horticulturae

MDPI

Article

Estimation of Apple Leaf Nitrogen Concentration Using Hyperspectral Imaging-Based Wavelength Selection and Machine Learning

Sihyeong Jang [1], Jeomhwa Han [1,*], Junggun Cho [1], Jaehoon Jung [1], Seulki Lee [1], Dongyong Lee [1] and Jingook Kim [2]

[1] Fruit Research Division, National Institute of Horticultural and Herbal Science, Wanju 55365, Republic of Korea; jangsh6968@korea.kr (S.J.); jgcho@korea.kr (J.C.); jhyskok028@korea.kr (J.J.); lsk0729@korea.kr (S.L.); dongle1013@korea.kr (D.L.)
[2] Institute of Agriculture and Life Sciences, Gyeongsang National University, Jinju 52828, Republic of Korea; jgkim119@gnu.ac.kr
* Correspondence: najuflower@korea.kr

Abstract: In apple cultivation, the total nitrogen content is an important indicator of plant growth, fruit quality, and yield. Timely monitoring of growth becomes imperative, since an imbalance, either in deficiency or excess nitrogen, can result in physiological disorders, adversely impacting both the quantity and quality of fruit. Leaf nitrogen content can be determined using simple chlorophyll meters or destructive testing; however, these methods are time-consuming. However, by employing spectral imaging technology, it is possible to swiftly predict leaf nitrogen content. This study estimated the total nitrogen content in apple trees via hyperspectral imaging and machine learning-based regression analysis (partial least-squares regression (PLSR), support vector regression (SVR), and eXtreme gradient boosting regression (XGBoost). Additionally, to reduce computational costs and improve reproducibility, spectral binning was divided into three stages (4, 8, and 16 bins), and models were compared with a 2-binning estimation model. The analysis focused on green, red, red edge, and near-infrared (NIR) spectra, with 5–10 selected wavelengths, and the SVR-based prediction model showed a similar or greater performance to that of the full spectrum. At 4- and 8-binning, the selected wavelengths were similar to those at 2-binning, maintaining similar prediction model performance. However, at 16 bp, the performance of the prediction model decreased owing to spectral data loss, leading to a significant reduction in wavelengths for nitrogen content estimation. These results can support informed nitrogen fertilization decisions, enabling precise, real-time monitoring of nitrogen content for enhanced plant growth, fruit quality, and yield in apple trees. Additionally, the selected wavelengths can be considered in the development of new types of multispectral sensors.

Keywords: leaf nitrogen concentration; hyperspectral imaging; apple tree; machine learning; variable selection

Citation: Jang, S.; Han, J.; Cho, J.; Jung, J.; Lee, S.; Lee, D.; Kim, J. Estimation of Apple Leaf Nitrogen Concentration Using Hyperspectral Imaging-Based Wavelength Selection and Machine Learning. *Horticulturae* **2024**, *10*, 35. https://doi.org/10.3390/horticulturae10010035

Academic Editor: Caixi Zhang

Received: 20 November 2023
Revised: 22 December 2023
Accepted: 27 December 2023
Published: 28 December 2023

1. Introduction

Apple (*Malus pumila* Mill.) is a perennial crop belonging to the Rosaceae family, and careful selection of suitable cultivation sites based on geographical and environmental conditions is needed because of its long-term cultivation in a single location. Furthermore, in the cultivation of fruit trees, it is crucial to supply the right amount of nutrients during key stages. During these stages, nitrogen is the most critical factor influencing both vegetative growth and the quality and quantity of fruit. Insufficient nitrogen weakens plant growth, resulting in poor fruit development and a significant decrease in yield and quality [1]. In contrast, an excess of nitrogen causes assimilated nutrients to be consumed primarily for the growth of stems and leaves, causing the plant to grow excessively and leading to fruit disorders such as bitter pits or corky tissue [2]. As the fruit size increases, coloration

becomes inadequate and maturation is delayed, resulting in rapid quality deterioration during storage. Furthermore, prolonged vegetative growth leads to a decrease in nutrient accumulation during storage and delays plant maturation, increasing susceptibility to frost damage. Therefore, timely fertilization is crucial for effective cultivation management [3–6].

Remote sensing technology, which observes characteristics and phenomena using sensors that are mounted on platforms such as satellites and aircraft without physical contact with a target, is gaining attention. This technology utilizes reflected or radiated electromagnetic energy to observe desired subjects. Recently, advancements in drone technology, satellites, and high-resolution sensor technology, coupled with the integration of big data and AI, have been utilized in various fields, such as geology, marine science, defense, and the environment, where on-site surveys are challenging, not only for urban and territorial planning [7]. In agriculture, various methods, including real-time monitoring of crop nutrients [8], monitoring of moisture levels [9,10], disease and pest diagnosis [11], crop yield assessment [12–14], cultivation area estimation [15], early prediction of harvest timing, and forecasting of harvest quantity and quality [16,17], are actively utilized. Such applications signify a significant stride toward enhancing agricultural productivity, efficient resource management, and sustainability. By enabling predictive modeling and precise agricultural management, agriculture can progress toward more sustainable and efficient farming activities, contributing to income growth and environmental preservation. A typical RGB sensor covering the visible spectrum represents information for only three to ten wavelengths, whereas multispectral sensors, including near-infrared sensors, provide information for the same range of wavelengths. In contrast, hyperspectral sensors can capture information for as few as several dozen to several hundred wavelengths. However, the increasing size of spectral data leads to higher costs, complex data processing, and challenges such as signal-to-noise ratio (SNR) degradation [18]. Therefore, postacquisition preprocessing and minimization of data loss are necessary for effective data handling. Regression analysis is a technique that utilizes one or more independent variables (x) to explain the dependent variable of interest using a mathematical function. The types of regression analysis include linear regressions, such as simple and multiple linear regressions, and nonlinear regression analyses, such as tree-based and polynomial regressions. Linear regression has the advantages of simple computations, easy model interpretation, and rapid analysis [19]; however, linear regression is sensitive to outliers and may result in decreased model accuracy when the relationships between variables are not linear. Addressing this limitation increases model accuracy by accounting for nonlinear relationships through higher-order terms, compensating for the disadvantages of linear regression [20]. However, as the number of interaction terms increases, the calculation becomes more complex and requires more time, and a higher bias can lead to overfitting (bias–variance tradeoff). Therefore, it is challenging to definitively state which regression analysis is better based on the independent variables concerning the subject of analysis. Finally, it is important to compare the performance of the models, calculated using both linear and nonlinear regression analyses and to select a model with high reproducibility. This approach ensures the selection of a reliable model through a performance comparison.

Based on the described developments, focused research is underway to apply similar methodologies to orchard cultivation. Studies utilizing hyperspectral imaging have been conducted to predict carbohydrate content, which is associated with fruit quality, achieving a high prediction performance of over 75% [21]. Additionally, research focused on predicting potassium levels using a combination of various vegetation indices derived from hyperspectral imaging has been carried out. Among the diverse vegetation indices, the combination of red edge and blue wavelengths in the derived DVI (Difference Vegetation Index) exhibited the highest performance, with an R^2 value of 0.899 [22].

In this study, we developed a model to predict the leaf nitrogen content of apple trees via hyperspectral imaging by (1) conducting regression analysis (partial least-squares regression, support vector regression, and eXtreme gradient boosting regression) using both the full spectrum and selected wavelengths, followed by a comparison of the eval-

uation performances; and (2) reducing the spectral resolution through spectral binning and performing regression analysis using both the full spectrum and selected wavelengths, followed by a subsequent comparison of the evaluation performance.

2. Materials and Methods

This study was conducted over two years, from 2021 to 2022, at the experimental field of the National Institute of Horticultural & Herbal Science located in Wanju-gun, Jeollabuk-do, Republic of Korea (35°49′42.8″ N, 127°01′52.9″ E). Two-year-old nursery stocks of 'Hongro/M.9' were used for the experiment, and they were subsequently grafted onto potted rootstocks. The potting mixture was prepared by mixing horticultural soil, loess soil, and perlite at a ratio of 5:4:1. The plants were planted at intervals of 3 m × 2 m, with each treatment plot accommodating 38 trees. Nitrogen fertilization was carried out by dividing ammonium nitrate (NH_4NO_3) into fertilizer amounts of 171 g/year, 43 g/year, and 0 g/year for each plot, after which the fertilizer was diluted in 2 L of water.

