





Article

Hyperspectral Imaging-Based Multiple Predicting Models for Functional Component Contents in *Brassica juncea*

Jae-Hyeong Choi ^{1,2}, Soo Hyun Park ¹, Dae-Hyun Jung ^{1,3}, Yun Ji Park ¹, Jung-Seok Yang ¹, Jai-Eok Park ¹, Hyein Lee ¹ and Sang Min Kim ^{1,2,*}

¹ Smart Farm Research Center, KIST Gengneung Institute of Natural Products, Gangneung 25451, Korea

² Department of Bio-Medical Science & Technology, KIST School, University of Science and Technology, Seoul 02792, Korea

³ Department of Smart Farm Science, Kyung Hee University, Yongin 17104, Korea

* Correspondence: kims@kist.re.kr; Tel.: +82-33-650-3640; Fax: +82-33-650-3679

Abstract: Partial least squares regression (PLSR) prediction models were developed using hyperspectral imaging for noninvasive detection of the five most representative functional components in *Brassica juncea* leaves: chlorophyll, carotenoid, phenolic, glucosinolate, and anthocyanin contents. The region of interest for functional component analysis was chosen by polygon selection and the extracted average spectra were used for model development. For pre-processing, 10 combinations of Savitzky–Golay filter (S. G. filter), standard normal variate (SNV), multiplicative scatter correction (MSC), 1st-order derivative (1st-Der), 2nd-order derivative (2nd-Der), and normalization were applied. Root mean square errors of calibration (RMSEP) was used to assess the performance accuracy of the constructed prediction models. The prediction model for total anthocyanins exhibited the highest prediction level ($R_V^2 = 0.8273$; $RMSEP = 2.4277$). Pre-processing combination of SNV and 1st-Der with spectral data resulted in high-performance prediction models for total chlorophyll, carotenoid, and glucosinolate contents. Pre-processing combination of S. G. filter and SNV gave the highest prediction rate for total phenolics. SNV inclusion in the pre-processing conditions was essential for developing high-performance accurate prediction models for functional components. By enabling visualization of the distribution of functional components on the hyperspectral images, PLSR prediction models will prove valuable in determining the harvest time.

Keywords: hyperspectral image; partial least squares regression; prediction models; root mean square error of prediction; standard normal variate; total anthocyanins; total carotenoids; total chlorophylls; total glucosinolates; total phenolics



Citation: Choi, J.-H.; Park, S.H.; Jung, D.-H.; Park, Y.J.; Yang, J.-S.; Park, J.-E.; Lee, H.; Kim, S.M.

Hyperspectral Imaging-Based Multiple Predicting Models for Functional Component Contents in *Brassica juncea*. *Agriculture* **2022**, *12*, 1515. <https://doi.org/10.3390/agriculture12101515>

Academic Editors: Gniewko Niedbała and Sebastian Kujawa

Received: 17 August 2022

Accepted: 19 September 2022

Published: 21 September 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

Brassica juncea is a member of the family Brassicaceae whose leaves contain a variety of functional components including chlorophyll, carotenoid, flavonoid, and phenolic components, as well as glucosinolate and anthocyanin components [1–5]. Functional component content in crop plants can vary extensively within a wide range, depending on the cultivation conditions. The bioactivities of such functional components and their applications as pharmacological agents, health functional foods [6], and cosmetic materials [7] have been extensively studied. For example, chlorophylls and anthocyanins are components that confer the characteristic green and red colors to plant organs, respectively, and they may also indicate crop management or growth conditions in wilted or diseased plant organs [8,9]. In addition, carotenoids have been investigated with respect to their biological effects in anti-obesity activity [10], and flavonoids, phenolics, and anthocyanins have been investigated for anti-obesity [11], antioxidant, and antimicrobial activities [12,13]. The glucosinolate derivative, isothiocyanate, has been investigated for anticancer [14,15] and antimicrobial activities [16].

Recently, there has been growing interest in detection and quantification of specific bioactive substances for targeted applications. Concomitantly, studies involving the development of faster and efficient detection methods have attracted attention. The most widely used detection method involves the collection of plant samples coupled with invasive extraction using organic solvents, followed by analysis of the extracts. Such detection methods involving organic solvents and chemical reagents are time consuming, less efficient, and adversely affect the environment. To overcome these disadvantages of chemical methods, many scientists have conducted chemometrics studies based on spectroscopic knowledge to analyze the functional components in plants, such as High-Performance Liquid Chromatography (HPLC) and Near-Infrared (NIR) [17], or Fourier Transform-NIR (FT-NIR) spectroscopy [18,19]. However, these methods are invasive and inefficient, and data processing is time consuming. Alternatively, recent studies have focused on non-invasive methods. For example, a hyperspectral imaging system was used to explore various parameters including plant-derived functional components, such as total chlorophylls and total carotenoids [20], total capsaicinoids [17], total glucosinolates [21], total flavonoids [22–24], and total polysaccharides [24], as well as microbial contamination of fish flesh [25], fruit moisture content [26], and sulfite dioxide residues on fruit surfaces [27]. Among the studies aimed at developing prediction models for functional components, prediction of total anthocyanins [28] at $R_C^2 = 0.883$ and $R_V^2 = 0.830$ and total polyphenols [29] at $R_C^2 = 0.820$ and $R_V^2 = 0.551$ have been reported. However, in these studies, only one or two functional components were targeted for prediction model construction by using hyperspectral images.

In this study, it was hypothesized that multiple components could be predicted non-invasively and spontaneously from hyperspectral images from the leaves of *B. juncea*. In order to develop a prediction model for multiple components, a partial least squares regression (PLSR) method was used, with 10 pre-processing combinations. Compared to other regression models, the PLSR method enables the development of a predictive model with high accuracy using only a relatively small amount of data. For this reason, it is a representative method widely used to make industrialized products because it reduces the data acquisition time, and the model application is simple. Total chlorophylls, total carotenoids, total phenolics, total glucosinolates, and total anthocyanins were selected as target functional components for prediction, because these components are the most general functional components in plants which have a beneficial effect for human health and can be estimated by spectrophotometric method after simple solvent extraction of plant samples. *B. brassica* was cultivated under various conditions to obtain functional components with a wide range of concentrations. The harvested leaves were used for hyperspectral image and functional component quantification. In addition, visualization software was developed to detect real-time distribution of target components. The results in this study suggest the possibility of noninvasive multiple component prediction from one hyperspectral image. In addition, models and visualization software were applied to agricultural systems (such as growth chamber, indoor farm, and greenhouse) to monitor real-time functional components.

