




Article

Near-Infrared Hyperspectral Imaging (NIR-HSI) for Nondestructive Prediction of Anthocyanins Content in Black Rice Seeds

Hanim Z. Amanah ^{1,2}, Collins Wakholi ¹, Mukasa Perez ¹, Mohammad Akbar Faqeerzada ¹,
Salma Sultana Tunny ¹ , Rudiati Evi Masithoh ², Myoung-Gun Choung ³, Kyung-Hwan Kim ⁴,
Wang-Hee Lee ^{1,5}  and Byoung-Kwan Cho ^{1,5,*} 

- ¹ Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea; hanim_za@ugm.ac.id (H.Z.A.); wcoln@yahoo.com (C.W.); piremuk@yahoo.com (M.P.); akbar.faqeerzada@gmail.com (M.A.F.); sstunny97@gmail.com (S.S.T.); wanghee@cnu.ac.kr (W.-H.L.)
- ² Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Gadjah Mada University, Yogyakarta 55281, Indonesia; evi@ugm.ac.id
- ³ Department of Herbal Medicine Resource, Dogye Campus, Kangwon National University, Samcheok 25949, Korea; cmg7004@kangwon.ac.kr
- ⁴ Department of Gene Engineering, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Korea; biopiakim@korea.kr
- ⁵ Department of Smart Agriculture Systems, College of Agricultural and Life Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea
- * Correspondence: chobk@cnu.ac.kr; Tel.: +82-42-8216-715



Citation: Amanah, H.Z.; Wakholi, C.; Perez, M.; Faqeerzada, M.A.; Tunny, S.S.; Masithoh, R.E.; Choung, M.-G.; Kim, K.-H.; Lee, W.-H.; Cho, B.-K. Near-Infrared Hyperspectral Imaging (NIR-HSI) for Nondestructive Prediction of Anthocyanins Content in Black Rice Seeds. *Appl. Sci.* **2021**, *11*, 4841. <https://doi.org/10.3390/app11114841>

Academic Editors: Daniel Cozzolino and Koen Janssens

Received: 20 March 2021
Accepted: 20 May 2021
Published: 25 May 2021

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



Copyright: © 2021 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/>).

Abstract: Anthocyanins are an important micro-component that contributes to the quality factors and health benefits of black rice. Anthocyanins concentration and compositions differ among rice seeds depending on the varieties, growth conditions, and maturity level at harvesting. Chemical composition-based seeds inspection on a real-time, non-destructive, and accurate basis is essential to establish industries to optimize the cost and quality of the product. Therefore, this research aimed to evaluate the feasibility of near-infrared hyperspectral imaging (NIR-HSI) to predict the content of anthocyanins in black rice seeds, which will open up the possibility to develop a sorting machine based on rice micro-components. Images of thirty-two samples of black rice seeds, harvested in 2019 and 2020, were captured using the NIR-HSI system with a wavelength of 895–2504 nm. The spectral data extracted from the image were then synchronized with the rice anthocyanins reference value analyzed using high-performance liquid chromatography (HPLC). For comparison, the seed samples were ground into powder, which was also captured using the same NIR-HSI system to obtain the data and was then analyzed using the same method. The model performance of partial least square regression (PLSR) of the seed sample developed based on harvesting time, and mixed data revealed the model consistency with R^2 over 0.85 for calibration datasets. The best prediction models for 2019, 2020, and mixed data were obtained by applying standard normal variate (SNV) pre-processing, indicated by the highest coefficient of determination (R^2) of 0.85, 0.95, 0.90, and the lowest standard error of prediction (SEP) of 0.11, 0.17, and 0.16 mg/g, respectively. The obtained R^2 and SEP values of the seed model were comparable to the result of powder of 0.92–0.95 and 0.09–0.15 mg/g, respectively. Additionally, the obtained beta coefficients from the developed model were used to generate seed chemical images for predicting anthocyanins in rice seed. The root mean square error (RMSE) value for seed prediction evaluation showed an acceptable result of 0.21 mg/g. This result exhibits the potential of NIR-HSI to be applied in a seed sorting machine based on the anthocyanins content.

Keywords: micro-component; black rice; seed quality; near-infrared; hyperspectral imaging

1. Introduction

Rice (*Oryza sativa* L.) is a vital cereal food in Asia and is consumed by almost half of the world's population [1]. The chemical composition of rice is characterized by a high content of carbohydrates, mainly starch (56–74%), a moderate content of protein (8–11%), and a minor content of lipids (2–4%) and minerals (1–3%) [2]. There are many different kinds of rice, including white rice, which is widely consumed, and pigmented rice. Pigmented rice, such as purple and black rice, contains important bioactive compounds beneficial for human health. The bran of black rice is rich in fiber and many kinds of phytochemicals, such as tocopherols, tocotrienols, oryzanols, vitamin B complex, and other phenolic compounds [3,4]. The protein content of black rice was also reported to be higher than white rice. In addition, black rice contains anthocyanins and proanthocyanidin in its aleurone layer, meaning that black rice is considered an essential health product.

Anthocyanins are a water-soluble flavonoid compound that is responsible for the appearance of certain colors in nature and may appear as red, purple, or blue, depending on the pH. This color change is due to the reversible structural transformation of anthocyanins in a certain range of pH values [5,6]. Anthocyanins were reported to be an essential component of traditional herbal medicine [7], and have a wide range of pharmacological applications against various stress conditions and chronic diseases, e.g., inflammation, cognitive decline, neural dysfunction, capillary fragility and permeability, platelet aggregation, cardiovascular complication, liver damage, lipid peroxidation, and cancer tumor growth [8]. Moreover, in vivo tests reported that supplementation of black rice containing anthocyanins in the diet reduces atherosclerotic lesions in hypercholesterolemic rabbits and apolipoprotein-E-deficient mice. These biological activities were related to anthocyanins' potencies as antiradical scavenging and antioxidant activities [9].