2.1. Hyperspectral Data

The hyperspectral imaging system was composed of a hyperspectral sensor (Fx10, Specim Spectral Imaging Ltd., Finland) that operates in the wavelength range of 400–1000 nm, with 224 channels, a field of view of 38°, and a spectral resolution of 5.5 nm based on the 2-binning line scan method. In addition, the system included a rotator mounted at the bottom (RS10, Specim Spectral Imaging Ltd., Oulu, Finland) and a reference board with 99% reflectance (Spectralon, Labsphere, Inc., North Sutton, NH, USA) to correct for variations in sunlight. In addition, a rotator (RS10; Specim Spectral Imaging Ltd., Oulu, Finland) attached to the bottom and a reference board (Spectralon, Labsphere, Inc., North Sutton, NH, USA) with 99% reflectance to compensate for solar variability were used in the system setup. To prevent image distortion during hyperspectral imaging, the rotation radius of the rotator was set to 30°. Prior to capturing the main image, a dark current image was acquired to eliminate noise caused by the heat generated during sensor operation. The reference board was then placed beside the subject, and images were acquired using dedicated imaging software (Lumo Scanner, Specim Spectral Imaging Ltd., Oulu, Finland). The acquired images were processed using hyperspectral image processing software (ENVI 5.3, Exelis Visual Information Solutions, Boulder, CO, USA). Before image processing, the images were subjected to a preprocessing phase that included dark current correction and radiometric correction. Normalized images were subsequently applied to a vegetation index, specifically the NDVI-GNDVI, as described in Equation (1), to separate the canopy area from the background.

$$\text{NDVI} - \text{GNDVI} = \left(\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}\right) - \left(\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}\right) \quad (1)$$

The images were converted into vegetation indices utilizing density slices to separate the canopy from the background based on a designated threshold. Subsequently, the canopy section was designated as the region of interest, and the reflectance values were extracted (Figure 1).

As depicted in Figure 2, in shadowed regions, where light absorption and reflection are minimal, noise occurs. When comparing spectral curves between areas with shadows and those without, reflectance values exhibit a difference of approximately 0.1 to 0.5 or more depending on the wavelength. Since such differences can lead to data errors in predicting nitrogen content, histograms were generated for each wavelength. Subsequently, threshold values were set to minimize the impact of shadows and delineate the regions. To reduce the spectral resolution of the extracted hyperspectral data, the original 2-binning images were partitioned into 4-binning, 8-binning, and 16-binning.

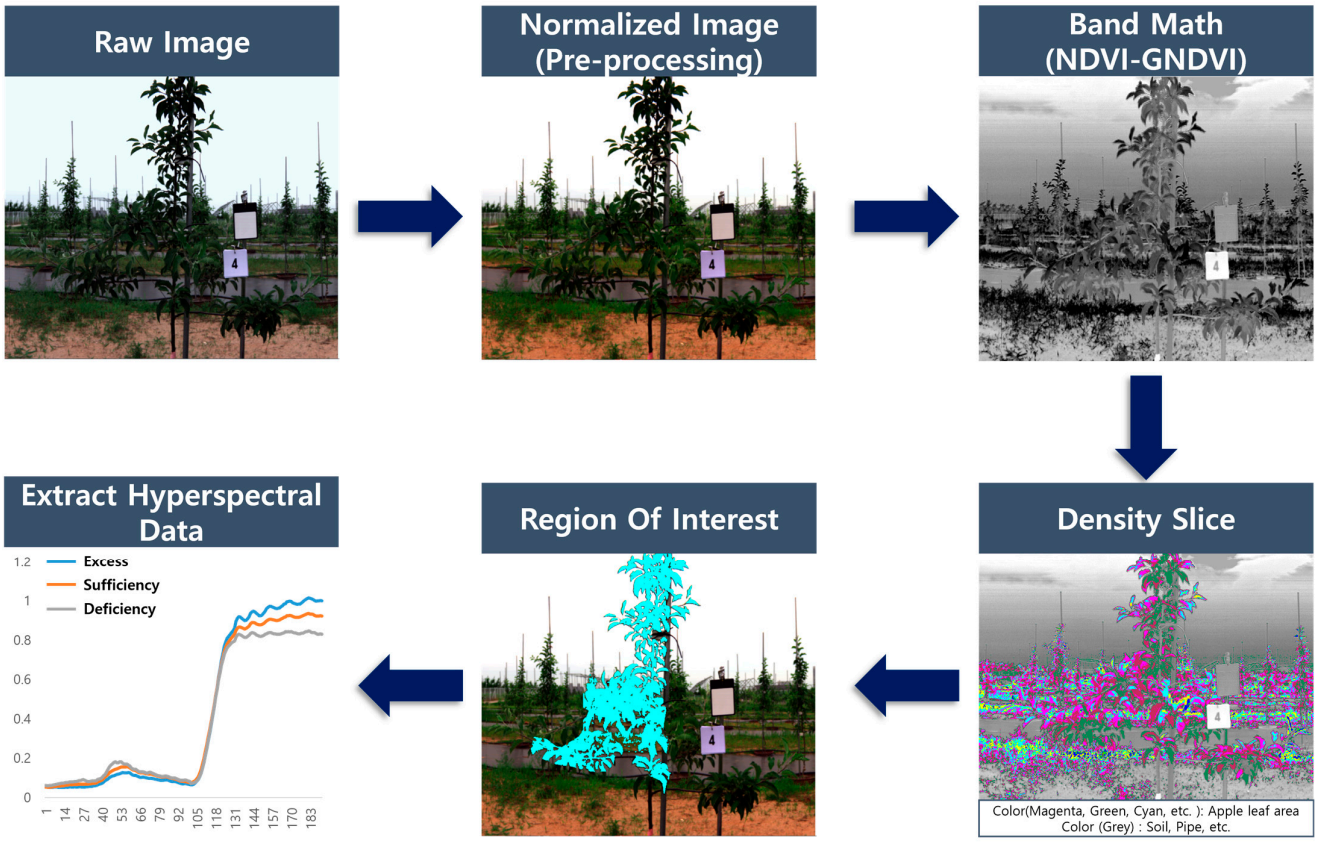

Figure 1. Description of hyperspectral imaging processing.

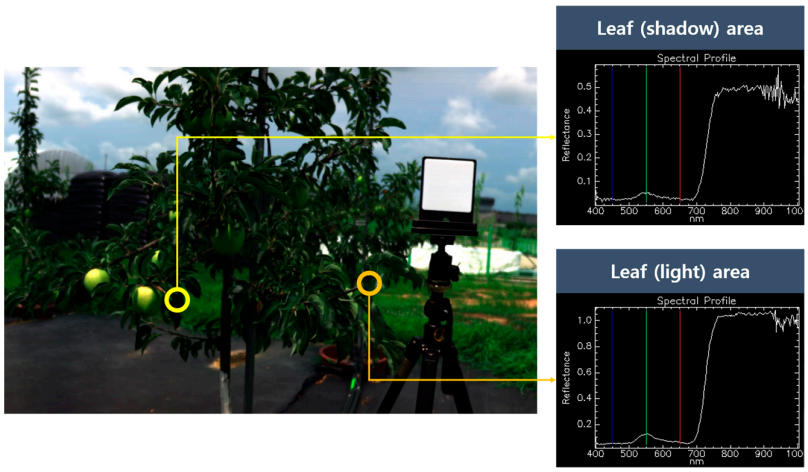

Figure 2. Comparison of spectral curves (Red line: 650 nm, Green line: 550 nm, Blue line: 450 nm) for apple leaves with varying degrees of light.

2.2. Apple Leaf Nitrogen Content Measurement

At each time point, a total of 21 leaf samples were collected, with seven samples from each treatment group. Nitrogen content data were acquired for the leaves, with a focus on mature leaves, and a total of 10 leaves were collected. The leaf nitrogen content was measured in accordance with the soil and plant analysis methods stipulated by the National Institute of Agricultural Sciences (2000). The collected leaves were dried in a dryer at 60 °C for 5 days (60 h). A 1 g sample of the dried material was digested with a mixture of nitrogen and perchloric acid at a ratio of 85:15, amounting to 10 mL. Upon completion of digestion, the solution was allowed to cool to room temperature. The residual liquid in

the container was then rinsed with distilled water and filtered through a volumetric flask. The leaf nitrogen content was subsequently measured using a carbon/nitrogen elemental analyzer (NL/Primacs SNC-100, Skalar Analytical B.V., Breda, The Netherlands).

2.3. Hyperspectral Data Transformations

Various issues arise when capturing hyperspectral images in open fields owing to differences in environmental factors. These include changes in atmospheric conditions, uneven lighting sources [23,24], and noise caused by the heat of the sensor itself [25]. Therefore, accurate analysis of hyperspectral images obtained in open fields requires preprocessing, which involves setting and optimizing the hyperspectral sensor and imaging equipment according to the conditions. The first derivative method is a preprocessing technique used to reduce the noise caused by light. This involves calculating the rate of change between data points by differentiating the raw data. This method extracts features through the gradient of the data rather than the raw reflectance values, thereby reducing the noise caused by environmental fluctuations and improving the accuracy of the data. Additionally, the Savitzky–Golay filter is a method for smoothing data at regular intervals [26]. The output value at each data point is determined by finding, through least-squares fitting, the polynomial of order k that best fits the surrounding points. This method is commonly used, because it reduces noise due to various light conditions and atmospheric states while maintaining spectral characteristics, making it an effective preprocessing technique [27].