2. Materials and Methods

2.1. Plant Growth Conditions

To obtain *Brassica juncea* (L.) Czern. leaf samples with varying concentrations of functional components, three cultivation environments were implemented: an indoor farm, a greenhouse, and an open field (Figure 1). *B. juncea* was cultivated in an indoor farm in a hydroponic system under red, blue, and green LED light combination at 18–23 °C; Hoagland nutrient solution with an electrical conductivity (EC) value of 1.5 dS/m was used as a growth medium. Greenhouse and open field cultivation were carried out on the soil with fertilizer components at 20–28 °C for greenhouse and 15–20 °C under the sunlight for open field, respectively. The leaves of *B. juncea* were collected after 6 weeks of cultivation in each environment. Considering the growth phase and varied distributions of

leaf colors, 15–20 full-grown leaves from each cultivation environment were harvested. A total of 55 leaves were sampled and stored at $-20\text{ }^{\circ}\text{C}$ until further analysis. Spectral data were obtained for all the samples by using hyperspectral imaging. After four days of freeze-drying and subsequent grinding, each sample's powder was divided into 20 mg portions in triplicate for analysis of the five functional components. Content values with a large deviation for each functional component were excluded, and the mean of the remaining values was used as the component content value.



Figure 1. Cultivation features of *B. juncea* in indoor farm (A), greenhouse (B), and outdoor field (C). The plants were harvested at the sixth week of cultivation. The growth temperature for crop cultivation was maintained at $18\text{--}23\text{ }^{\circ}\text{C}$ in indoor farm (A), $20\text{--}28\text{ }^{\circ}\text{C}$ in greenhouse (B), and $15\text{--}20\text{ }^{\circ}\text{C}$ in outdoor field.

2.2. Total Chlorophyll and Carotenoid Contents

Measurements were made using extracts prepared in triplicate by adding 2 mL of 90% MeOH containing 10% water (*v/v*) to 20 mg of each sample, followed by sonication for 1 h at $40\text{ }^{\circ}\text{C}$. The resulting crude extract was centrifuged at 4000 rpm and $4\text{ }^{\circ}\text{C}$ for 20 min to separate plant debris and the supernatant. The supernatant was filtered using a $0.45\text{ }\mu\text{m}$ membrane filter; 1.5 mL of this filtrate was collected to use in quantification of functional components. For analysis, 150 μL from the 1.5 mL of the filtered supernatant was mixed with 90% MeOH containing 10% water to prepare 1.5 mL of a $10\times$ diluted solution. Following the method described previously [30], the absorbance of the diluted sample solution was measured at 665.2, 652.4, and 470.0 nm wavelengths using a spectrophotometer (Cary 60 UV-Vis, Agilent Technologies, Santa Clara, CA, USA). The specification of spectrophotometer includes Xenon Flash Lamp (80 Hz) as a light source, measuring wavelengths from 190 nm to 1100 nm with a resolution of 1.5 nm. The scanning speed of this equipment is 24,000 nm per min. Absorbance (A) at each wavelength was used in Equations (1)–(3) to calculate total chlorophyll a, total chlorophyll b, and total carotenoids respectively.

$$\text{Chla } (\mu\text{g mL}^{-1}) = 16.82 A_{665.2} - 9.28 A_{652.4} \quad (1)$$

$$\text{Chlb } (\mu\text{g mL}^{-1}) = 36.92 A_{652.4} - 16.54 A_{665.2} \quad (2)$$

$$\text{Car } (\mu\text{g mL}^{-1}) = \frac{(1000 A_{470.0} - 1.91 \text{Chla} - 95.15 \text{Chlb})}{225} \quad (3)$$

where $A_{665.2}$ is absorbance at 665.2 nm; $A_{652.4}$ is absorbance at 652.4 nm; $A_{470.0}$ is absorbance at 470 nm; Chla stands for total chlorophyll a; Chlb stands for total chlorophyll b; Chla + Chlb stands for total chlorophyll content; and Car stands for total carotenoid content. The data are expressed as mean \pm standard deviation mg g^{-1} dry weight (DW) from biological triplicates.

2.3. Total Phenolic Contents

A previously described method [31] was used after modification for the detection of total phenolic content. For analysis, 100 μL from the 1.5 mL of supernatant (Section 2.2) was mixed with 100 μL of Folin-Ciocalteu reagent and 1.5 mL of distilled water; the mixture was incubated for 5 min at room temperature; then, 300 μL of 7.5% Na_2CO_3 solution was added and the mixture was allowed to react for 1 h at room temperature. Absorbance was

measured at 765.0 nm using a spectrophotometer. A standard curve prepared using 25, 50, 100, and 250 ppm gallic acid standard solutions was used to calculate total phenolic content in the samples from their measured absorbance. The data are expressed as mean \pm standard deviation mg g^{-1} DW from biological triplicates.

2.4. Total Glucosinolate Contents

The extract prepared by the method described in 2.2 was used to obtain 1.5 mL of supernatant of the crude extract. For analysis, 50 μL from the 1.5 mL of supernatant was mixed with 150 μL of distilled water and 1.5 mL of 2 mM sodium tetrachloropalladate (II); the mixture was left to react for 1 h at room temperature. The absorbance was measured at 425.0 nm using a spectrophotometer. The absorbance value was then used to calculate total glucosinolate content according to Equation (4) [32]:

$$y \text{ (}\mu\text{mol g}^{-1}\text{)} = 1.40 + 118.86 A_{425.0} \quad (4)$$

where $A_{425.0}$ is the absorbance at 425.0 nm. The data are expressed as mean \pm standard deviation $\mu\text{mol g}^{-1}$ DW from biological triplicates.

2.5. Total Anthocyanin Contents

The method described in a previous study [33] was modified and used for the estimation of the total anthocyanins. Sample extracts were prepared in triplicate. Briefly, 2 mL of acidic MeOH containing 1% HCl (v/v) was added to 20 mg of powder sample (Section 2.1) and sonicated for 1 h at 60 °C. The resulting crude extract was centrifuged and filtered in the same way as in the pre-processing of samples for total chlorophylls (Section 2.2). Then, 300 μL of the 1.5 mL of filtered supernatant was diluted by adding MeOH containing 1% HCl, and a 5 \times diluted solution of 1.5 mL volume was prepared. Absorbance of this solution was measured at 530.0 and 600.0 nm wavelengths, and the values obtained were used to calculate total anthocyanin contents according to Equation (5), as reported previously [33].

$$y \text{ (}\text{mg g}^{-1}\text{)} = (A_{530.0} - A_{600.0}) \frac{V \times n \times Mw}{\epsilon \times m} \quad (5)$$

where $A_{530.0}$ is the absorbance at 530.0 nm; $A_{600.0}$ is the absorbance at 600.0 nm; V is the total volume of the extracted solution; n is the dilution ratio; Mw is the molecular weight of cyanidin-3-glucoside (i.e., 449.4); ϵ is the molar extinction coefficient of anthocyanin ($29,600 \text{ M}^{-1} \text{ cm}^{-1}$); and m is the mass of the sample. The data are expressed as mean \pm standard deviation mg g^{-1} DW from biological triplicates.