The anthocyanins content in black rice is greatly affected by genotype and the environmental growth condition [10] and reported ranging from 0.52 to 3.47 mg/g. Cyanidin-3-glucoside (C3G) and peonidin-3-glucoside (Pn3G) are the most common forms of anthocyanins found in black rice [11]. Other forms of anthocyanins observed in black rice are cyanidin-3,5-diglucoside, cyanidin-3-rutinoside, cyanidin-3-gentiobioside, malvidin-3-glucoside, and peonidin-3-rutinoside [12,13]. Anthocyanins are distributed in all parts of black rice seeds but are mostly accumulated in the bran. The anthocyanins level in bran is more than fifteen times higher than that in the embryo. Furthermore, the C3G and Pn3G content in bran were reported as being 98% and 93% of their total content in whole black rice, whereas only 2% and 7% were available in embryos, respectively [14–16].

Chromatography techniques, including paper chromatography and high-performance liquid chromatography (HPLC), have extensively studied the qualitative and quantitative chemical components, including anthocyanins, in agricultural products [17]. These techniques are a destructive method in sample preparation and require chemicals solvent for the pre-analysis of extraction, which is time-consuming, costly, and labor-intensive. In addition, these methods also generate chemical waste, which is harmful to the environment. Moreover, the large-scale detection of anthocyanins attributed to a single seed is quite difficult with these techniques. Thus, chemical-based seed selection for the quality control in factories remains a challenging process. Hence, the development of non-destructive methods related to spectroscopy and image processing analysis allows for overcoming the drawbacks mentioned above.

Near-infrared and mid-infrared spectroscopy methods have been widely used to detect anthocyanins in agricultural products such as soybeans [18], flowering tea [19], and sweet cherries [20]. Fourier transform near-infrared (FT-NIR) and Fourier transform infrared (FT-IR) can also be applied to predict anthocyanins content in black soybean based on the seed's spectra [21]. Many researchers have favored the hyperspectral imaging (HSI) system, which integrates the spectroscopic and imaging techniques due to the capability in acquiring spectral and spatial information simultaneously. In recent years, HSI has been adopted to assess chlorogenic acid content in *Flos lonicerae* [22], to determine deoxynivalenol in bulk wheat kernel [23], and to monitor the quality of strawberries during storage [24].

The potential capability of HSI to predict macro-components in grain, such as starch in corn and protein in wheat, based on single kernel data, has also been proven [25,26]. For micro-component, particularly in anthocyanins detection, visible hyperspectral imaging (Vis-HSI) has been applied to slices of purple potato [27,28] and Lychee pericarp [29]. PLSR models based on the full wavelength in the lychee pericarp sample resulted in performance over 0.9 for both calibration and prediction models, while the pre-processing method (SNV, MSC, and SG) slightly improves the performance model to predict anthocyanins in purple-fleshed sweet potato slices from 0.876 to 0.889. Furthermore, this hyperspectral method also successfully visualizes the anthocyanins distribution throughout the sample.

HSI in the NIR region also has been used to predict micro-component, including total anthocyanins, flavonoids, and phenolics of dry black goji berries (*Lycium ruthenicum* Murr.) [30]. The PLS regression model using a full range of spectra (975–1646 nm) resulted in the performance of the calibration model of total anthocyanins and phenolics over 0.90 and 0.85 for total flavonoids. This technique also resulted in high performance of the prediction model of 0.89, 0.84, and 0.85 for anthocyanins, flavonoids, and phenolics, respectively. In addition, the result of the PLS regression method was comparable to the other techniques, including LS-SVM and CNN, that were applied in the study. However, in this sample, anthocyanins were uniformly distributed throughout the sample, and the content was higher than those in black rice.

Anthocyanins are a micro-component in rice, mainly available in its aleurone, and are not distributed uniformly throughout the rice seed. Anthocyanins produce a dark color (purple to black) in the seed surface that may absorb the light emitted from the instrument. However, most of these chemicals are present on the rice seed's surface, which might indicate one of the advantages of measurement using a non-destructive system, which works based on the interaction between light and matter. At present, no research has been done on the application of near-infrared hyperspectral imaging (NIR-HSI) for evaluating anthocyanins content in a single seed of rice. Perhaps, the lack of such studies is due to the low concentration of anthocyanins in pigmented rice, and the market demand for this rice is not high. However, with the gradual increase in functional food demand in the last decade, the research interest related to anthocyanins in rice has risen. Thus, this study aims to evaluate the potential of near-infrared hyperspectral imaging (NIR-HSI) for predicting anthocyanins in individual rice seeds. For comparison, this study also developed a model for the prediction of anthocyanins based on the powder sample of the same rice variety.

