review of Mid-Infrared and near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. *Molecules* **2017**, *22*, 168. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Pappas, C.S.; Takidelli, C.; Tsantili, E.; Tarantilis, P.A.; Polissiou, M.G. Quantitative Determination of Anthocyanins in Three Sweet Cherry Varieties Using Diffuse Reflectance Infrared Fourier Transform Spectroscopy. *J. Food Compos. Anal.* **2011**, *24*, 17–21. [\[CrossRef\]](#)
46. Li, X.; Zhou, R.; Xu, K.; Xu, J.; Jin, J.; Fang, H.; He, Y. Rapid Determination of Chlorophyll and Pheophytin in Green Tea Using Fourier Transform Infrared Spectroscopy. *Molecules* **2018**, *23*, 1010. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Parmar, H. *Spectroscopic Study of Some Conducting Polymers*, 1st ed.; Sardar Patel University: Anand, India, 2012.
48. Schulz, H.; Baranska, M. Identification and Quantification of Valuable Plant Substances by IR and Raman Spectroscopy. *Vib. Spectrosc.* **2007**, *43*, 13–25. [\[CrossRef\]](#)
49. Pimentel-Gomes, F. *Statistics Course Experimental*, 1st ed.; FEALQ: Piracicaba, Brazil, 2009.
50. Cozzolino, D.; Vadell, A.; Ballesteros, F.; Galiotta, G.; Barlocco, N. Combining Visible and Near-Infrared Spectroscopy with Chemometrics to Trace Muscles from an Autochthonous Breed of Pig Produced in Uruguay: A Feasibility Study. *Anal. Bioanal. Chem.* **2006**, *385*, 931–936. [\[CrossRef\]](#)
51. Arena, C.; Tsonev, T.; Doneva, D.; DeMicco, V.; Michelozzi, M.; Brunetti, C.; Centritto, M.; Fineschi, S.; Velikova, V.; Loreto, F. The Effect of Light Quality on Growth, Photosynthesis, Leaf Anatomy and Volatile Isoprenoids of a Monoterpene-Emitting Herbaceous Species (*Solanum lycopersicum* L.) and an Isoprene-Emitting Tree (*Platanus orientalis* L.). *Environ. Exp. Bot.* **2016**, *130*, 122–132. [\[CrossRef\]](#)
52. Llorach, R.; Martínez-Sánchez, A.; Tomás-Barberán, F.A.; Gil, M.I.; Ferreres, F. Characterisation of Polyphenols and Antioxidant Properties of Five Lettuce Varieties and Escarole. *Food Chem.* **2008**, *108*, 1028–1038. [\[CrossRef\]](#)
53. Steidle Neto, A.J.; de Oliveira Moura, L.; de Carvalho Lopes, D.; de Almeida Carlos, L.; Martins, L.M.; De Castro Louback Ferraz, L. Non-Destructive Prediction of Pigment Content in Lettuce Based on Visible-NIR Spectroscopy. *J. Sci. Food Agric.* **2016**, *97*, 2015–2022. [\[CrossRef\]](#)
54. de Oliveira Moura, L.; de Carvalho Lopes, D.; Steidle-Neto, A.J.; de Castro Louback Ferraz, L.; de Almeida Carlos, L.; Martins, L.M. Evaluation of Techniques for Automatic Classification of Lettuce Based on Spectral Reflectance. *Food Anal. Methods* **2016**, *9*, 1799–1806. [\[CrossRef\]](#)
55. Boshkovski, B.; Doupis, G.; Zapolska, A.; Kalaitzidis, C.; Koubouris, G. Hyperspectral Imagery Detects Water Deficit and Salinity Effects on Photosynthesis and Antioxidant Enzyme Activity of Three Greek Olive Varieties. *Sustainability* **2022**, *14*, 1432. [\[CrossRef\]](#)
56. Montes, J.M.; Technow, F.; Bohlinger, B.; Becker, K. Grain Quality Determination by Means of near Infrared Spectroscopy in *Jatropha curcas* L. *Ind. Crops Prod.* **2013**, *43*, 301–305. [\[CrossRef\]](#)

57. Coast, O.; Shah, S.; Ivakov, A.; Gaju, O.; Wilson, P.B.; Posch, B.C.; Bryant, C.J.; Negrini, A.C.A.; Evans, J.R.; Condon, A.G.; et al. Predicting Dark Respiration Rates of Wheat Leaves from Hyperspectral Reflectance. *Plant Cell Environ.* **2019**, *42*, 2133–2150. [\[CrossRef\]](#)
58. Moriwaki, T.; Falcioni, R.; Tanaka, F.A.O.; Cardoso, K.A.K.; Souza, L.A.; Benedito, E.; Nanni, M.R.; Bonato, C.M.; Antunes, W.C. Nitrogen-Improved Photosynthesis Quantum Yield Is Driven by Increased Thylakoid Density, Enhancing Green Light Absorption. *Plant Sci.* **2019**, *278*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Calviño-Cancela, M.; Martín-Herrero, J. Spectral Discrimination of Vegetation Classes in Ice-Free Areas of Antarctica. *Remote Sens.* **2016**, *8*, 856. [\[CrossRef\]](#)