2.6. Hyperspectral Imaging

A hyperspectral imaging camera (MicroHSI 410 SHARK, Corning Inc., Corning, NY, USA) was used. The detailed specifications of the hyperspectral camera used in this study are summarized in Table 1. As shown in Figure 2, the hyperspectral imaging system was equipped with eight halogen lamps ($15 \text{ W} \times 8$) as light sources; for data scanning, the hyperspectral camera was moved by the conveyor at the top of a dark chamber that blocked all external light. Considering the time and speed for a single scan, two or three samples were measured per one hyperspectral image at a moving speed of 100 mm s^{-1} .

Table 1. MicroHSI™ 410-SHARK specifications.

Item.	Specification
No. of spatial pixels	1408 spatial pixels
Focal Length, f-number	16 mm, f/1.4 standard
Spectral Range	400–1000 nm
Full FOV	28.6 degrees (500 mrad) standard

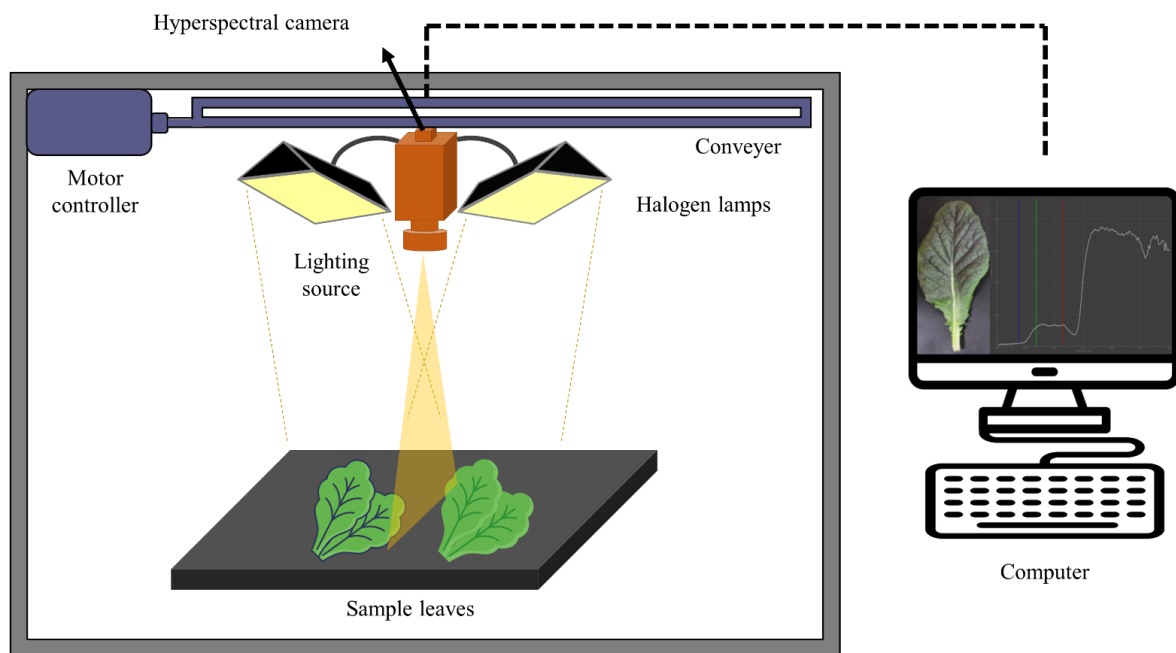


Figure 2. Hyperspectral imaging system (HSI) and extraction of spectral data from computer.

2.7. Data Processing and Prediction Models

The spectral data required for model development were extracted from the hyperspectral image data of *B. juncea* leaves, for which the spectral library in the Python 3.8 environment was used. After obtaining the RGB images of three bands (Blue: 456.55 nm, Red: 544.61 nm, and Green: 660.67 nm), the area of the sample for functional component analysis was chosen by polygon selection and set as the region of interest (ROI) (Figure 3). An area of 1 cm width on the leaf edge was not included in the ROI or used in the analysis, as it was presumed to contain much background noise. Further, the area of the leaf vein was also excluded in the experiments due to the relatively low content of functional components. The average spectra of the ROI across all samples was used for PLSR model development after checking the correlation with each functional component based on absorbance wavelengths.



Figure 3. Extraction of the region of interest (ROI) by polygon selection.

To predict each functional component, the extracted average spectra were applied to the Unscrambler X software v.11 (CAMO, Oslo, Norway) for model development through PLSR with full cross validation. The maximum number of PLS components used was limited up to 20. Various data pre-processing protocols were combined with PLSR and tested to ensure outstanding prediction performance of the developed model with reduced

noise signal in the spectral data, including the Savitzky–Golay filter (S. G. filter) with 3 or 7 smoothing points, standard normal variate (SNV), multiplicative scatter correction (MSC), mean normalization, 1st-order derivative (1st-Der), and 2nd-order derivative (2nd-Der). The applied pre-processing combinations are listed in Table 2.

Table 2. Pre-processing combinations tested in this study.

Methods	Pre-Processing Conditions
1	Raw data
2	Raw data, S. G. filter (interval = 3)
3	Raw data, S. G. filter (interval = 7)
4	Raw data, S. G. filter (interval = 3), SNV
5	Raw data, S. G. filter (interval = 3), MSC
6	Raw data, 1st-Der
7	Raw data, 2nd-Der
8	Raw data, SNV, 1st-Der
9	Raw data, SNV, 2nd-Der
10	Raw data, Normalization

The performance of the PLSR-based prediction models developed in this study was evaluated based on the R^2 and root mean square errors (RMSE) of calibration and validation, respectively, using Equations (6) and (7). From the prediction models developed with various pre-processing combinations, the model exhibiting comparatively high R_V^2 and low RMSE for validation was selected.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

where y_i is the measured value of component obtained by analysis; \hat{y}_i is the value predicted by the model; \bar{y} is the mean value of component from the analysis; and n is the number of samples, which was a total of 55 numbers for this experiment.

2.8. Development of Visualization Software for Applying Predictive Models

Visualization software was produced and utilized so that the developed model could be recognized intuitively. PyQT5, OpenCV, Pillow, Spectral, Matplotlib, Numpy, Loguru, and Pandas libraries were used in a Python 3.8 environment. The GUI was constructed by using PyQT5. The predicted values of the components were calculated by multiplying the weights at each wavelength on the spectrum obtained at each pixel from hyperspectral image data. The predicted values for each pixel were visualized with a jet colormap. Color bar means the predicted value of functional component by PLSR models.