2. Materials and Methods

2.1. Rice Sample

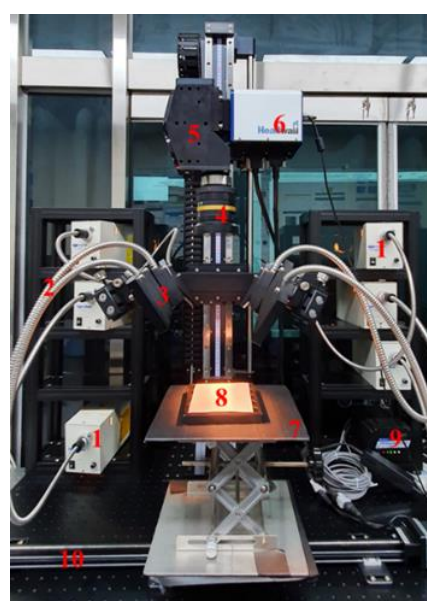
Thirty-two samples of black rice (listed in Table 1) were obtained from Rural Development Administration (RDA), Jeollabuk-do, South Korea, harvested in 2019 and 2020. Fourteen varieties were grown in both 2019 and 2020, two rice types (Heugjinju Byeo and Cheongpungheugchal) were only cultivated in 2019, and two varieties (Gancheok Byeo and Cheonghaejinmi) only harvested in 2020. The weight of 100 seeds of rice ranged from 1.46 to 2.44 g, indicating rice size variability based on the variety. The rice seed's color and size within a single variety were also varied, which resulted in the possibility of different anthocyanins contents among different seeds. Therefore, the variability of seeds in the biochemical composition will result in the variation of spectral data extracted from the image of HSI data. On the other hand, investigating anthocyanins based on a single seed is constrained by the minimum weight requirement of the HPLC procedure. Hence, in this research, twenty grams of rice seeds were collected from each rice variety of each year, and among them, 36 seeds were selected randomly for the image data acquisition using the NIR-HSI system. After acquiring the image, all of the seeds were ground to obtain a uniform particle size under 250 μm . The powder sample was then placed on the sample plate and scanned using the same NIR-HSI system for data acquisition. After both seed and powder image acquisition, the powder samples were then sent for chemical analysis using the HPLC method.

Table 1. List of examined black rice varieties of two years harvesting time.

No	Harvesting Time	
	2019	2020
1	Jejubukjeju-2002-561	Jejubukjeju-2002-561
2	Heuginju Byeo	Gancheok Byeo
3	Heugnam Byeo	Heugnam Byeo
4	Heugkwang Byeo	Heugkwang Byeo
5	Sinmyungheugchal	Sinmyungheugchal
6	Heugseol	Heugseol
7	Seonhyangheukmi	Seonhyangheukmi
8	Boseogheugchal	Boseogheugchal
9	Sinnongheugchal	Sinnongheugchal
10	Josaengheugchal	Josaengheugchal
11	Cheongpungheugchal	Chungnam 2 ho
12	Chungnam 2 ho	Heugsujeong
13	Heugsujeong	Jeughyangchal 1 ho
14	Heughyangchal 1 ho	Jahyangna 861
15	Jahyangna 861	Seonhyangheugmi
16	Seonhyangheugmi	Cheonghaejinmi

2.2. NIR Hyperspectral Imaging System

A laboratory-based line scan NIR-HSI system (Figure 1) was used to scan and collect the sample data in reflectance mode. The system consists of six 100 W tungsten-halogen line-light sources (Light Bank, Ushio Inc., Tokyo, Japan) connected to the optical fiber used to illuminate the sample during measurement. The sensing module was composed of a line-scan spectrograph (NIR, Headwall Photonics, Fitchburg, MA, USA) that covered a spectral range of 895–2504 nm, a mercury cadmium telluride (MCT) detector (Model: Xeva-2.5–320; Xenics, Heverlee, Belgium) to detect light reflected by the sample, as well as a high-performance camera (Headwall Photonics, Fitchburg, MA, USA) with 320 (spatial) × 256 (spectral) pixel resolution and an objective lens (focal length 25 mm f/1.4). During data acquisition, the sample plate was placed on the motorized table to move the sample through the camera's capture view at a particular speed controlled by a DC motor. The software for data acquisition was developed using Microsoft Visual Basic (version 6.0) on a Windows platform.



NIR-HSI component

1. Light sources
2. Fiber optic
3. Light
4. Camera lens
5. Detector
6. Spectrograph
7. Motorized table
8. Sample plate
9. DC motor
10. Conveyor

Figure 1. Laboratory-based scan of near-infrared (NIR) hyperspectral imaging system.

2.3. NIR-HSI Data Acquisition and Extraction

The images of the samples (I_{raw}) were captured using the NIR-HSI system based on the arrangement as follows. Two rice varieties (36 seeds of each) were arranged on the 10×10 grid black square plate, transferred to the moving table controlled by a stepping motor, and scanned line by line using the HSI system. In total, there were 16 scanned plates for 32 rice varieties. The powder was placed into a sample holder plate containing 9 samples of each and adjusted to 5 mm thickness for HSI scanning. The sample arrangement process of powder was carried out by flapping the sample surface using a rod stick without compression. The distance between the sample and the camera was set as 25 cm in order to cover the spatial range of the sample. The spatial and spectral data were obtained when the sample passed the camera field of view (FOV) and were saved in a 3D hypercube containing two spatial dimensions and one spectral dimension. The dark (I_{dark}) and white (I_{white}) reference images were acquired to calculate the reflectance value and to correct noise which may come from the environment. The dark reference image (0% reflectance) was collected by covering the camera lens using an opaque cap and turning off the light. In comparison, the white reference (99% reflectance) was acquired by using a white Teflon sheet. The HSI corrected image (I_c) was calculated based on the HSI raw data image (I_{raw}) using Equation (1), as follows:

$$I_c = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}} \quad (1)$$

The sample spectral information was extracted from the corrected image (I_c) after removing the background by applying the region of interest (ROI) selection step. Both processes, image correction and background removal, were conducted using MATLAB software (The Math Works, Natick, MA, USA, R2019a). To simplify the image correction and spectral extraction process and to explain the data analysis strategy, the flow chart of this process is presented in Figure 2.