2.4. Variable Selection Method

Hyperspectral data, which contain abundant continuous spectral information, complicate computational analysis [28] and can lead to overfitting owing to unnecessary variables [29], consequently diminishing the performance of regression models [30]. To address these issues, methods have been developed to eliminate variables with little relevance to the dependent variable among numerous independent variables or to find combinations of predictive variables. These methods involve combining meaningful information to extract new features, thereby removing unnecessary information or noise and extracting important information for analysis. Among the variable selection methods, competitive adaptive reweighted sampling (CARS) uses PLS-based regression coefficients as criteria to evaluate the importance of variables. Subsets are randomly generated via Monte Carlo sampling, and N variables are selected through competition, followed by wavelength selection based on an exponentially decreasing function and adaptive reweighted sampling, with the lowest root-mean-square error (RMSE) chosen through cross-validation [31]. The successive projections algorithm (SPA) employs a forward selection approach, constructs subsets of variables with minimal collinearity, calculates the distance between the variables and their orthogonal projections, and selects those with the maximum orthogonal distance. Selection was based on the lowest RMSECV in multiple linear regression (MLR) [32]. The random frog (R-Frog) model, which is based on partial least-squares regression (PLSR), randomly selects variable sets and calculates selection probabilities through repeated iterations using the reversible jump Markov chain Monte Carlo algorithm. Wavelengths with higher selection probabilities were chosen as feature variables.

2.5. Regression Analysis Based on Machine Learning Models

PLSR develops models using least-squares regression between dependent variables by creating latent variables that maximize the covariance between a linear combination of independent and dependent variables. This approach addresses the issue of low regression coefficient estimates owing to the high correlations among the independent variables. Gradient boosting regression, an analysis method utilizing the boosting technique within ensemble models, progressively adds three models that predict and calculate residuals (the differences between predicted and actual values), thereby reducing errors. However, this process can be time-consuming and can be mitigated by extreme gradient boosting

(XGBoost) regression analysis. Unlike gradient boosting regression, XGBoost supports parallel and distributed processing, allowing it to handle large datasets rapidly. It learns efficiently and concisely through pruning and can use various objective functions to reduce time. A support vector machine (SVM) is an algorithm that maps data to a high-dimensional space and determines a decision boundary by maximizing the margin, which represents the distance between the decision boundary and data points (MATLAB R2023a, MathWorks, Natick, MA, USA). The performances of these regression models were validated using 10-fold cross-validation and evaluated based on the coefficient of determination (R^2) and RMSE. R^2 is a statistical metric in regression analysis that indicates how effectively a model explains the variability of the dependent variable. A value close to one suggests that the model effectively explains the variability of the dependent variable, while a value close to zero indicates that the model fails to adequately explain the variability of the dependent variable. The RMSE is a metric used to measure the difference between predicted and actual values. The method involves squaring the differences between each predicted value and its corresponding actual value, calculating the mean of these squared differences, and then taking the square root of that mean. A lower RMSE indicates that the model's predictions are closer to the actual values, signifying higher model performance. Figure 3 presents a flowchart summarizing the processes, including image preprocessing and analysis methods, conducted in this study.

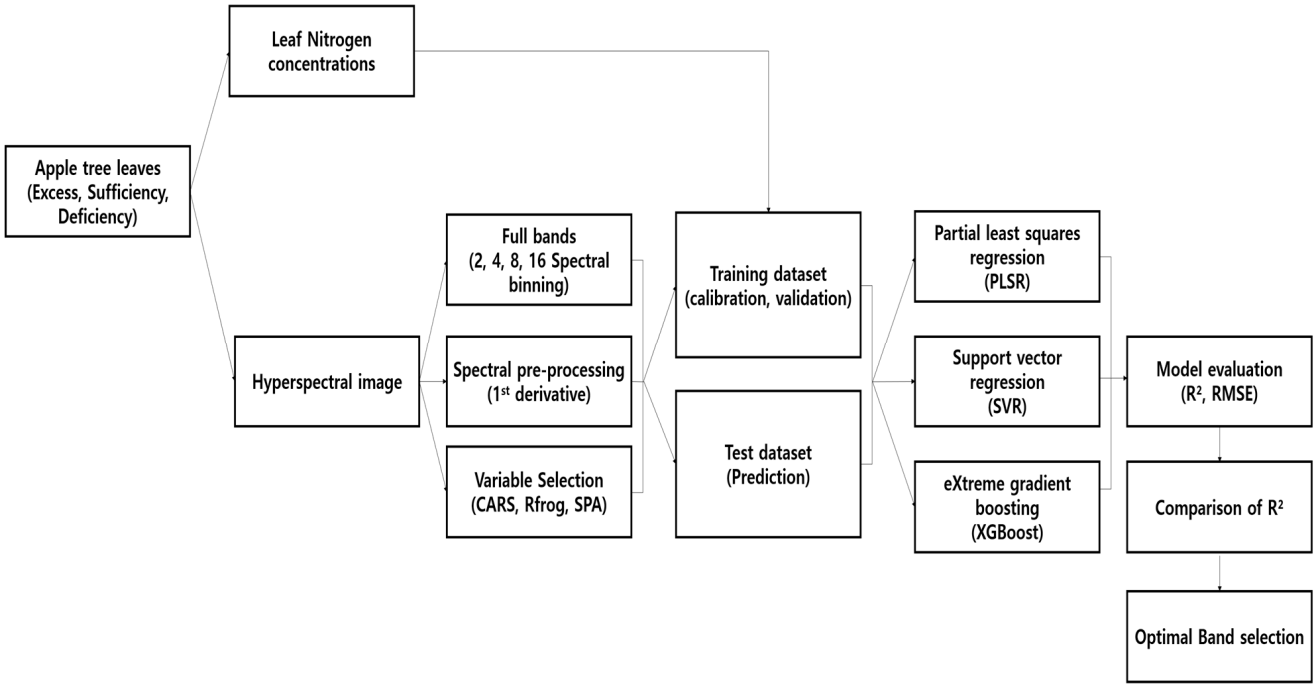

Figure 3. Flowchart for estimating the apple leaf nitrogen concentration using hyperspectral imaging.

3. Results

3.1. Nitrogen Content

Table 1 presents a comparative analysis of the leaf nitrogen content based on nitrogen application rates for 2021 and 2022. In 2021, the excessive treatment, adequate, and insufficient groups exhibited nitrogen content ranges of 2.58–3.67%, 1.48–2.77%, and 1.26–2.51%, respectively. Except for the first period, statistically significant differences were observed in the second to the seventh periods between the treatment groups. Although the eighth to tenth periods did not significantly differ between the excessive and adequate groups, the insufficient group exhibited notable differences. In 2022, the excessive treatment, adequate, and insufficient groups exhibited nitrogen content ranges of 2.48–3.32%, 2.15–2.84%, and 1.85–2.51%, respectively. As the growth period progressed, there was a decrease, irre-

spective of the treatment group. Except for the seventh period (16 August), statistically significant differences were observed among the treatment groups. For the seventh period, it was presumed that a significant amount of rainfall before conducting the growth assessment did not affect nitrogen fertilization. A comparison of the nitrogen content between 2021 and 2022 revealed that these differences arose because of differences in growth year and between the first and second years, which affected the root growth and consequently inhibited nitrogen absorption [33].

Table 1. Analysis of variance (ANOVA) of the apple leaf nitrogen concentration according to growth stage.

	2021 (*n* = 147)			2022 (*n* = 196)		
	Excess (%)	**Sufficient (%)**	**Deficiency (%)**	**Excess (%)**	**Sufficient (%)**	**Deficiency (%)**
1st	2.62 ± 0.03 a	2.61 ± 0.15 a	2.51 ± 0.28 a	3.31 ± 0.27 a	2.48 ± 0.50 b	2.08 ± 0.20 c
2nd	2.95 ± 0.15 a	2.71 ± 0.08 b	2.41 ± 0.12 c	2.86 ± 0.15 a	2.38 ± 0.19 b	2.05 ± 0.13 c
3rd	2.58 ± 0.28 a	1.48 ± 0.04 b	1.26 ± 0.06 c	3.04 ± 0.14 a	2.45 ± 0.14 b	2.12 ± 0.43 c
4th	2.61 ± 0.26 a	1.71 ± 0.11 b	1.33 ± 0.12 c	3.02 ± 0.33 a	2.35 ± 0.22 b	2.04 ± 0.44 c
5th	3.11 ± 0.24 a	2.18 ± 0.19 b	1.50 ± 0.20 c	3.32 ± 0.24 a	2.84 ± 1.71 b	2.51 ± 0.18 c
6th	3.67 ± 0.43 a	2.48 ± 0.21 b	1.61 ± 0.23 c	2.99 ± 0.11 a	2.61 ± 0.09 b	2.29 ± 0.22 c
7th	3.45 ± 0.36 a	2.56 ± 0.37 b	1.84 ± 0.18 c	2.81 ± 0.24 a	2.52 ± 0.61 a	1.97 ± 0.18 b
8th	3.38 ± 0.45 a	2.54 ± 0.20 a	1.49 ± 0.20 b	2.62 ± 0.15 a	2.31 ± 0.17 b	1.99 ± 0.26 c
9th	3.54 ± 0.64 a	2.77 ± 0.16 a	1.61 ± 0.20 b	2.87 ± 0.40 a	2.17 ± 0.38 b	1.98 ± 0.17 b
10th	3.29 ± 0.70 a	2.65 ± 0.21 a	1.68 ± 0.28 b	2.48 ± 0.34 a	2.15 ± 0.15 b	1.85 ± 0.21 c

Lowercase letters indicate significant differences (p-value < 0.05) between different nitrogen fertilization.