3. Results and Discussion

3.1. Analysis of Plant Pigments and Metabolites

Table 3 summarizes the results of the quantification of five main functional components found in the leaves of *B. juncea* (i.e., total chlorophyll, total carotenoid, total phenolic, total glucosinolate, and total anthocyanin content). With respect to total chlorophyll and total anthocyanin content, which are responsible for leaf color, the former was relatively low when the anthocyanin content was high. Total chlorophyll content ranged from 2.13 to 11.70 mg g⁻¹ DW, with a mean value of 6.33 ± 2.21 mg g⁻¹ DW. The amount of total carotenoids in *B. juncea* is generally low; it had the lowest mean value (0.91 mg g⁻¹ DW) among the five functional components under study herein. The total phenolic contents ranged from 2.11 to 9.56 mg g⁻¹ DW, with a mean of 4.85 ± 1.94 mg g⁻¹ DW. A common feature of the family Brassicaceae is the abundance of glucosinolate components; therefore,

B. juncea leaves exhibited the highest minimum, maximum, and mean values for total glucosinolate content among all five functional components analyzed. Anthocyanins are responsible for the red and blue color of *B. juncea* leaves, and the color intensity varies depending on the cultivation conditions. A high level of anthocyanins reportedly requires blue light, cool climate, and a large daily temperature range [34–36]. Consistently, in the field experiment, *B. juncea* cultivated in the cool outdoor field environment exhibited a relatively higher level of anthocyanin production, whereas *B. juncea* cultivated in the glass greenhouse under higher temperatures resulted in intense light-green-to-green leaves according to hyperspectral imaging. Thus, total anthocyanin content varied greatly from 0 to 33.80 mg g⁻¹ DW, with a mean of 5.41 ± 6.75 mg g⁻¹ DW.

Table 3. Minimum, maximum, and mean content for the five functional components in *B. juncea*.

Parameter	Min *	Max *	Mean *	Standard Deviation *
total chlorophylls ^a	2.13 *	11.70 *	6.33 *	2.21 *
total carotenoids ^a	0.21	1.59	0.91	0.29
total phenolics ^a	2.11	9.56	4.85	1.94
total glucosinolates ^b	8.62	52.89	24.49	9.41
total anthocyanins ^a	0	33.80	5.41	6.75

* Units of measurement, ^a: mg g⁻¹ DW; ^b: μmol g⁻¹ DW.

3.2. Average Spectra and Correlation Analysis

The average spectra are shown in Figure 4A. The graphs showing the correlation of the contents of the five functional components at each wavelength of spectral data are shown in Figure 4B–F. Except for total phenolic contents, the other four functional components studied herein exhibited the highest negative correlation coefficient in the range of 400–600 nm of the visible light region of the spectrum. Total phenolic content showed a positive correlation in the wavelength bands corresponding to blue and green regions. This accounted for the high total phenolic content in *B. juncea* leaves with a relative increase in green areas, and an increase in total chlorophylls, total carotenoids, total glucosinolates, and total anthocyanins with a relative reduction in green areas. The potential use of prediction models for functional components quantitation via PLSR analysis was verified using correlation coefficient values of approximately 0.5 for all components, evenly distributed across the wavelength bands.

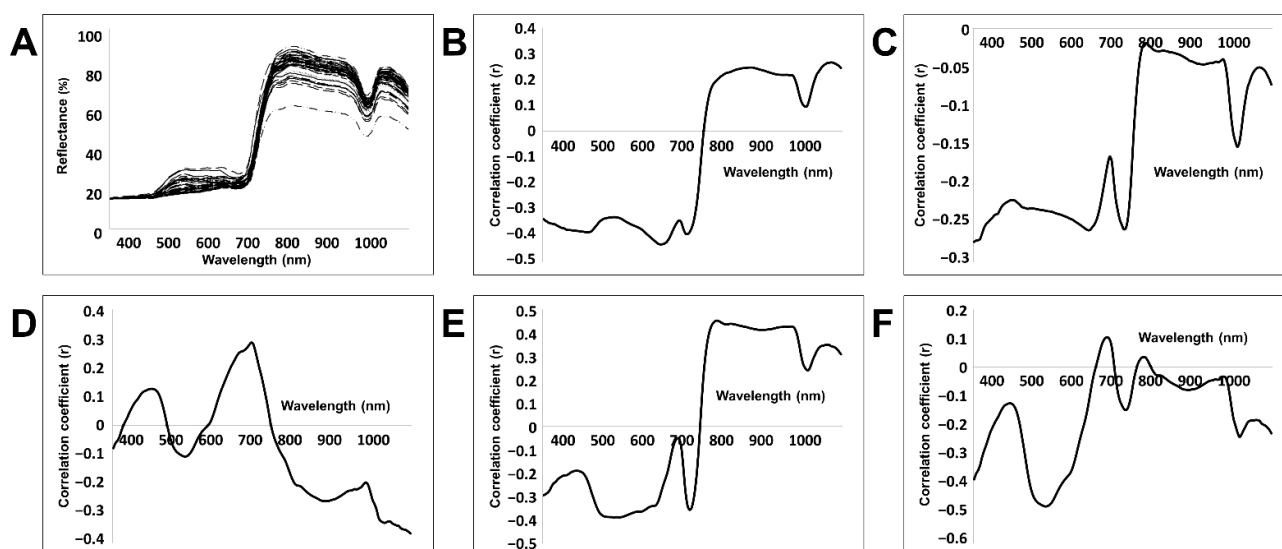


Figure 4. Average spectra (A) and correlation coefficient graphs for the contents of the main functional components present in leaves of *B. juncea* (total chlorophylls (B); total carotenoids (C); total phenolics (D); total glucosinolates (E); total anthocyanins (F)).