60. Horikawa, Y.; Mizuno-Tazuru, S.; Sugiyama, J. Near-Infrared Spectroscopy as a Potential Method for Identification of Anatomically Similar Japanese Diploxylons. *J. Wood Sci.* **2015**, *61*, 251–261. [\[CrossRef\]](#)
61. Damerum, A.; Selmes, S.L.; Biggi, G.F.; Clarkson, G.J.J.; Rothwell, S.D.; Truco, M.J.; Michelmores, R.W.; Hancock, R.D.; Shellcock, C.; Chapman, M.A.; et al. Elucidating the Genetic Basis of Antioxidant Status in Lettuce (*Lactuca sativa*). *Hortic. Res.* **2015**, *2*, 15055. [\[CrossRef\]](#)
62. Mihoubi, W.; Sahli, E.; Gargouri, A.; Amiel, C. FTIR Spectroscopy of Whole Cells for the Monitoring of Yeast Apoptosis Mediated by P53 Over-Expression and Its Suppression by *Nigella sativa* Extracts. *PLoS ONE* **2017**, *12*, e0180680. [\[CrossRef\]](#)
63. Baker, N.R.; Harbinson, J.; Kramer, D.M. Determining the Limitations and Regulation of Photosynthetic Energy Transduction in Leaves. *Plant Cell Environ.* **2007**, *30*, 1107–1125. [\[CrossRef\]](#)
64. Albu, M.G.; Ghica, M.V.; Giurginca, M.; Trandafir, V.; Popa, L.; Cotrut, C. Spectral Characteristics and Antioxidant Properties of Tannic Acid Immobilized on Collagen Drug-Delivery Systems. *Rev. Chim.* **2009**, *60*, 666–672.
65. Chen, S.; Zhang, F.; Ning, J.; Liu, X.; Zhang, Z.; Yang, S. Predicting the Anthocyanin Content of Wine Grapes by NIR Hyperspectral Imaging. *Food Chem.* **2015**, *172*, 788–793. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Kokaly, R.F.; Skidmore, A.K. Plant Phenolics and Absorption Features in Vegetation Reflectance Spectra near 1.66 μm . *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 55–83. [\[CrossRef\]](#)
67. Willick, I.R.; Lahlali, R.; Vijayan, P.; Muir, D.; Karunakaran, C. Wheat Flag Leaf Epicuticular Wax Morphology and Composition in Response to Moderate Drought Stress Are Revealed by SEM, FTIR-ATR and Synchrotron X-ray Spectroscopy. *Physiol. Plant.* **2018**, *162*, 316–332. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Lang, C.; Costa, F.R.C.; Camargo, J.L.C.; Durgante, F.M.; Vicentini, A. Near Infrared Spectroscopy Facilitates Rapid Identification of Both Young and Mature Amazonian Tree Species. *PLoS ONE* **2015**, *10*, e0134521. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Colthup, N.B. *Introduction to Infrared and Raman Spectroscopy*, 3rd ed.; Colthup, L.H., Daly, S.E.W., Eds.; Academic Press: London, UK, 1990.
70. Jensen, J.S.; Egebo, M.; Meyer, A.S. Identification of Spectral Regions for the Quantification of Red Wine Tannins with Fourier Transform Mid-Infrared Spectroscopy. *J. Agric. Food Chem.* **2008**, *56*, 3493–3499. [\[CrossRef\]](#)
71. Rodrigues, M.; de Oliveira, R.B.; Dos Santos, G.L.A.A.; de Oliveira, K.M.; Reis, A.S.; Furlanetto, R.H.; Júnior, L.A.Y.B.; Coelho, F.S.; Nanni, M.R. Rapid Quantification of Alkaloids, Sugar and Yield of Tobacco (*Nicotiana tabacum* L.) Varieties by Using VIS–NIR–SWIR Spectroradiometry. *Spectrochim. Acta Part A Mol. Biomol.* **2022**, *274*, 121082. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Barnes, M.L.; Breshears, D.D.; Law, D.J.; van Leeuwen, W.J.D.; Monson, R.K.; Fojtik, A.C.; Barron-Gafford, G.A.; Moore, D.J.P. Beyond Greenness: Detecting Temporal Changes in Photosynthetic Capacity with Hyperspectral Reflectance Data. *PLoS ONE* **2017**, *12*, e0189539. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Jin, J.; Wang, Q. Selection of Informative Spectral Bands for PLS Models to Estimate Foliar Chlorophyll Content Using Hyperspectral Reflectance. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3064–3072. [\[CrossRef\]](#)