3.2. Spectral Characteristics

Figure 4 shows the spectral curves of both the raw and first derivative postprocessing data. A comparison of the raw data across treatment groups revealed that the reflectance in the visible light spectrum was greater in the insufficient treatment group than in the excessive treatment group. This culminated in a peak within the green wavelength range of 530–570 nm. In the nonvisible spectrum, the reflectance was lower in the near-infrared region after reaching 800 nm. These observations can be attributed to the predominant absorption of chlorophyll in the visible spectrum, while the absorption at >700 nm in the nonvisible spectrum tends to occur, resulting in significant differences in reflectance depending on the vegetative state. For the first derivative, differences were detected at green wavelengths of 520 nm and 560 nm, at a red edge wavelength of 750 nm, and at a near-infrared wavelength of 800 nm.

3.3. The Development of a Leaf Nitrogen Content Prediction Model Based on a Full Spectrum

Table 2 presents the results of the leaf nitrogen content prediction model using both the raw and preprocessed spectroscopic data. For the raw data prediction model, PLSR exhibited an R^2 of 0.619 and an RMSE of 0.261%, whereas SVR showed an R^2 of 0.682 and an RMSE of 0.245%. XGBoost demonstrated the best performance, with an R^2 of 0.756 and an RMSE of 0.216%. For the first derivative, the SVR exhibited the highest performance, with an R^2 of 0.739 and an RMSE of 0.224%. However, whereas the XGBoost calibration model had an R^2 value of 0.99, its prediction model had an R^2 value of 0.59, indicating potential overfitting.

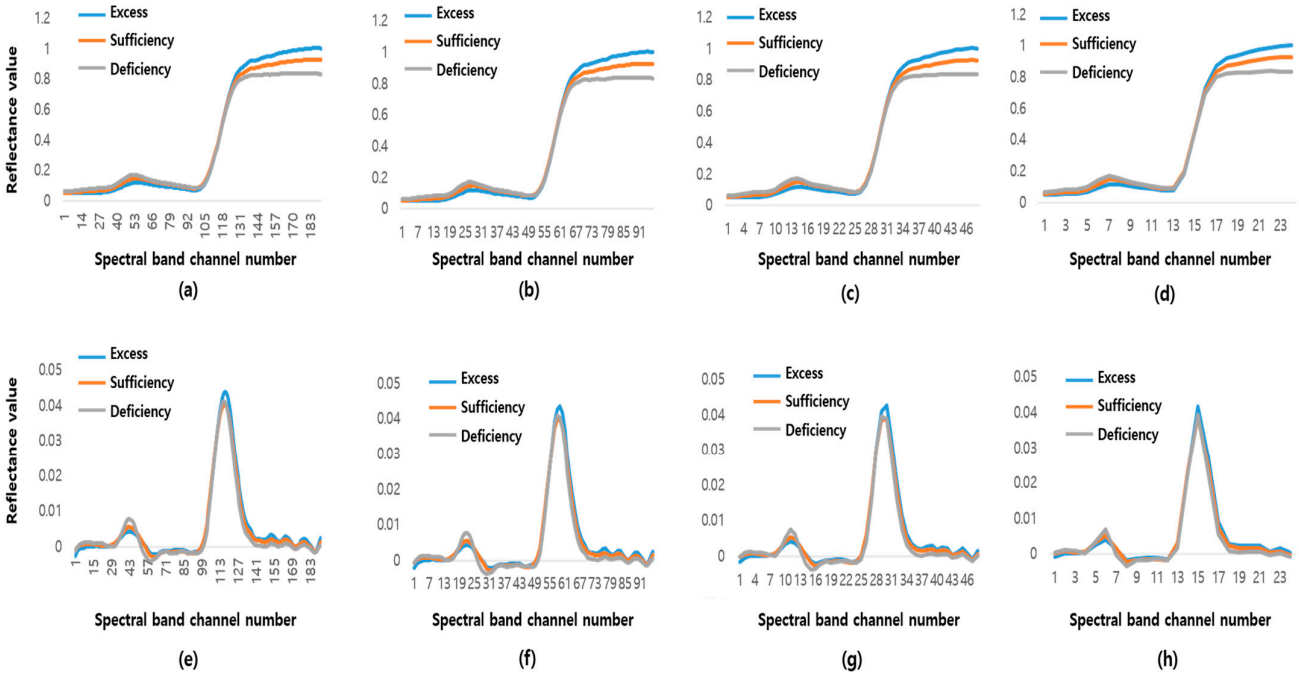

Figure 4. Reflectance curves (raw, 1st derivative) of the apple leaves, obtained via hyperspectral imaging based on 2 (**a**,**e**), 4 (**b**,**f**), 8 (**c**,**g**), and 16 (**d**,**h**) binning.

Table 2. Estimation of regression model performance using the full spectrum.

		Calibration		Validation		Prediction	
		R^2	RMSE (%)	R^2	RMSE (%)	R^2	RMSE (%)
Raw	PLSR	0.639	0.267	0.607	0.278	0.619	0.261
	SVR	0.79	0.21	0.706	0.245	0.682	0.245
	XGBoost	0.892	0.151	0.715	0.24	0.756	0.215
1st dev	PLSR	0.688	0.248	0.587	0.289	0.696	0.24
	SVR	0.699	0.248	0.638	0.271	0.739	0.224
	XGBoost	0.999	0.003	0.65	0.268	0.596	0.288

Table 3 presents the results of the variable selection method for CARS, Rfrog, and SPA. From the raw spectral data, CARS selected seven wavelengths, Rfrog selected ten wavelengths, and SPA selected six wavelengths, primarily from the green, red edge, and near-infrared (NIR) (800, 870 nm) regions. In the case of variable selection using the first derivative, CARS, Rfrog, and SPA retained the same number of wavelengths as in the raw data.

Table 3. Selection of wavelengths using the full-spectrum variable selection algorithm.

	Variable Selection Method	Spectral Band Channel Numbers
RAW	CARS	553, 556, 687, 748, 767, 845, 896
	Rfrog	483, 505, 510, 553, 652, 668, 823, 839, 877
	SPA	703, 727, 743, 847, 890, 898
1st dev	CARS	561, 572, 671, 673, 695, 735, 764
	Rfrog	561, 572, 596, 671, 673, 697, 732, 735
	SPA	507, 569, 676, 695, 730, 740

The results of a PLSR analysis using spectral data for predicting leaf nitrogen content based on variable selection are presented in Table 4. With respect to the raw spectral data, the CARS method achieved the highest prediction performance when seven wavelengths

were selected. With four latent variables, the calibration model exhibited an R^2 of 0.603 and an RMSE of 0.285%, the validation model exhibited an R^2 of 0.566 and an RMSE of 0.296%, and the prediction model exhibited an R^2 of 0.608 and an RMSE of 0.274%. In the case of the first derivative, the Rfrog method showed the highest performance for the calibration model, with an R^2 of 0.674 and an RMSE of 0.256%; for the validation model, with an R^2 of 0.616 and an RMSE of 0.296%; and for the prediction model, with an R^2 of 0.7 and an RMSE of 0.236%. According to the SVR analysis, the highest performance of the model for the raw spectral data was observed for the SPA variable selection, for which the calibration R^2 was 0.797, the RMSE was 0.202%, the prediction R^2 was 0.765, and the RMSE was 0.208%. The first derivative showed the highest performance in the Rfrog method (calibration: R^2 = 0.674, RMSE = 0.256%; prediction: R^2 = 0.678, RMSE = 0.248%). Although the proposed model exhibited lower performance than the raw images did, the prediction model's performance was greater than that of PLSR. XGBoost regression analysis showed the highest performance with CARS algorithm-based variable selection in the raw data (calibration: R^2 = 0.797, RMSE = 0.202%; prediction: R^2 = 0.765, RMSE = 0.208%). Other variable selection algorithms also showed good evaluation performance, with an R^2 that was greater than 0.7. In the case of the first derivative, while all the variable selection algorithms exhibited 90% performance in the calibration model, a lower performance below 60% in the validation model suggested the occurrence of overfitting.

Table 4. Estimation of regression model performance using selected wavelengths.