3.3. Development of PLSR Models Using Spectral Data Extracted from HSI

The PLSR prediction models for the prediction of five important functional components in plant leaf tissues were developed by applying various pre-processing combinations. The performance of each prediction model according to the pre-processing combination for each of the five functional components is shown in Table 4. Of all the five functional components studied, the highest prediction accuracy was recorded for total anthocyanins in almost all pre-processing methods. Notably, among the ten pre-processing combinations tested, the 6th combination registered R_V^2 and RMSEP values of 0.8273 and 2.4277, respectively, indicating the highest level of prediction performance. This may be because the distribution of anthocyanins, referred to as ‘purple magic’ by Kim et al. [37], accounted for the largest purple areas over the plant body. When dark-red areas on *B. juncea* leaves were visibly greater, the anthocyanin content was higher, which may have influenced the spectrum extracted from hyperspectral images. In turn, the highest prediction performance for total phenolic contents was registered for the 4th pre-processing combination, i.e., S. G. filter and SNV. Apart from these two components, the highest prediction performance for total chlorophyll, total carotenoid, and total glucosinolate content was observed for the 8th pre-processing conditions with SNV and 1st-Der. Pre-processing condition SNV is shared by the 4th and the 8th combinations, which is a well-known pre-processing method of normalization based on the standard deviation of the overall spectrum to eliminate the influence of light scattering. Presumably, the variability of the spectrum—which may arise from the vibration of the migration module while producing the hyperspectral image data—had been calibrated. In addition, pre-processing by 1st-Der, which is shared by the 6th and the 8th combinations, is effective in calibrating the baseline variations originating from the difference in relative intensity of the light sources. This is because the method differentiates the spectrum to place an emphasis on the changes within the absorption bands to amplify the spectral variation, while only the variation is shown. This is presumed to account for the better performance of the prediction model with the 8th pre-processing combination for total chlorophylls, total carotenoids, and total glucosinolates.

Table 4. PLSR prediction model outcomes on functional components for each pre-processing combination. The prediction model with the best performance among each pre-processing method was determined to have the lowest RMSEP values (marked in bold).

Pre-Processing Method	Total Chlorophylls				Total Carotenoids			
	Calibration		Validation		Calibration		Validation	
	R_C^2 *	RMSEC *	R_V^2 *	RMSEP *	R_C^2	RMSEC	R_V^2	RMSEP
1	0.6475	1.3538	0.4066	1.6701	0.6419	0.1267	0.4485	0.1525
2	0.6747	1.2651	0.4145	1.6100	0.6403	0.1270	0.4570	0.1514
3	0.6732	1.2437	0.3893	1.6406	0.6455	0.1272	0.4073	0.1574
4	0.7001	1.2236	0.4818	1.5154	0.6569	0.1056	0.4331	0.1343
5	0.6795	1.2340	0.4589	1.5475	0.6346	0.1077	0.4165	0.1363
6	0.6546	1.3105	0.4278	1.6300	0.5979	0.1086	0.3172	0.1431
7	0.2853	1.7787	0.2121	1.8574	0.2853	1.7787	0.2121	1.8574
8	0.6842	1.1832	0.5350	1.4045	0.6775	0.1040	0.4683	0.1315
9	0.2755	1.7886	0.2160	1.8528	0.1807	0.1500	0.1080	0.1555
10	0.6822	1.2316	0.4384	1.5721	0.6512	0.1087	0.3983	0.1398

Pre-Processing Method	Total Phenolics				Total Glucosinolates				Total Anthocyanins			
	Calibration		Validation		Calibration		Validation		Calibration		Validation	
	R_C^2	RMSEC	R_V^2	RMSEP	R_C^2	RMSEC	R_V^2	RMSEP	R_C^2	RMSEC	R_V^2	RMSEP
1	0.7657	0.9461	0.6625	1.1195	0.8587	3.4785	0.7429	4.5713	0.9296	1.5415	0.8194	2.5925
2	0.8322	0.7837	0.6955	1.0571	0.8525	3.5799	0.7365	4.6769	0.7903	2.6361	0.6861	3.2571
3	0.7620	0.9911	0.6258	1.2247	0.8352	3.8748	0.7226	4.8702	0.6910	3.3728	0.6255	3.7414
4	0.8204	0.8151	0.6909	1.0474	0.8465	3.6502	0.7635	4.4275	0.9312	1.5469	0.8378	2.6025
5	0.7563	0.9475	0.6213	1.1641	0.8481	3.6354	0.7350	4.6763	0.8835	1.9652	0.7871	2.7335
6	0.7869	0.8908	0.6398	1.1517	0.8237	3.8265	0.7119	4.8708	0.9144	1.6814	0.8273	2.4277
7	0.6466	1.0884	0.4591	1.3577	0.9591	1.8771	0.6729	5.2963	0.9273	1.8615	0.7718	3.3002
8	0.8157	0.8652	0.6983	1.0794	0.8508	3.4289	0.7827	4.0647	0.9072	2.1886	0.7938	3.1842
9	0.7981	0.8477	0.4419	1.4260	0.7414	4.7484	0.6175	5.7491	0.8060	2.6448	0.7150	3.2259
10	0.7659	0.9329	0.6647	1.0989	0.8552	3.5580	0.7419	4.9688	0.9102	1.7609	0.8034	2.7358

* R_C^2 — R^2 value in calibration; RMSEC—RMSE value in calibration; R_V^2 — R^2 value in validation; RMSEP—RMSE value in validation. Bold means the values showing the best performance.

Notably, the prediction models for total chlorophyll and total carotenoid content led to lower performance in prediction performance, with $R^2 < 0.3$ for the 7th and 9th pre-processing combinations sharing pre-processing by 2nd-Der, which shows the characteristics of the changes in spectral slope to effectively calibrate the baseline and remove any micro-noise that may appear in the system, along with pre-processing by 1st-Der. Nevertheless, it is possible to interpret that the removal of noise resulted in the removal of the effects of absorption bands created by micro-components, thereby lowering prediction performance. In previous studies, total phenolics had been predicted at $R_C^2 = 0.820$ and $R_V^2 = 0.551$ by Caporaso et al. [29], and total anthocyanins had been predicted at $R_C^2 = 0.883$ and $R_V^2 = 0.830$ by Liu et al. [28]. However, the results reported herein were higher, at $R_C^2 = 0.8204$ and $R_V^2 = 0.6909$ for total phenolics, and at $R_C^2 = 0.9144$ and $R_V^2 = 0.8273$ for total anthocyanins. Furthermore, the 4th and 8th pre-processing combinations in the PLSR models for the five functional components resulted in the most outstanding prediction performance, implying that the use of pre-processing with SNV in the prediction spectrum for the five components is essential for the development of high-prediction performance models. Figure 5 depicts the results of the prediction model with the highest prediction performance among all models developed. The selected models did not have a very high prediction rate for functional components; nevertheless, the results verify their potential for predicting the concentrations of multiple functional components in actual cultivation conditions of *B. juncea*. In addition, the development of other regression models such as machine learning, mixed model, or principal components analysis have the potential to lead better predictive performance. It is judged that the combination of pre-processing methods used in this study is sufficiently worthy of reference for application to other regression models.