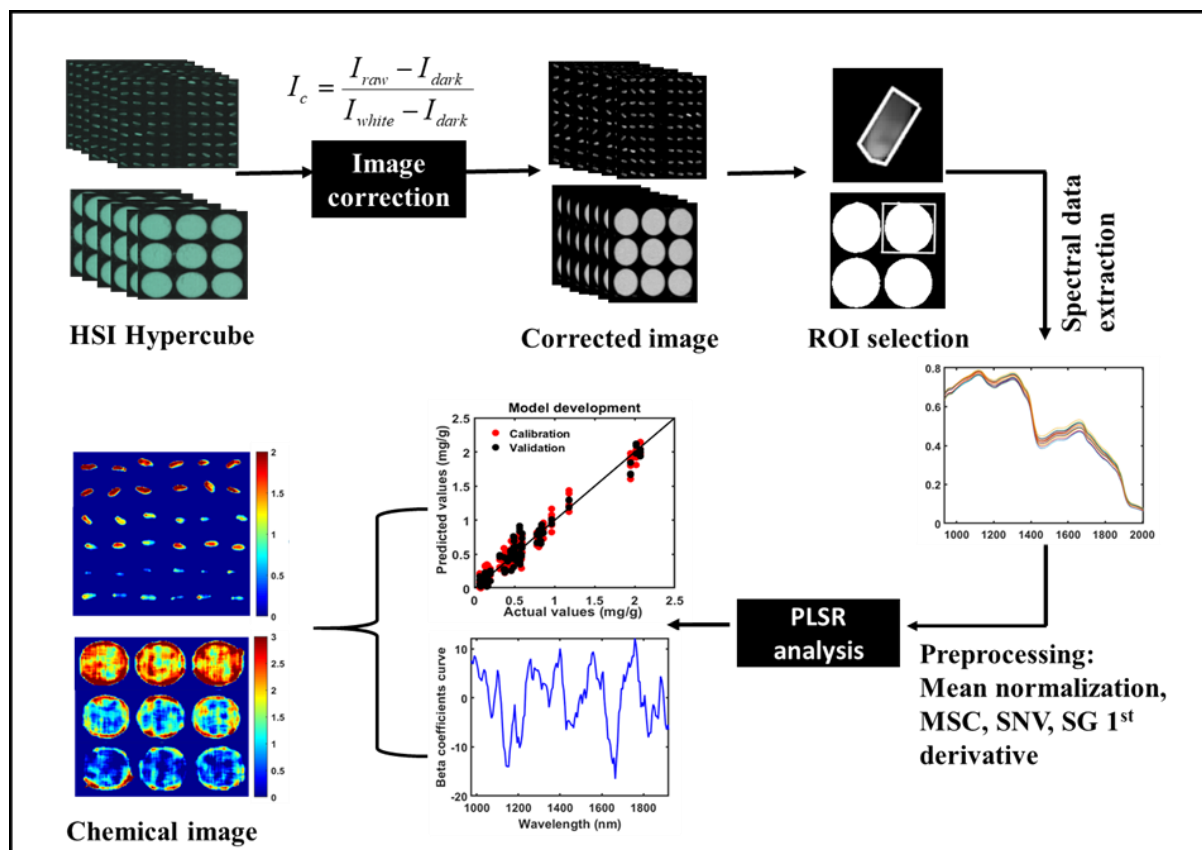


Figure 2. The strategy of hyperspectral data extraction and analysis.

2.4. Data Pre-Processing

The spectral data mostly contained a considerable level of noise, possibly generated from the camera and environmental effect. The acquired spectral data were subjected to several pre-processing methods, including mean normalization, multiplication scatters correction (MSC), standard normal variate (SNV), and Savitzky–Golay first derivative (Figure 2). Normalization was used to fit spectral data within a comparable range (0–1) to compensate for inconsistencies due to the thickness of the sample and optical source length. Standard normal variate (SNV) pre-processing was used to correct spectral information changes caused by scattering and particle size variability [31]. MSC was used to correct the spectra scattering intensity [31], while Savitzky–Golay’s first derivatives were applied to correct the baseline effect and eliminate overlapping peaks [32].

2.5. High-Performance Liquid Chromatography (HPLC) Test for Reference Analysis

The anthocyanins content in black rice was measured using high-performance liquid chromatography (HPLC) by injected 20 μ L sample extract into the system. The extraction process was carried out by dissolving the powder of black rice (1 g) into 30 mL of 1% HCL/40% CH₃OH at 4 °C for 24 h. Before analysis, all of the samples were filtered through a 0.45 μ m membrane filter. The HPLC system for quantification of anthocyanins was operated using an Agilent 1200 series HPLC (Boeblingen, Germany) system comprising of a quaternary pump, an Agilent 1200 series diode-array detector (DAD) operating at 530 nm, ChemStation software, a 1200-well-plate autosampler, and a Tosoh ODS-120T column (250 mm \times 4.6 mm i.d., Tokyo, Japan) protected with a Nova-Pak C₁₈ guard insert column (Water, Milford, MA, USA). During the measurement, the column temperature was maintained at 30 °C.