			Calibration		Validation		Prediction	
			R^2	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
PLSR	Raw	CARS	0.603	0.285	0.566	0.296	0.608	0.274
		Rfrog	0.629	0.274	0.599	0.284	0.575	0.284
		SPA	0.467	0.358	0.345	0.376	0.356	0.395
	1st Dev	CARS	0.597	0.285	0.563	0.296	0.538	0.305
		Rfrog	0.608	0.281	0.565	0.295	0.585	0.291
		SPA	0.608	0.281	0.575	0.292	0.594	0.283
SVR	Raw	CARS	0.754	0.226	0.729	0.236	0.754	0.213
		Rfrog	0.758	0.236	0.707	0.243	0.742	0.217
		SPA	0.797	0.202	0.687	0.251	0.765	0.208
	1st Dev	CARS	0.671	0.258	0.614	0.278	0.662	0.264
		Rfrog	0.674	0.256	0.616	0.278	0.678	0.248
		SPA	0.669	0.259	0.623	0.276	0.666	0.258
XGBoost	Raw	CARS	0.846	0.177	0.705	0.244	0.756	0.218
		Rfrog	0.892	0.148	0.731	0.233	0.732	0.222
		SPA	0.844	0.178	0.729	0.233	0.7	0.236
	1st Dev	CARS	0.889	0.154	0.539	0.304	0.708	0.231
		Rfrog	0.999	0.005	0.611	0.28	0.753	0.212
		SPA	0.999	0.002	0.565	0.28	0.575	0.286

3.4. The Development of a Leaf Nitrogen Content Prediction Model Based on a 4-Binning Full Spectrum

Table 5 presents the results of the leaf nitrogen content prediction model using both raw and first derivative wavelength data based on the spectral resolution of 4-binning. For the raw data prediction model, the PLSR showed an R^2 of 0.618 and an RMSE of 0.271%. For SVR, the results had an R^2 of 0.652 and an RMSE of 0.255%. XGBoost demonstrated an R^2 value of 0.755 and an RMSE of 0.216%. Compared with that of 2-binning, the prediction model performance in terms of SVR decreased by 4%; however, there was no significant decrease in performance for PLSR or XGBoost. For the first derivative, the PLSR showed R^2 and RMSE values of 0.591 and 0.284%, respectively, which were notably lower than those

of the 2-binning method. However, the performance of the prediction model improved for both the SVR and XGBoost models.

Table 5. Estimation of regression model performance using the full spectrum based on spectral 4-binning.

		Calibration		Validation		Prediction	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
Raw	PLSR	0.643	0.269	0.59	0.287	0.617	0.271
	SVM	0.811	0.2	0.707	0.244	0.652	0.255
	XGBoost	0.884	0.155	0.711	0.241	0.755	0.216
1st dev	PLSR	0.654	0.264	0.596	0.285	0.643	0.263
	SVM	0.704	0.245	0.638	0.27	0.748	0.219
	XGBoost	0.998	0.015	0.601	0.284	0.655	0.257

Table 6 presents the results of the variable selection method for the CARS, Rfrog, and SPA algorithms based on spectral resolution 4-binning. For the raw data, eight wavelengths were selected by CARS, and ten wavelengths were chosen by Rfrog. In the case of SPA, six wavelengths were determined. These selections were primarily from the green, red edge, and NIR (800, 870 nm) regions, similar to the 2-binning results. For the first derivative variable selection, eight wavelengths were chosen by CARS, ten by Rfrog, and six by SPA.

Table 6. Selection of the wavelength for 4-binning using the variable selection algorithm.

	Variable Selection Method	Spectral Band Channel Numbers
RAW	CARS	555, 560, 695, 745, 795, 835, 845, 905
	Rfrog	540, 555, 690, 695, 745, 770, 795, 825, 905
	SPA	725, 740, 745, 845, 860, 890
1st dev	CARS	490, 680, 685, 690, 705, 735, 740, 885
	Rfrog	560, 565, 570, 605, 680, 685, 700, 735, 740, 795
	SPA	435, 505, 675, 705, 730

Based on the 4-binning for variable selection, PLSR analysis showed that the raw data achieved the highest performance with Rfrog (calibration: R^2 = 0.639, RMSE = 0.270%; validation: R^2 = 0.572, RMSE = 0.293%; prediction: R^2 = 0.612, RMSE = 0.272%). On the other hand, SPA-selected variables did not include wavelengths from the visible light region, such as green and red, leading to the assumption that predictions using these variables might exhibit a lower performance (Table 7). For the first derivative, the highest performance was observed with CARS (calibration: R^2 = 0.631, RMSE = 0.274%; validation: R^2 = 0.596, RMSE = 0.290%; prediction: R^2 = 0.610, RMSE = 0.275), suggesting that the performance of the prediction models varies depending on the selection of red edge and NIR wavelengths compared to those of Rfrog and SPA. According to the SVR analysis, the raw data exhibited the highest performance for CARS, the calibration model (R^2 = 0.746, RMSE = 0.229%), and the prediction model (R^2 = 0.760, RMSE = 0.212%). This trend was consistent with the first derivative, which exhibited the highest performance for CARS (calibration: R^2 = 0.690, RMSE = 0.254; prediction: R^2 = 0.651, RMSE = 0.271%). For the XGBoost prediction model results, both the raw data and the first derivative demonstrated the highest performance when variable selection was conducted using CARS. However, for the first derivative, overfitting was evident, consistent with the spectral resolution of 2-binning.

Table 7. Estimation of regression model performance using selected wavelengths based on spectral 4-binning.

			Calibration		Validation		Prediction	
			R^2	RMSE	R^2	RMSE	R^2	RMSE
PLSR	Raw	CARS	0.596	0.284	0.530	0.307	0.617	0.270
		Rfrog	0.639	0.270	0.572	0.293	0.612	0.272
		SPA	0.613	0.281	0.585	0.290	0.556	0.290
	1st Dev	CARS	0.631	0.274	0.581	0.290	0.610	0.275
		Rfrog	0.596	0.285	0.582	0.290	0.578	0.291
		SPA	0.536	0.312	0.516	0.324	0.505	0.315
SVR	Raw	CARS	0.746	0.229	0.726	0.238	0.760	0.212
		Rfrog	0.751	0.228	0.729	0.237	0.753	0.214
		SPA	0.698	0.246	0.674	0.256	0.743	0.223
	1st Dev	CARS	0.690	0.254	0.670	0.261	0.651	0.271
		Rfrog	0.700	0.245	0.648	0.266	0.588	0.281
		SPA	0.614	0.284	0.599	0.289	0.577	0.296
XGBoost	Raw	CARS	0.849	0.175	0.738	0.229	0.751	0.217
		Rfrog	0.852	0.174	0.722	0.236	0.747	0.214
		SPA	0.803	0.200	0.652	0.264	0.730	0.222
	1st Dev	CARS	0.999	0.005	0.623	0.256	0.712	0.243
		Rfrog	0.924	0.130	0.589	0.287	0.705	0.234
		SPA	0.900	0.144	0.569	0.295	0.656	0.251

3.5. The Development of a Leaf Nitrogen Content Prediction Model Based on an 8-Binning Full Spectrum

Table 8 presents the results of the leaf nitrogen content prediction models using raw and first derivative full-spectrum data based on the 8-binning methods. The raw data showed that PLSR at the 4-binning sites had an R^2 of 0.617 and an RMSE of 0.271%, and the highest performance was observed here, with an R^2 of 0.657 and an RMSE of 0.243%. XGBoost had an R^2 of 0.752 and an RMSE of 0.218%, similar to the 4-binning methods used in both PLSR and XGBoost, whereas SVR had a lower performance. In the case of the first derivative, the SVR prediction model showed a decreased performance compared to that of 2- and 4-binning, whereas PLSR improved (R^2 = 0.626, RMSE = 0.267%). XGBoost showed an improved performance compared to previous spectral resolutions but still exhibited overfitting.

Table 8. Estimation of regression model performance using the full spectrum based on spectral 8-binning.

		Calibration		Validation		Prediction	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
Raw	PLSR	0.643	0.269	0.593	0.286	0.617	0.271
	SVM	0.775	0.218	0.699	0.248	0.687	0.243
	XGBoost	0.880	0.158	0.735	0.23	0.752	0.218
1st dev	PLSR	0.657	0.263	0.596	0.285	0.653	0.259
	SVM	0.695	0.250	0.640	0.270	0.726	0.231
	XGBoost	0.960	0.094	0.558	0.299	0.663	0.255

Table 9 presents the results for variable selection using methods such as CARS, Rfrog, and SPA based on the 8-binning spectral resolution. In the raw data, CARS selected 8 wavelengths, Rfrog chose 10 wavelengths, and SPA identified 5 wavelengths. In the first derivative data, CARS was used for four wavelengths, Rfrog was used for ten wavelengths, and SPA was used for four wavelengths.

Table 9. Selection of the wavelength for 8-binning using a variable selection algorithm.