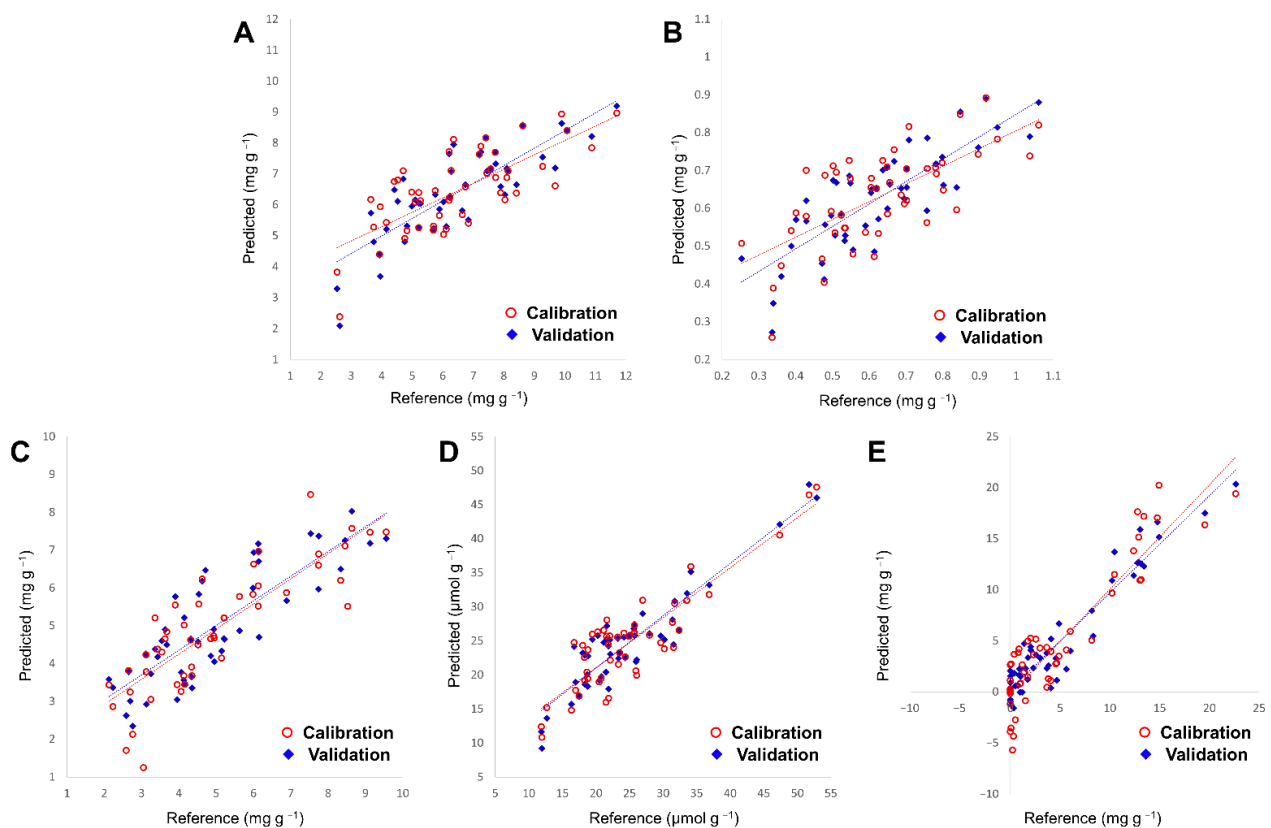


Figure 5. Prediction models for each functional component: (A) total chlorophylls, (B) total carotenoids, (C) total phenolics, and (D) total glucosinolates with the 8th pre-processing combination; (E) total anthocyanins with the 4th pre-processing combination. Fifth-five samples were used to establish each prediction model.

Hyperspectral imagery spectroscopy is a technology that uses spectroscopic techniques and imaging which can offer the sample's optical information. Therefore, even if the amounts of components are the same, the spectral properties will not be the same if the distribution of these components in the plant are different. Thus, direct application of models in this study to other plants is difficult. New learning data from other plants are required to make prediction model of new plants. However, it is noteworthy that the kind of pre-processing method and the methodology presented in this paper can be applied to new plants to develop prediction models for multiple components quickly. These trials can expand the scientific meaning and industrial uses related to engineering based on a hyperspectral imaging system.

3.4. Application of the Functional Component Prediction Model for Visualization

For the prediction models developed for the five plant functional components under study, visualization could probably be achieved in the form of a distribution map based on the prediction of component values by the prediction model from the spectrum values in the unit of pixels obtained from the hyperspectral image. Figure 6 shows such visualization based on PLSR prediction models for each pixel representation of the concentrations of functional components according to the colors on the color map. Hence, the variation of different colors from high to low concentrations of functional components in the leaf can be detected intuitively with respect to the distribution using the color map. Similarly, cultivation in the open air under actual sunlight conditions is likely to allow enhanced accuracy of the prediction model with training data, which in turn is likely to enable monitoring of the functional components during cultivation. This should enable farmers to determine the best time for harvest based on the prediction of key functional components.

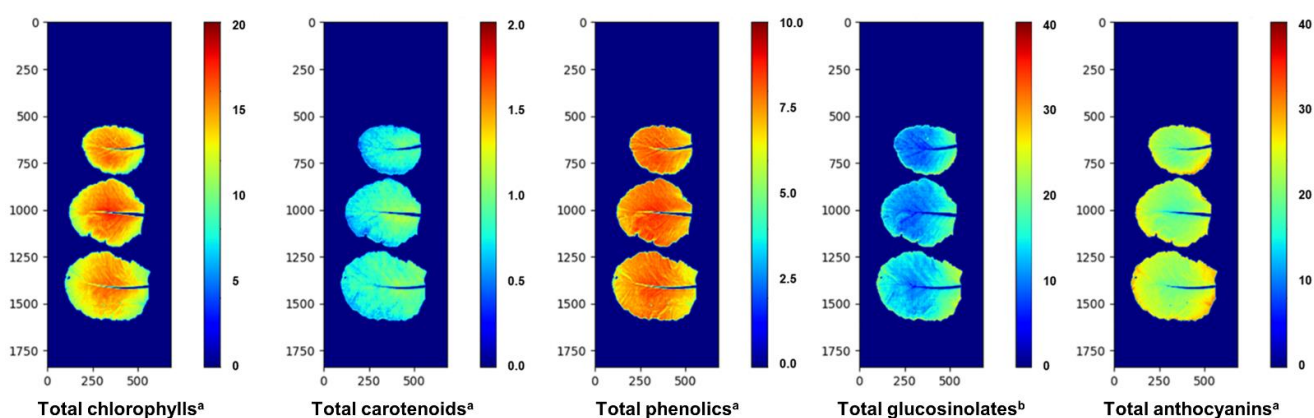


Figure 6. Distribution maps of plant functional components predicted by the developed PLSR models. X- and Y-axes indicate the size of images in pixels. Color bar means the predicted value of functional component by the developed PLSR model (units of measurement, ^a: mg g⁻¹ DW; ^b: μmol g⁻¹ DW).