The mobile phase included solvent A (5% formic acid/H₂O *v/v*) and solvent B (5% formic acid/acetonitrile *v/v*), entering the system at a flow rate of 0.7 mL/min. The applied gradient program was: 0–20 min, 10–30% B; 20–25 min, 30–60% B; 25–26 min, 60–10% B; and 26–35 min, 10% B. The anthocyanins concentration was obtained by comparing the HPLC peak area to external standard calibration curves. The standard calibration curve ($R^2 \geq 0.999$) was provided by injecting 0.5–1 μ g of anthocyanins into 20 μ L of 1% HCL/40% CH₃OH. The total anthocyanins content and individual type of anthocyanins were analyzed, and the result was presented in mg/g.

2.6. Model Development

The partial least squares regression (PLSR) technique was applied to predict total anthocyanins content in black rice. This multivariate method combines multiple regression and feature-based extraction on the principal component analysis method, which may predict the behavior of dependent variables based on the large datasets of independent variables. The model of PLSR relies on the linear relationship between X and Y variables, which creates a possibility to predict the component in the X variable [33]. The equations to express the PLS model are as follows:

$$X = TP^T + E \quad (2)$$

$$Y = UQ^T + F \quad (3)$$

$$U = TB + G \quad (B = (T^T T)^{-1} T^T U) \quad (4)$$

where X is the independent variable matrix representing the spectral data, while Y represents the dependent variable related to the anthocyanins content of black rice. T and U are score matrices, and P and Q are the X and Y loading matrices, respectively. The E matrices denote the error matrices for the X and Y data. Finally, the inner relation between the spectral data and the anthocyanins content was constructed using least squares (Equation (4)). This technique was previously used in our study to predict anthocyanins in

intact soybean using FT-NIR and FT-IR methods; thus, it is recommended because of the excellent performance.

In this research, 1152 spectral data of seeds (36 seeds \times 32 rice samples) and every seed sample were randomly grouped into 9. The spectral data of each group were then averaging, which resulted in 288 spectral data (9 spectra \times 32 rice sample). The averaging process was carried out to accommodate the probability of rice seed physical characteristics difference within a single rice sample. The weight of 100 seeds of rice ranged from 1.46 to 2.44 g, indicating rice size variability based on the variety. Average seed weight, which is closely related to seed size, has been proven to affect the phytochemical content in soybean [34], and is thought to have a similar effect on rice seed. The 288 spectral data for powder samples were obtained from 32 varieties, with 9 data from each. The entire dataset was divided into a calibration and validation dataset. The calibration dataset consisted of 70% of the data, while the validation dataset consisted of 30% of the data.

The wavelength area used to develop the model ranged from 900 to 1800 nm. This waveband selection was carried out based on the consideration that the illumination of a tungsten-halogen light source equipped in the NIR-HSI only effective in that region even though the sensing module of the instrument has a spectral range coverage of 895–2504 nm. The reflectance intensity of the white image using the Teflon sheet (99% reflection) over 1800 nm was low and flat, meaning that the data obtained from this area is unreliable or noisy.

2.7. Image Processing

The capability of creating chemical images is one of the benefits of using hyperspectral imaging in the prediction of chemical components in food and agricultural product. These advantages provide a possibility to visualize the distribution of the evaluated chemical in the examined sample. The chemical image was developed based on the beta coefficient of the created model. The process of creating the chemical image was done by transferring the 3D hyperspectral image to a two-dimensional matrix and multiplying the matrix using the PLS regression coefficient. The resulted matrix was then converted into a 3D image. The visualization of the different anthocyanins concentrations in the samples was carried out by summing the corresponding pixels of all band images. The following equation explains the process to obtain a chemical image (I_{chem}):

$$I_{chem} = \sum_{i=1}^n I_i B_i + C \quad (5)$$

According to the equation, the hypercube image at the particular band (i) is denoted by I_i . The PLS regression coefficient is represented by B_i , and the constant value is indicated by C .

3. Results and Discussion

3.1. The Spectral Characteristic of Black Rice

The rice spectral data based on the seed and powder sample in the 900–1800 nm wavelength showed a similar pattern (Figure 3), characterized by several broad peaks in the same region. The band ranged from 1000 to 1200 were assigned to CH second overtone, which may come from aromatic or aliphatic compounds. The range of wavelengths between 1200 and 1600 nm is the region for the vibration of OH first overtone, NH first overtone, and CH combinations [35]. The next area (1600–1800 nm) is assigned to the OH combination band of water and CH first overtone [36].

Figure 3 also shows a distinct intensity based on the different concentrations of anthocyanins in black rice. In the first range (1140–1350 nm), the CH second overtone band arises. In the band range between 1400 and 1500, the CH combination from the aromatic compound and OH first overtone stretching exist. Despite containing many CH groups, cyanidin-3-glucoside, which is the most abundant anthocyanins in rice, is also rich in OH groups in the structure [37].