	Variable Selection Method	Spectral Band Channel Numbers
RAW	CARS	550, 560, 670, 680, 740, 760, 850, 900
	Rfrog	520, 550, 570, 670, 680, 740, 760, 850, 900
	SPA	580, 610, 730, 880, 900
1st dev	CARS	670, 680, 690, 730
	Rfrog	470, 560, 610, 670, 680, 720, 730, 740, 790
	SPA	660, 680, 730, 890

Based on the 8-binning, PLSR analysis using the raw data and variables selected by CARS showed the highest performance for the latent variable 3. The calibration model had an R^2 of 0.612 and an RMSE of 0.281%, the validation model had an R^2 of 0.577 and an RMSE of 0.292%, and the prediction model had an R^2 of 0.580 and an RMSE of 0.282% (Table 10). In the case of the first derivative data, which are the same as those for 4-binning, the highest performance was observed for Rfrog. The calibration model had an R^2 of 0.663 and an RMSE of 0.261%, the validation model had an R^2 of 0.631 and an RMSE of 0.273%, and the prediction model had an R^2 of 0.693 and an RMSE of 0.238%. However, as the spectral resolution decreased, the performance of the prediction models also decreased. According to the SVR analysis, regardless of whether the raw or first derivative data were used, the performance of the prediction models was greater than 0.7. Among these models, the best performance was observed with the raw data using variables that were selected by CARS, a calibration model R^2 of 0.745 and an RMSE of 0.232%, a validation model R^2 of 0.722 and an RMSE of 0.241%, and a prediction model R^2 of 0.754 and an RMSE of 0.238%. For XGBoost, although Rfrog achieved the highest performance, the prediction results were lower than those of SVR.

Table 10. Estimation of regression model performance using selected wavelengths based on spectral 8-binning.

			Calibration		Validation		Prediction	
			R^2	RMSE	R^2	RMSE	R^2	RMSE
PLSR	Raw	CARS	0.612	0.281	0.577	0.292	0.580	0.282
		Rfrog	0.609	0.282	0.573	0.294	0.573	0.284
		SPA	0.586	0.291	0.558	0.300	0.515	0.300
	1st Dev	CARS	0.608	0.283	0.577	0.293	0.621	0.266
		Rfrog	0.663	0.261	0.631	0.273	0.693	0.238
		SPA	0.604	0.284	0.588	0.289	0.623	0.266
SVR	Raw	CARS	0.745	0.232	0.722	0.241	0.754	0.214
		Rfrog	0.750	0.229	0.727	0.239	0.756	0.213
		SPA	0.711	0.241	0.651	0.265	0.702	0.236
	1st Dev	CARS	0.672	0.273	0.656	0.20	0.653	0.258
		Rfrog	0.723	0.233	0.665	0.259	0.712	0.232
		SPA	0.642	0.279	0.613	0.280	0.603	0.276
XGBoost	Raw	CARS	0.856	0.172	0.711	0.241	0.721	0.228
		Rfrog	0.858	0.171	0.708	0.242	0.729	0.226
		SPA	0.790	0.207	0.614	0.279	0.655	0.256
	1st Dev	CARS	0.819	0.194	0.512	0.314	0.737	0.222
		Rfrog	0.999	0.002	0.571	0.295	0.705	0.233
		SPA	0.882	0.161	0.475	0.325	0.640	0.256

3.6. The Development of a Leaf Nitrogen Content Prediction Model Based on a 16-Binning Full Spectrum Study

Table 11 presents the results of the leaf nitrogen content prediction model using full-spectrum data with a spectral resolution of 16 bp. For the raw data, the prediction

model results showed that R^2 for PLSR was 0.598, and that for XGBoost, it was 0.728. These values are approximately 2% and 3% lower than the spectral resolutions of 2-, 4-, and 8-binning, indicating a decrease in the prediction model's performance. However, the results obtained using SVR exhibited the highest performance among the spectral resolutions, with an R^2 value of 0.709. With respect to the first derivative data, PLSR, SVR, and XGBoost exhibited the lowest performances compared with those of the 2-, 4-, and 8-binning resolutions. Notably, XGBoost exhibited overfitting which was similar to the previous 2-, 4-, and 8-binning results.

Table 11. Estimation of regression model performance using the full spectrum based on spectral 16-binning.

		Calibration		Validation		Prediction	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
Raw	PLSR	0.633	0.272	0.592	0.287	0.598	0.279
	SVM	0.743	0.231	0.705	0.246	0.709	0.233
	XGBoost	0.85	0.176	0.673	0.256	0.728	0.224
1st dev	PLSR	0.623	0.275	0.591	0.287	0.583	0.284
	SVM	0.667	0.267	0.626	0.281	0.692	0.24
	XGBoost	0.999	0.003	0.653	0.264	0.685	0.24

Table 12 presents the results of variable selection using CARS, Rfrog, and SPA based on a spectral resolution of 16 bins. For the raw data, CARS selected 7 wavelengths, while Rfrog selected 10 wavelengths. SPA identified 10 wavelengths. In the case of the first derivative data, variable selection resulted in nine wavelengths for CARS, ten wavelengths for Rfrog, and four wavelengths for SPA.

Table 12. Selection of the wavelength for 16-binning using the variable selection algorithm.

	Variable Selection Method	Spectral Band Channel Numbers
RAW	CARS	540, 560, 740, 760, 800, 840, 900
	Rfrog	460, 480, 540, 560, 720, 740, 780, 800, 840, 900
	SPA	540, 560, 700, 740, 760, 780, 800, 880
1st dev	CARS	500, 540, 560, 700, 720, 740, 760, 780, 900
	Rfrog	480, 560, 580, 660, 680, 700, 740, 760, 800, 880
	SPA	480, 500, 520, 740, 760, 800

Table 13 presents the results of the analysis conducted using PLSR, SVR, and XGBoost based on the selected variables at a spectral resolution of 16 bins. For PLSR, using variables selected by SPA, the calibration model had an R^2 of 0.667 and an RMSE of 0.259, the validation model had an R^2 of 0.603 and an RMSE of 0.282, and the prediction model demonstrated the highest performance, with an R^2 of 0.704 and an RMSE of 0.236. In the case of SVR, the prediction model using variables selected by Rfrog (random frog) from the raw data demonstrated the highest performance (R^2 = 0.748, RMSE = 0.215%). Finally, the results of the XGBoost regression analysis showed that, for the raw data, the prediction performance using variables selected by CARS was the highest (R^2 = 0.746, RMSE = 0.216%). For the first derivative data, although overfitting was resolved compared to previous spectral resolutions, the overall performance was lower than that of the raw data.

Table 13. Estimation of regression model performance using selected wavelengths based on spectral 16-binning.

			Calibration		Validation		Prediction	
			R^2	RMSE	R^2	RMSE	R^2	RMSE
PLSR	Raw	CARS	0.612	0.281	0.566	0.296	0.648	0.257
		Rfrog	0.632	0.273	0.567	0.295	0.614	0.275
		SPA	0.667	0.259	0.603	0.282	0.704	0.236
	1st Dev	CARS	0.638	0.270	0.600	0.284	0.650	0.255
		Rfrog	0.573	0.366	0.558	0.368	0.667	0.322
		SPA	0.620	0.277	0.600	0.284	0.604	0.274
SVR	Raw	CARS	0.741	0.230	0.716	0.240	0.743	0.217
		Rfrog	0.746	0.226	0.702	0.245	0.748	0.215
		SPA	0.736	0.231	0.696	0.247	0.718	0.227
	1st Dev	CARS	0.646	0.272	0.620	0.282	0.662	0.251
		Rfrog	0.664	0.274	0.648	0.279	0.684	0.245
		SPA	0.652	0.283	0.638	0.283	0.669	0.251
XGBoost	Raw	CARS	0.831	0.185	0.687	0.251	0.746	0.216
		Rfrog	0.842	0.182	0.690	0.249	0.739	0.220
		SPA	0.826	0.188	0.669	0.258	0.733	0.221
	1st Dev	CARS	0.893	0.151	0.537	0.306	0.615	0.266
		Rfrog	0.743	0.233	0.527	0.309	0.565	0.284
		SPA	0.896	0.150	0.535	0.306	0.645	0.254

4. Discussion

This paper presents the results of a prediction model for apple tree leaf nitrogen content using full-spectrum wavelengths. For the raw data, the R^2 values for PLSR, SVR, and XGB ranged from 0.633 to 0.643, 0.743 to 0.811, and 0.850 to 0.892, respectively. For the first derivative, the R^2 values for the PLSR and SVR ranged from 0.623 to 0.688 and 0.667 to 0.704, respectively, and overfitting was observed with XGB. When compared with results from previous research, the raw data showed that PLSR had an R^2 of 0.773 and the first derivative data had an R^2 of 0.774 [34]. The improvement in performance can, exactly, be attributed to the higher spectral resolution. Despite maintaining the same wavelength range, the increased spectral resolution introduces a greater number of wavelengths. This, in turn, contributes to a higher count of independent variables in the prediction model, ultimately leading to improved performance.