74. Özgenç, Ö.; Durmaz, S.; Hakki, I.; Eksi-kocak, H. Determination of Chemical Changes in Heat-Treated Wood Using ATR-FTIR and FT Raman Spectrometry. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2017**, *171*, 395–400. [\[CrossRef\]](#)
75. Kubicki, J.D.; Yang, H.; Sawada, D.; O'Neill, H.; Oehme, D.; Cosgrove, D. The Shape of Native Plant Cellulose Microfibrils. *Sci. Rep.* **2018**, *8*, 4–11. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Sexton, T.; Sankaran, S.; Cousins, A.B. Predicting Photosynthetic Capacity in Tobacco Using Shortwave Infrared Spectral Reflectance. *J. Exp. Bot.* **2021**, *72*, 4373–4383. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Zeng, J.; Ping, W.; Sanaeifar, A.; Xu, X.; Luo, W.; Sha, J.; Huang, Z.; Huang, Y.; Liu, X.; Zhan, B.; et al. Quantitative Visualization of Photosynthetic Pigments in Tea Leaves Based on Raman Spectroscopy and Calibration Model Transfer. *Plant Methods* **2021**, *17*, 4. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Straková, P.; Larmola, T.; Andrés, J.; Ilola, N.; Launiainen, P.; Edwards, K.; Minkkinen, K.; Laiho, R. Quantification of Plant Root Species Composition in Peatlands Using FTIR Spectroscopy. *Front. Plant Sci.* **2020**, *11*, 597. [\[CrossRef\]](#)
79. Prats-Mateu, B.; Felhofer, M.; de Juan, A.; Gierlinger, N. Multivariate Unmixing Approaches on Raman Images of Plant Cell Walls: New Insights or Overinterpretation of Results? *Plant Methods* **2018**, *14*, 52. [\[CrossRef\]](#)
80. Gierlinger, N.; Keplinger, T.; Harrington, M. Imaging of Plant Cell Walls by Confocal Raman Microscopy. *Nat. Protoc.* **2012**, *7*, 1694–1708. [\[CrossRef\]](#)
81. Braga, P.; Crusiol, L.G.T.; Nanni, M.R.; Caranhato, A.L.H.; Fuhrmann, M.B.; Nepomuceno, A.L.; Neumaier, N.; Farias, J.R.B.; Koltun, A.; Gonçalves, L.S.A.; et al. Vegetation Indices and NIR-SWIR Spectral Bands as a Phenotyping Tool for Water Status Determination in Soybean. *Precis. Agric.* **2021**, *22*, 249–266. [\[CrossRef\]](#)

82. Ling, B.; Goodin, D.G.; Raynor, E.J.; Joern, A. Hyperspectral Analysis of Leaf Pigments and Nutritional Elements in Tallgrass Prairie Vegetation. *Front. Plant Sci.* **2019**, *10*, 142. [[CrossRef](#)]
83. Atzberger, C. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sens.* **2013**, *5*, 949–981. [[CrossRef](#)]
84. Nanni, M.R.; Cezar, E.; da Silva Junior, C.A.; Silva, G.F.C.; da Silva Gualberto, A.A. Partial Least Squares Regression (PLSR) Associated with Spectral Response to Predict Soil Attributes in Transitional Lithologies. *Arch. Agron. Soil Sci.* **2018**, *64*, 682–695. [[CrossRef](#)]
85. El-Hendawy, S.; Al-Suhaibani, N.; Mubushar, M.; Tahir, M.U.; Marey, S.; Refay, Y.; Tola, E. Combining Hyperspectral Reflectance and Multivariate Regression Models to Estimate Plant Biomass of Advanced Spring Wheat Lines in Diverse Phenological Stages under Salinity Conditions. *Appl. Sci.* **2022**, *12*, 1983. [[CrossRef](#)]
86. da Silva Junior, C.A.; Nanni, M.R.; Shakir, M.; Teodoro, P.E.; de Oliveira-Júnior, J.F.; Cezar, E.; de Gois, G.; Lima, M.; Wojciechowski, J.C.; Shiratsuchi, L.S. Soybean Varieties Discrimination Using Non-Imaging Hyperspectral Sensor. *Infrared Phys. Technol.* **2018**, *89*, 338–350. [[CrossRef](#)]
87. Zar, J.H. *Biostatistical Analysis*, 5th ed.; Pearson Education: Upper Saddle River, NJ, USA, 2010; ISBN 0-13-100846-3.
88. Gitelson, A.; Chivkunova, O.; Zhigalova, T.; Solovchenko, A. In Situ Optical Properties of Foliar Flavonoids: Implication for Non-Destructive Estimation of Flavonoid Content. *J. Plant Physiol.* **2017**, *218*, 258–264. [[CrossRef](#)] [[PubMed](#)]
89. Fan, K.; Li, F.; Chen, X.; Li, Z.; Mulla, D.J. Nitrogen Balance Index Prediction of Winter Wheat by Canopy Hyperspectral Transformation and Machine Learning. *Remote Sens.* **2022**, *14*, 3504. [[CrossRef](#)]