4. Conclusions

Using hyperspectral imaging, PLSR models were developed for the prediction of the multiple contents of five key functional components in the leaves of *B. juncea*: chlorophylls, carotenoids, phenolics, glucosinolates, and anthocyanins. To develop the models, the region of interest for analysis of the functional components was uniformly chosen by polygon selection from the hyperspectral image data, and the average spectra were extracted. Various pre-processing combinations were applied to the spectral data to develop a model which showed the most outstanding prediction performance. The resulting PLSR prediction models had $R^2 \geq 0.8$ for total phenolic, total glucosinolate, and total anthocyanin contents, and valid models were thus obtained for each functional component. In addition, the models for total chlorophylls and total carotenoids had R_C^2 of 0.6842 and 0.6775, respectively, which implied the potential for development of more efficient models via further data processing in the future. Overall, among the ten pre-processing condi-

tions tested here, the 8th pre-processing condition combining SNV and 1st-Der resulted in the highest performance of prediction for functional components. Additionally, the 4th combination containing SNV exhibited outstanding prediction performance for total phenolic content. Hence, the most efficient pre-processing condition for the development of a prediction model exhibiting a high level of performance was SNV, shared by the 4th and 8th pre-processing combinations. In addition, the ‘multiple-chemical’ analysis from one spectral-image data could be effectively performed. Considering the application of the models developed here, it is possible to draw a distribution map for functional components by applying the spectral data in pixels from the hyperspectral images to the prediction models. The variation of different colors from high to low concentrations of functional components in the leaf can thus be detected intuitively with respect to their distribution, as the data for the content value of each functional component was associated with the color map. Based on our findings, the application of the model to measure hyperspectral images under actual cultivation in natural sunlight conditions is likely to allow real-time monitoring of the components of interest. Furthermore, the prediction for the functional components will contribute to accurate determination of the best time for harvest.

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

Funding: This work was supported by the Korean Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and by the Korean Smart Farm R&D Foundation (KosFarm) through the Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA); the Ministry of Science and ICT (MSIT); and the Rural Development Administration (RDA) (421034-04).

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kumar, V.; Thakur, A.K.; Barothia, N.D.; Chatterjee, S.S. Therapeutic potentials of *Brassica juncea*: An overview. *Cell Med.* **2011**, *1*, 2.1–2.16. [[CrossRef](#)]
2. Tian, Y.; Deng, F. Phytochemistry and biological activity of mustard (*Brassica juncea*): A review. *CyTA-J. Food* **2020**, *18*, 704–718. [[CrossRef](#)]
3. Malabed, R.S.; Noel, M.G.; Aton, B.C., III; Toribio, E.A.F. Characterization of the glucosinolates and isothiocyanates in mustard (*Brassica juncea* L.) extracts and determination of its myrosinase activity and antioxidant capacity. In Proceedings of the De La Salle University Research Congress 2014, Manila, Philippines, 6–8 March 2014; pp. 1–7.
4. Oulad El Majdoub, Y.; Alibrando, F.; Cacciola, F.; Arena, K.; Pagnotta, E.; Matteo, R.; Micalizzi, G.; Dugo, L.; Dugo, P.; Mondello, L. Chemical characterization of three accessions of *Brassica juncea* L. extracts from different plant tissues. *Molecules* **2020**, *25*, 5421. [[CrossRef](#)]
5. Cartea, M.E.; Francisco, M.; Soengas, P.; Velasco, P. Phenolic compounds in *Brassica* vegetables. *Molecules* **2010**, *16*, 251–280. [[CrossRef](#)] [[PubMed](#)]
6. Manohar, P.R.; Pushpan, R.; Rohini, S. Mustard and its uses in Ayurveda. *Indian J. Tradit. Knowl.* **2009**, *8*, 400–404.
7. Lee, J.E.; Kim, A.J. Antioxidant activity, whitening and anti-wrinkle effects of leaf and seed extracts of *Brassica juncea* L. *Asian J. Beauty Cosmetol.* **2020**, *18*, 283–295. [[CrossRef](#)]
8. Méthy, M.; Olioso, A.; Trabaud, L. Chlorophyll fluorescence as a tool for management of plant resources. *Remote Sens. Environ.* **1994**, *47*, 2–9. [[CrossRef](#)]
9. Jezek, M.; Zörb, C.; Merkt, N.; Geilfus, C.M. Anthocyanin management in fruits by fertilization. *J. Agric. Food Chem.* **2018**, *66*, 753–764. [[CrossRef](#)]
10. Mounien, L.; Tourniaire, F.; Landrier, J.F. Anti-obesity effect of carotenoids: Direct impact on adipose tissue and adipose tissue-driven indirect effects. *Nutrients* **2019**, *11*, 1562. [[CrossRef](#)]
11. Kawser Hossain, M.; Abdal Dayem, A.; Han, J.; Yin, Y.; Kim, K.; Kumar Saha, S.; Yang, G.M.; Choi, H.Y.; Cho, S.G. Molecular mechanisms of the anti-obesity and anti-diabetic properties of flavonoids. *Inter. J. Mol. Sci.* **2016**, *17*, 569. [[CrossRef](#)] [[PubMed](#)]
12. Kim, M.H. Antioxidant and antibacterial activity of extracts from *Brassica juncea* Czerniak et coss., *Celosia cristata* L., and *Beta vulgaris* L. *J. Korean Soc. Food Cult.* **2012**, *27*, 719–729. [[CrossRef](#)]