Hyperspectral data, represented as continuous curves, constitute a complex dataset because of differences in reflectance values, even within adjacent wavelength bands in the same spectral range. These results suggest that nonlinear regression analysis methods, such as SVR and XGB, are more advantageous in terms of prediction performance and interpretability than linear regression analyses, such as PLSR [35]. Additionally, Savitzky–Golay filtering, a preprocessing method which is used to reduce the noise caused by light, smooths the data by adjusting the polynomial order and window size. However, the first derivative, which represents the rate of change in adjacent wavelengths rather than the inherent value of the reflectance, is sensitive to spectral changes and peak enhancement [36]. This sensitivity is beneficial but can be problematic when noise is present, as it leads to significant changes in the gradient. Such drawbacks are evident in the results of this experiment, where a lower prediction performance was observed or overfitting occurred in the tree-based boosting method, XGB, owing to the sensitivity of the first derivative data.

A comparison of the variable selection algorithms revealed that the primary selections were made at the blue (470–490 nm), green (550 nm), red edge (680–740 nm), and NIR wavelengths. In the visible light spectrum, wavelengths that were closely associated with chlorophyll were chosen. For chlorophyll a, the highest absorption occurred at the boundary of the red and red edge wavelengths, approximately at 670 nm, whereas chloro-

phyll b exhibited maximum absorption at 470 nm and reflection at the green wavelength. Furthermore, in the nonvisible spectrum, specifically at the red edge and near-infrared (NIR) wavelengths, differences in reflectance values reflect the nutritional status of leaves, which typically increase in value when the nutritional state is favorable [37]. The structural characteristics of leaves vary with nitrogen levels: a higher nitrogen content results in an increase in the leaf surface area. Additionally, the leaf epidermis thickens, and cells in the mesophyll tissue increase in size and become more densely arranged, leading to an increase in the chlorophyll content [38,39]. Thickening of epidermal tissue facilitates active gas exchange, resulting in enhanced photosynthesis. Based on the spectral characteristics corresponding to the structural changes in the leaves, the analysis results considering the full spectrum revealed that for the PLSR models, R^2 = 0.619, which was lower than that of CARS, Rfrog, and SPA. In contrast, for the SVR models, CARS had an R^2 of 0.754, Rfrog had an R^2 of 0.742, and SPA had an R^2 of 0.765, indicating a greater performance than those of the models using the full spectrum. In the case of XGB, the performance across various variable selection algorithms ranged from 0.7 to 0.756, showing effectiveness that is similar to the results obtained using full-spectrum analysis. These results indicate differences based on the variable selection algorithm. PLSR, which creates new variables through linear combinations of independent variables, seems to lack an adequate explanation of the selected variables. In contrast, the use of the radial basis function kernel that is based on Gaussian functions in SVR, along with various loss functions (such as the mean square error and mean absolute error) and gradient boosting in tree-based XGB, allows for the interpretation of nonlinear relationships between independent and predicted variables, unlike in PLSR. The improvement in predictive performance through the optimization of prediction models, including hyperparameter tuning for each analysis method, suggested that fewer variables can yield similar or better results in the prediction models. Another method explored in previous research reduces variables that are involved in predicting nitrogen content using various vegetation indices and the red edge wavelength. The results showed that the R^2 based on the BPNN model was 0.77 [40]. However, since vegetation indices require a combination of multiple wavelengths, lowering the spectral resolution might lead to changes in the values of these indices. Therefore, reducing spectral resolution is considered inadequate as an alternative for variable reduction in this context.

When comparing the wavelengths that were selected based on the 2-binning criterion with those selected through spectral binning at 4, 8, and 16 bp, it was observed that for the number of wavelengths selected by CARS in the raw data, similar or adjacent wavelengths were chosen regardless of the spectral resolution. When comparing the XGB prediction models that exhibited the highest performance for each spectral resolution, the lowest value was observed for the 16-binning model, with an R^2 of 0.743, and the highest was observed for the 4-binning model, with an R^2 of 0.760, indicating a similar performance with a difference of only 1.7%. In the case of Rfrog, unlike CARS, the selected wavelengths varied slightly according to the spectral resolution. However, a 2% difference in the coefficient of determination was observed based on the spectral resolution in the XGB prediction model. For SPA, in the case of 2- and 4-binnings, only the red edge and NIR wavelength regions were selected, which differed from the wavelengths chosen by CARS and Rfrog, which showed a difference in evaluation performance of approximately 3% to 5% compared with previous variable selection methods. Additionally, the lowest R^2 value (0.65) was observed after eight binning cycles, which seems to be due to the decrease in performance attributed to whether the 680 nm wavelength, located between the red and red edge wavelengths, was selected under the same binning criteria for CARS and Rfrog. When the first derivative spectral data were used for variable selection through spectral binning, there was no similarity in the wavelengths that were selected based on spectral resolution in contrast to the raw data. Consequently, the regression analysis, particularly for SVR, exhibited substantial deviations, with R^2 values ranging from 0.577 to 0.712. This is because as the spectral resolution decreases, leading to a reduction in the number of wavelengths, continuous spectral data loss occurs. While the raw data retain the inherent spectroscopic

characteristics of the canopy, the first derivative data, owing to data loss during spectral binning, respond sensitively to even minor changes. As a result, the prediction performance was unstable and varied with spectral resolution. Furthermore, a high spectral resolution does not necessarily translate into an improved performance in predictive models.

Figure 5 presents the mapping of hyperspectral images using the wavelengths selected based on CARS. The results are divided into red to green colors based on the nitrogen content range, indicating that the leaf nitrogen content ranged from a minimum of 0% to a maximum of 4%. Spectral binning, which combines wavelength bands, can reduce the number of wavelength bands and lead to the loss of continuous spectral data, potentially degrading the performance of the prediction models [41]. However, this process can also reduce the costs associated with data processing and analysis. Additionally, by combining adjacent wavelength bands, the SNR can be enhanced, and the inclusion of similar spectral data can be minimized. Therefore, appropriate spectral binning may offer advantages such as a reduced data processing speed owing to a reduction in high-dimensional spectral data and enhanced predictive performance [42–44].

Figure 5. Mapping of leaf nitrogen concentration in apple trees using the SVR model and variable selection of CARS.

5. Conclusions

In this study, various predictive models were developed and compared via regression analysis with both the full spectrum and selected significant wavelengths to predict the leaf nitrogen content in apple trees via hyperspectral imaging. In addition, spectral binning was used to reduce the spectral resolution to 5, 10, or 20 nm, and regression analysis was conducted using only the wavelengths identified through variable selection. The predictive performance at these reduced spectral resolutions was compared to that at the original spectral resolution to determine the optimal spectral resolution. The study showed that reducing the spectral resolution reduces the number of wavelengths, leading to data loss. However, the intrinsic shape of the spectral curve is maintained, suggesting that performance can be preserved, even with a lower spectral resolution. However, hyperspectral imaging has a narrow spectral resolution, allowing for detailed interpretation of physiological responses in crops across numerous wavelengths. However, due to the high cost of equipment and various constraints during image acquisition, to address these issues, the spectral resolution was decreased to achieve satisfactory results. These results imply that the development of a miniaturized multispectral sensor can be practical and cost-effective, potentially serving as an alternative to hyperspectral sensors. Furthermore, utilizing geographic information systems, including sensors and drones, could enhance the precision of monitoring apples that are cultivated in extensive orchards. Through stable cultivation management, this approach could secure both quantity and quality, providing a reliable means for ensuring stable crop yields and quality control.

Author Contributions: Conceptualization, S.J. and J.C.; methodology, S.J. and J.C.; validation, J.J. and J.K.; formal analysis, S.J.; investigation, S.L. and D.L.; writing—original draft preparation, J.H.; writing—review and editing, J.C. and J.H.; supervision, J.C.; project administration, J.H.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out with the support of "The Cooperative Research Program for Agriculture Science and Technology Development (Project No: PJ0156572023)", Rural Development Administration, Republic of Korea.

Data Availability Statement: All data are included in the manuscript.