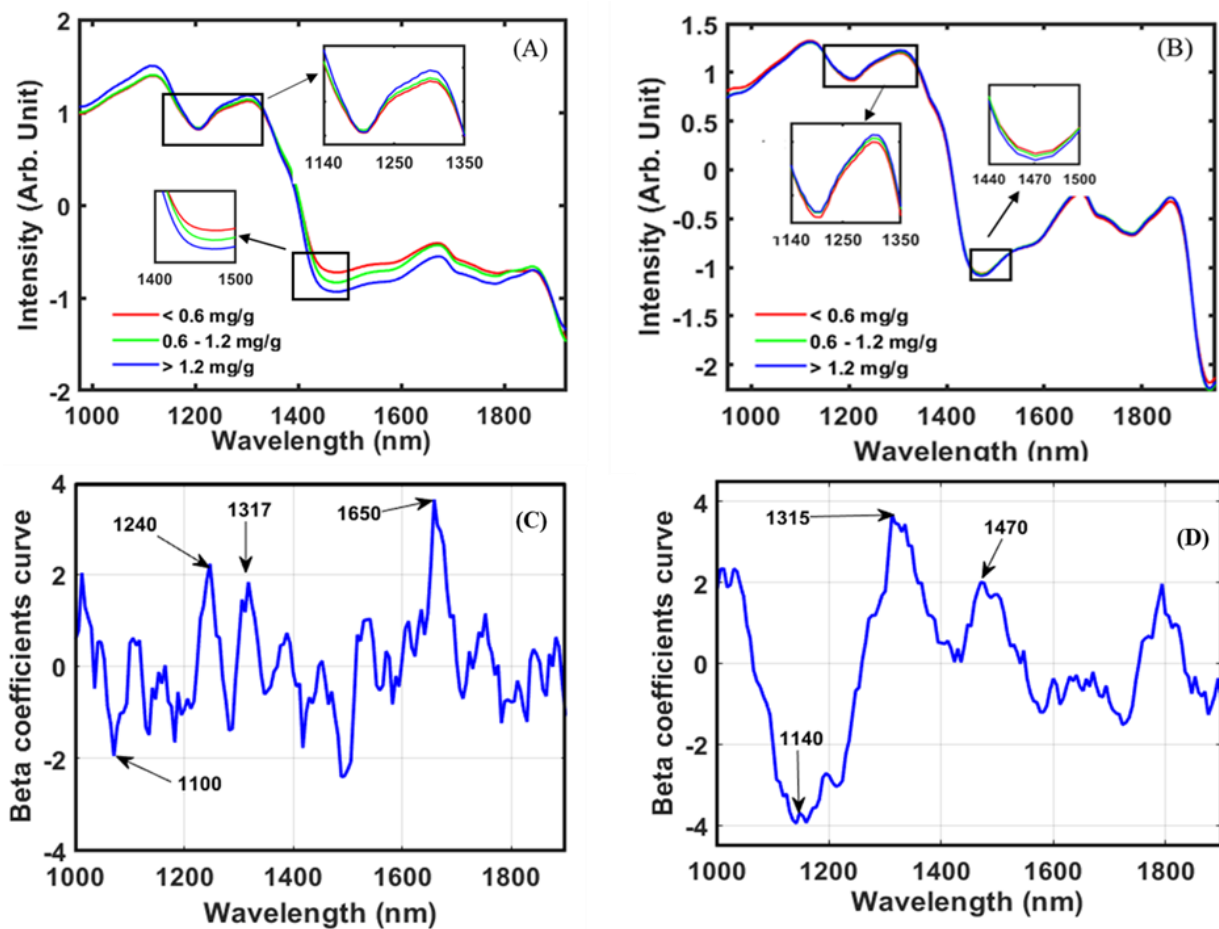


Figure 3. Mean of near-infrared pre-processed spectra using SNV of three rice group based on the anthocyanins content for seed sample (A) and powder sample (B), and beta coefficients curve of PLSR model for seed (C) and powder (D).

3.2. High-Performance Liquid Chromatography (HPLC) Result

The reference values for anthocyanins in this study were evaluated from thirty-two rice samples planted in different years (2019 and 2020), and the result is presented in Table 2.

Table 2. The reference value of anthocyanins (mg/g) obtained through HPLC analysis.

Year	Component	Number of Rice Varieties	Mean \pm SD	Max	Min	Range
2019	Total anthocyanins	16	0.41 \pm 0.32	1.18	0.07	1.11
	Cyanidin-3-glucoside	16	0.38 \pm 0.29	1.07	0.07	1.00
	Peonidin-3-glucoside	9	0.07 \pm 0.02	0.11	0.05	0.06
2020	Total anthocyanins	16	0.91 \pm 0.59	2.07	0.32	1.75
	Cyanidin-3-glucoside	16	0.81 \pm 0.56	1.89	0.26	1.63
	Peonidin-3-glucoside	16	0.10 \pm 0.03	0.18	0.06	0.12
2019–2020	Total anthocyanins	32	0.66 \pm 0.53	2.07	0.07	0.12
	Cyanidin-3-glucoside	32	0.59 \pm 0.49	1.89	0.07	1.82
	Peonidin-3-glucoside	25	0.09 \pm 0.03	0.18	0.05	0.13

SD: Standard deviation.

Sixteen rice varieties harvested in 2019 contain anthocyanins at levels between 0.07 and 1.18 mg/g, while 2020 samples contain anthocyanins at levels ranging from 0.32 to 2.07 mg/g. Cyanidin-3-glucoside (C3G) is the most abundant anthocyanins in the examined samples and can be evaluated in all rice varieties. Among 32 samples of black rice, twenty-five samples contain peonidin-3-glucoside (Pn3G) in a shallow concentration, ranging from 0.05 to 0.18 mg/g. A previous study by Park et al. (2008) [38] reported that

C3G and Pn3G were the major anthocyanins identified in the Korean black rice, while Lee (2010) [11] showed C3G as the highest anthocyanins type in black rice. C3G and Pn3G were also reported as the most common anthocyanin types in Chinese black rice [12]. In this research, the C3G concentration was similar to the total anthocyanins content. Thus, the prediction model for anthocyanins in this study was only based on the total anthocyanins content in rice.