13. Nawaz, H.; Shad, M.A.; Muzaffar, S. Phytochemical composition and antioxidant potential of Brassica. In *Brassica Germplasm: Characterization, Breeding and Utilization*; IntechOpen: London, UK, 2018; Volume 1, pp. 7–26. [[CrossRef](#)]
14. Okulicz, M. Multidirectional time-dependent effect of sinigrin and allyl isothiocyanate on metabolic parameters in rats. *Plant Foods Hum. Nutr.* **2010**, *65*, 217–224. [[CrossRef](#)]
15. Zhang, Y. Allyl isothiocyanate as a cancer chemopreventive phytochemical. *Mol. Nutr. Food Res.* **2010**, *54*, 127–135. [[CrossRef](#)] [[PubMed](#)]
16. Luciano, F.B.; Holley, R.A. Enzymatic inhibition by allyl isothiocyanate and factors affecting its antimicrobial action against *Escherichia coli*. *Inter. J. Food Microbiol.* **2009**, *131*, 240–245. [[CrossRef](#)]
17. Mo, C.; Hasegawa, M.; Lee, K.; Lim, J.G.; Kim, M.S.; Kang, S.; Lee, H.D.; Bae, H.; Kim, D.Y.; Cho, B.K. Development of a non-destructive on-line pungency measurement system for red-pepper powder. *J. Fac. Agric. Kyushu Univ.* **2013**, *58*, 137–144. [[CrossRef](#)]
18. Teye, E.; Huang, X.; Sam-Amoah, L.K.; Takrama, J.; Boison, D.; Botchway, F.; Kumi, F. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chem.* **2015**, *176*, 403–410. [[CrossRef](#)]
19. Sunoj, S.; Igathinathane, C.; Visvanathan, R. Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Comput. Electron. Agric.* **2016**, *124*, 234–242. [[CrossRef](#)]
20. Zhao, Y.R.; Li, X.; Yu, K.Q.; Cheng, F.; He, Y. Hyperspectral imaging for determining pigment contents in cucumber leaves in response to angular leaf spot disease. *Sci. Rep.* **2016**, *6*, 27790. [[CrossRef](#)]
21. Hernández-Hierro, J.M.; Esquerre, C.; Valverde, J.; Villacreces, S.; Reilly, K.; Gaffney, M.; González-Miret, M.L.; Heredia, F.J.; O'Donnell, C.P.; Downey, G. Preliminary study on the use of near infrared hyperspectral imaging for quantitation and localisation of total glucosinolates in freeze-dried broccoli. *J. Food Eng.* **2014**, *126*, 107–112. [[CrossRef](#)]
22. Onivogui, G.; Zhang, H.; Mlyuka, E.; Diaby, M.; Song, Y. Chemical composition, nutritional properties and antioxidant activity of monkey apple (*Anisophyllea laurina* R. Br. ex Sabine). *J. Food Nutr. Res.* **2014**, *2*, 281–287. [[CrossRef](#)]
23. Ihsan, M.; Saputro, A.H.; Handayani, W. Flavonoid distribution mapping system of velvet apple leaf based on hyperspectral imaging. In Proceedings of the 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), Bali, Indonesia, 12–13 October 2019; pp. 157–162.
24. He, J.; Chen, L.; Chu, B.; Zhang, C. Determination of total polysaccharides and total flavonoids in *Chrysanthemum morifolium* using near-infrared hyperspectral imaging and multivariate analysis. *Molecules* **2018**, *23*, 2395. [[CrossRef](#)]
25. Cheng, J.H.; Sun, D.W. Rapid quantification analysis and visualization of *Escherichia coli* loads in grass carp fish flesh by hyperspectral imaging method. *Food Bioproc. Technol.* **2015**, *8*, 951–959. [[CrossRef](#)]
26. Choi, J.Y.; Kim, J.; Kim, J.; Jeong, S.; Kim, M.; Park, S.; Moon, K.D. Hyperspectral imaging technique for monitoring moisture content of blueberry during the drying process. *Korean J. Food Preserv.* **2021**, *28*, 445–455. [[CrossRef](#)]
27. Bai, X.; Xiao, Q.; Zhou, L.; Tang, Y.; He, Y. Detection of sulfite dioxide residue on the surface of fresh-cut potato slices using near-infrared hyperspectral imaging system and portable near-infrared spectrometer. *Molecules* **2020**, *25*, 1651. [[CrossRef](#)] [[PubMed](#)]
28. Liu, Y.; Sun, Y.; Xie, A.; Yu, H.; Yin, Y.; Li, X.; Duan, X. Potential of hyperspectral imaging for rapid prediction of anthocyanin content of purple-fleshed sweet potato slices during drying process. *Food Anal. Methods* **2017**, *10*, 3836–3846. [[CrossRef](#)]
29. Caporaso, N.; Whitworth, M.B.; Fowler, M.S.; Fisk, I.D. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chem.* **2018**, *258*, 343–351. [[CrossRef](#)]
30. Lichtenthaler, H.K.; Buschmann, C. Chlorophylls and carotenoids: Measurement and characterization by UV-VIS spectroscopy. *Curr. Protoc. Food Anal. Chem.* **2001**, *1*, F4.3.1–F4.3.8. [[CrossRef](#)]
31. Thomas, M.; Badr, A.; Desjardins, Y.; Gosselin, A.; Angers, P. Characterization of industrial broccoli discards (*Brassica oleracea* var. *italica*) for their glucosinolate, polyphenol and flavonoid contents using UPLC MS/MS and spectrophotometric methods. *Food Chem.* **2018**, *245*, 1204–1211. [[CrossRef](#)]
32. Mawlong, I.; Sujith Kumar, M.; Gurung, B.; Singh, K.; Singh, D. A simple spectrophotometric method for estimating total glucosinolates in mustard de-oiled cake. *Inter. J. Food Prop.* **2017**, *20*, 3274–3281. [[CrossRef](#)]
33. Yang, Y.C.; Sun, D.W.; Pu, H.; Wang, N.N.; Zhu, Z. Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biol. Technol.* **2015**, *103*, 55–65. [[CrossRef](#)]
34. Jeong, J.C.; Kim, S.J.; Hong, S.Y.; Nam, J.H.; Sohn, H.B.; Kim, Y.H.; Mekapogu, M. Growing environment influence the anthocyanin content in purple-and red-fleshed potatoes during tuber development. *Korean J. Crop. Sci.* **2015**, *60*, 231–238. [[CrossRef](#)]
35. Ninu, L.; Ahmad, M.; Miarelli, C.; Cashmore, A.R.; Giuliano, G. Cryptochrome 1 controls tomato development in response to blue light. *Plant J.* **1999**, *18*, 551–556. [[CrossRef](#)]
36. Gliberto, L.; Perrotta, G.; Pallara, P.; Weller, J.L.; Fraser, P.D.; Bramley, P.M.; Fiore, A.; Tavazza, M.; Giuliano, G. Manipulation of the blue light photoreceptor cryptochrome 2 in tomato affects vegetative development, flowering time, and fruit antioxidant content. *Plant Physiol.* **2005**, *137*, 199–208. [[CrossRef](#)] [[PubMed](#)]
37. Kim, H.S.; Yoo, J.H.; Park, S.H.; Kim, J.S.; Chung, Y.; Kim, J.H.; Kim, H.S. Measurement of environmentally influenced variations in anthocyanin accumulations in *Brassica rapa* subsp. *Chinensis* (Bok Choy) using hyperspectral imaging. *Front. Plant Sci.* **2021**, *12*, 693854. [[CrossRef](#)]