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

References

1. Kowalczyk, W.; Wrona, D.; Kowalczyk, W.; Przybyłko, S. Content of minerals in soil, apple tree leaves and fruits depending on nitrogen fertilization. *J. Elem.* **2016**, *22*, 67–77. [CrossRef]
2. Kowalczyk, W.; Wrona, D.; Przybyłko, S. Effect of nitrogen fertilization of apple orchard on soil mineral nitrogen content, yielding of the apple trees and nutritional status of leaves and fruits. *Agriculture* **2022**, *12*, 2169. [CrossRef]
3. Cheng, L.; Ma, F.; Ranwala, D. Nitrogen storage and its interaction with carbohydrates of young apple trees in response to nitrogen supply. *Tree Physiol.* **2004**, *24*, 91–98. [CrossRef] [PubMed]
4. Mohammad Sokri, S.; Babalar, M.; Barker, A.V.; Lesani, H.; Asgari, M.A. Fruit quality and nitrogen, potassium, and calcium content of apple as influenced by nitrate: Ammonium ratios in tree nutrition. *J. Plant Nutr.* **2015**, *38*, 1619–1627. [CrossRef]
5. Holb, I.J.; Gonda, I.; Vágó, I.; Nagy, P.T. Seasonal dynamics of nitrogen, phosphorus, and potassium contents of leaf and soil in environmental friendly apple orchards. *Commun. Soil Sci. Plant Anal.* **2009**, *40*, 694–705. [CrossRef]
6. Marsh, K.B.; Volz, R.K.; Cashmore, W.; Reay, P. Fruit colour, leaf nitrogen level, and tree vigour in 'Fuji' apples. *N. Z. J. Crop Hortic. Sci.* **1996**, *24*, 393–399. [CrossRef]
7. Chen, Y.; Ji, Y.; Zhou, J.; Chen, X.; Shen, W. Computation of signal-to-noise ratio of airborne hyperspectral imaging spectrometer. In Proceedings of the International Conference on Systems and Informatics (ICSAI2012), Yantai, China, 19–20 May 2012; IEEE Publications: Hoboken, NJ, USA, 2012; Volume 2012, pp. 1046–1049. [CrossRef]
8. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* **2017**, *9*, 1110. [CrossRef]
9. Mahesh, S.; Jayas, D.S.; Paliwal, J.; White, N.D.G. Hyperspectral imaging to classify and monitor quality of agricultural materials. *J. Stored Prod. Res.* **2015**, *61*, 17–26. [CrossRef]
10. Kim, Y.; Glenn, D.M.; Park, J.; Ngugi, H.K.; Lehman, B.L. Hyperspectral image analysis for water stress detection of apple trees. *Comput. Electron. Agric.* **2011**, *77*, 155–160. [CrossRef]
11. Cohen, Y.; Alchanatis, V.; Meron, M.; Saranga, Y.; Tsipris, J. Estimation of leaf water potential by thermal imagery and spatial analysis. *J. Exp. Bot* **2005**, *56*, 1843–1852. [CrossRef]
12. Sugiura, R.; Tsuda, S.; Tamiya, S.; Itoh, A.; Nishiwaki, K.; Murakami, N.; Shibuya, Y.; Hirafuji, M.; Nuske, S.; Hirafuji, M.; et al. Field phenotyping system for the assessment of potato late blight resistance using RGB imagery from an unmanned aerial vehicle. *Biosyst. Eng.* **2016**, *148*, 1–10. [CrossRef]
13. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [CrossRef]
14. Sankaran, S.; Khot, L.R.; Espinoza, C.Z.; Jarolmasjed, S.; Sathuvalli, V.R.; Vandemark, G.J.; Miklas, P.N.; Carter, A.H.; Pumphrey, M.O.; Knowles, N.R.; et al. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* **2015**, *70*, 112–123. [CrossRef]
15. Yuan, H.; Yang, G.; Li, C.; Wang, Y.; Liu, J.; Yu, H.; Feng, H.; Xu, B.; Zhao, X.; Yang, X.; et al. Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models. *Remote Sens.* **2017**, *9*, 309. [CrossRef]
16. Paul, G.C.; Saha, S.; Hembram, T.K. Application of phenology-based algorithm and linear regression model for estimating rice cultivated areas and yield using remote sensing data in Bansloi River Basin, Eastern India. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 10036.
17. Tennakoon, S.B.; Murty, V.V.N.; Eiumnoh, A. Estimation of cropped area and grain yield of rice using remote sensing data. *Int. J. Remote Sens.* **1992**, *13*, 427–439. [CrossRef]
18. Kang, Y.; Nam, J.; Kim, Y.; Lee, S.; Seong, D.; Jang, S.; Ryu, C. Assessment of regression models for predicting rice yield and protein content using unmanned aerial vehicle-based multispectral imagery. *Remote Sens.* **2021**, *13*, 1508. [CrossRef]
19. Li, M. Moving beyond the linear regression model: Advantages of the quantile regression model. *J. Manag.* **2015**, *41*, 71–98. [CrossRef]
20. Lu, R. Detection of bruises on apples using near–infrared hyperspectral imaging. *Trans. ASAE* **2003**, *46*, 523.
21. Kang, Y.S.; Park, K.S.; Kim, E.R.; Jeong, J.C.; Ryu, C.S. Estimation of the Total Nonstructural Carbohydrate Concentration in Apple Trees Using Hyperspectral Imaging. *Horticulturae* **2023**, *9*, 967. [CrossRef]
22. Guo, X.; Zhu, X.; Li, C.; Wei, Y.; Yu, X.; Zhao, G.; Sun, H. Hyperspectral Inversion of potassium content in apple leaves based on vegetation index. *Agric. Sci.* **2017**, *8*, 825–836. [CrossRef]
23. Manea, D.; Calin, M.A. Hyperspectral imaging in different light conditions. *Imaging Sci. J.* **2015**, *63*, 214–219. [CrossRef]

24. Moghadam, P.A.; Sharma, N.; Hefeeda, M. Enabling hyperspectral imaging in diverse illumination conditions for indoor applications. In Proceedings of the 12th ACM Multimedia Systems Conference, Istanbul, Turkey, 28 September–1 October 2021; pp. 23–35. [CrossRef]
25. Geladi, P.; Burger, J.; Lestander, T. Hyperspectral imaging: Calibration problems and solutions. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 209–217. [CrossRef]
26. Press, W.H.; Teukolsky, S.A. Savitzky-Golay smoothing filters. *Comput. Phys.* **1990**, *4*, 669–672. [CrossRef]
27. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [CrossRef]
28. Sun, T.; Zhang, J.; Zhang, Q.; Li, X.; Li, M.; Yang, Y.; Zhou, J.; Wei, Q.; Zhou, B.; Wei, Q.; et al. Integrative physiological, transcriptome, and metabolome analysis reveals the effects of nitrogen sufficiency and deficiency conditions in apple leaves and roots. *Environ. Exp. Bot.* **2021**, *192*, 104633. [CrossRef]
29. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.; Holmes, M.; Hanpin, M. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* **2010**, *667*, 14–32. [CrossRef]
30. Chong, I.G.; Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112. [CrossRef]
31. Yun, Y.H.; Li, H.D.; Deng, B.C.; Cao, D.S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends Anal. Chem.* **2019**, *113*, 102–115. [CrossRef]
32. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [CrossRef]
33. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 65–73. [CrossRef]
34. Ye, X.; Abe, S.; Zhang, S. Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. *Precis. Agric.* **2020**, *21*, 198–225. [CrossRef]
35. Ruffin, C.; King, R.L.; Younan, N.H. A combined derivative spectroscopy and Savitzky-Golay filtering method for the analysis of hyperspectral data. *GISci. Remote Sens.* **2008**, *45*, 1–15. [CrossRef]
36. Jabbar, H.K.; Khan, R.Z. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Comput. Sci. Commun. Instrum. Devices* **2015**, *70*, 163–172.
37. Slaton, M.R.; Raymond Hunt, E.; Smith, W.K. Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.* **2001**, *88*, 278–284. [CrossRef]
38. Merwin, I.A.; Stiles, W.C. Orchard groundcover management impacts on apple tree growth and yield, and nutrient availability and uptake. *J. Am. Soc. Hortic. Sci.* **1994**, *119*, 209–215. [CrossRef]
39. Reich, P.B.; Walters, M.B.; Tjoelker, M.G.; Vanderklein, D.; Buschena, C. Photosynthesis and respiration rates depend on leaf and root morphology and nitrogen concentration in nine boreal tree species differing in relative growth rate. *Funct. Ecol.* **1998**, *12*, 395–405. [CrossRef]
40. Li, W.; Zhu, X.; Yu, X.; Li, M.; Tang, X.; Zhang, J.; Xue, Y.; Zhang, C.; Jiang, Y. Inversion of Nitrogen Concentration in Apple Canopy Based on UAV Hyperspectral Images. *Sensors* **2022**, *22*, 3503. [CrossRef]
41. Leatherbarrow, R.J. Using linear and non-linear regression to fit biochemical data. *Trends Biochem. Sci.* **1990**, *15*, 455–458. [CrossRef]
42. Kim, M.S.; Chao, K.; Chan, D.E.; Jun, W.; Lefcourt, A.M.; Delwiche, S.R.; Kang, S.; Lee, K.; Lee, K. Line-scan hyperspectral imaging platform for agro-food safety and quality evaluation: System enhancement and characterization. *Trans. ASABE* **2011**, *54*, 703–711. [CrossRef]
43. Abrams, M.D.; Kubiske, M.E. Leaf structural characteristics of 31 hardwood and conifer tree species in central Wisconsin: Influence of light regime and shade-tolerance rank. *For. Ecol. Manag.* **1990**, *31*, 245–253. [CrossRef]
44. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative. *J. Mach. Learn. Res.* **2009**, *10*, 13.