3.3. Partial Least Square Regression (PLSR)

The multilinear regression model in this study was constructed using 288 spectra ranging from 900 to 1800 nm for both seed and powder samples. The regression results of the PLSR model for each pre-processing method developed using all samples comprising all varieties and harvesting time are presented in Table 3. The obtained model of seed samples showed an acceptable performance indicated by a high correlation coefficient (R^2), more than 0.85 and 0.75 for the calibration and validation dataset, respectively. The model performances of the powder sample were better than those of seed samples, denoted by higher R^2 , which were more than 0.9 for both the calibration and validation dataset for all pre-processed and raw data, except for Savitzky–Golay’s (SG) first derivative on the validation dataset. Anthocyanins in rice are not distributed uniformly throughout the seed, which affects sample homogeneity. In addition, the standard error of the models developed based on the seed sample was slightly higher than powder, indicating more scattered data.

Table 3. The partial least square regression (PLSR) model developed for predicting anthocyanins content in rice using short wave infrared hyperspectral technique by applying the different pre-processing methods.

Year	Pre-Processing	Seed Sample				Powder Sample			
		Rc^2	SEC	Rv^2	SEP	Rc^2	SEC	Rv^2	SEP
2019	Raw	0.85	0.12	0.81	0.13	0.95	0.07	0.93	0.09
	Mean Normalization	0.88	0.11	0.77	0.14	0.95	0.07	0.92	0.09
	MSC	0.92	0.09	0.80	0.13	0.96	0.06	0.92	0.10
	SNV	0.87	0.12	0.85	0.11	0.95	0.07	0.92	0.09
	SG first derivative	0.94	0.07	0.79	0.13	0.95	0.06	0.88	0.11
2020	Raw	0.95	0.13	0.88	0.21	0.95	0.13	0.93	0.16
	Mean Normalization	0.93	0.15	0.89	0.19	0.96	0.12	0.93	0.16
	MSC	0.95	0.13	0.92	0.17	0.96	0.12	0.93	0.17
	SNV	0.95	0.12	0.92	0.17	0.96	0.12	0.95	0.14
	SG first derivative	0.91	0.17	0.80	0.26	0.95	0.13	0.81	0.28
Mixed data	Raw	0.87	0.19	0.79	0.24	0.94	0.13	0.91	0.17
	Mean Normalization	0.91	0.15	0.79	0.23	0.94	0.13	0.90	0.17
	MSC	0.89	0.17	0.85	0.20	0.94	0.13	0.92	0.16
	SNV	0.93	0.14	0.90	0.16	0.94	0.13	0.92	0.15
	SG first derivative	0.85	0.20	0.75	0.26	0.93	0.14	0.84	0.22

MSC: Multiplication scatter correction; SNV: Standard normal variate; SG: Savitzky–Golay; SEC: Standard error of calibration; SEP: Standard error of prediction in validation.

Furthermore, the model performance developed using a mixed sample of 2019 and 2020 samples was similar to the model constructed based on the separated harvesting time. This phenomenon indicated the consistency of the model and proved the robustness of the PLSR method to predict anthocyanins in rice using NIR-HSI data. Overall, the SNV pre-processing method exhibited the highest performance (R^2) of 0.93 in the calibration dataset and 0.90 on the validation dataset for predicting rice anthocyanins based on seed spectra of mixed samples. This performance is comparable to the previous study conducted by Zhang et al. (2020) that predicted anthocyanins of dry goji berries and provided R^2 of 0.91–0.95 [30]. The NIR-HSI technique coupled with LS-SVM also was efficiently employed with R^2_p 0.96 and RMSE 0.146 mg/g to quantify total anthocyanins in mulberry fruit (*Morus* sp.) [39]. Another study conducted by Chen et al. (2015), which also proves the feasibility of NIR-HSI to measure anthocyanins in wine grapes, obtained the highest

accuracy ($R^2_p = 0.94$, RMSE = 0.0046 mg/g) by PLS with smoothed spectra [40]. Our result also revealed that model performance based on the powder was slightly better, denoted by a higher correlation coefficient of calibration (Rc^2) and validation (Rv^2) (0.94 and 0.93, respectively), and lower standard error on both calibration and validation datasets. This result was understandable due to powder samples being more homogenous than seed samples. Hence, the result will open up the door for using the NIR-HSI system to separate rice seeds based on the anthocyanins content.

The bands which contributed to the model development can be identified using the beta coefficient curve (Figure 3C,D). The specified bands of the seed and powder sample were similar, around 1140 to 1350 nm, which can be associated with the CH second overtone, and around 1450 nm representing the CH combination from the aromatic structure. Both bands were in the range of influential bands used for predicting anthocyanins in Jaboticaba fruit (*Myrciaria jaboticaba* (Vell.) O. Berg) [41]. The previous study by [40] also reported three important wavebands similar to our finding (1050, 1250, and 1400 nm), which contributed significantly to the development model predicting anthocyanins in wine grapes using NIR hyperspectral imaging. This range of wavebands between 1415 and 1512 nm is also reported as essential wavebands for flavonoid constituents [42].

3.4. Visualization Image Based on Anthocyanins Content in Rice

The effectiveness of the HSI image to visualize the chemical content in powder samples has been proven by several researchers. HSI has been successfully used to detect and quantify apricot in almond powder [43] and detect peanuts and walnuts in wheat powder [44]. In this study, PLS images were generated to visualize the content of anthocyanins in black rice powder and seeds. The chemical image provided the spatial distribution and the anthocyanins concentration of the sample, which are essential to assist in anthocyanins determination. Anthocyanins content in the powder sample was clearly confirmed at a range from 0.08 to 2.07 mg/g by the corresponding color bar shown in Figure 4A, in which the color changes from blue to red following the increase in anthocyanins concentration.

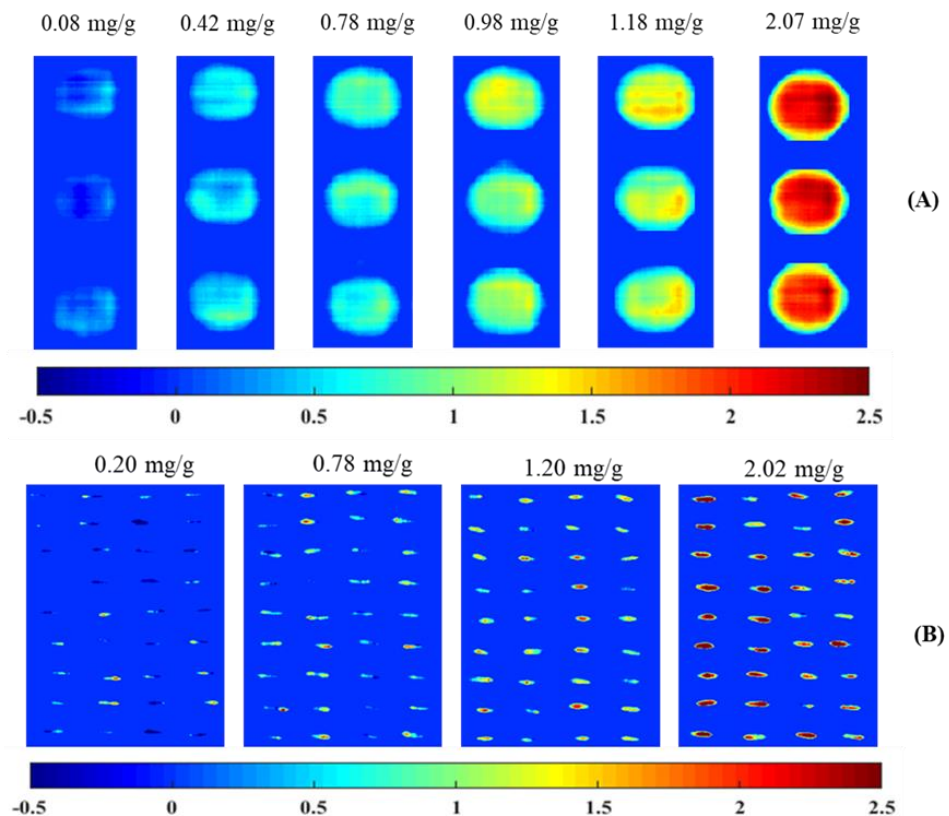


Figure 4. A visualization of anthocyanins based on their concentration in rice powder (A) and seeds (B).

Meanwhile, the color of seeds within a single variety varied, indicating the anthocyanins content variability (Figure 4B). The color gradation based on the concentration could not distinguish the group of seeds based on the anthocyanins content due to the variability of the anthocyanins content within a single variety. In this experiment, the anthocyanins content was examined based on the rice variety. The evaluation of the micro-components of rice could not be done for every single seed due to the minimum weight requirement in the extraction process. Hence, to justify the prediction on a single seed, the average anthocyanins content within each variety was calculated (Figure 5). The average value was compared to the reference value from HPLC and was followed by calculating the root mean square error (RMSE) to evaluate the prediction accuracy.

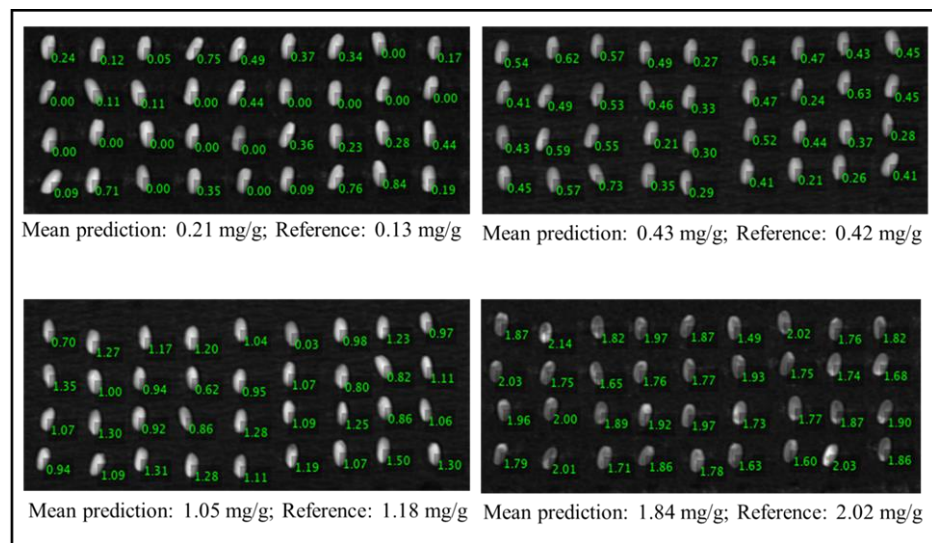


Figure 5. Anthocyanins prediction for every seed of the four examined rice varieties.

The average and the broadest error between mean prediction and reference values were 0.16 and 0.49 mg/g, respectively. The RSME value was 0.21 mg/g, slightly higher than the SEP value of the model of 0.16 mg/g. Among thirty-two varieties, only seven varieties revealed an error higher than the RMSE value, and four of them were varieties with high anthocyanin concentration (more than 1 mg/g). In this experiment, only six varieties contained anthocyanins concentrations of more than 1 mg/g, which possibly affected the model’s development. The prediction plot for all varieties is presented in Figure 6.

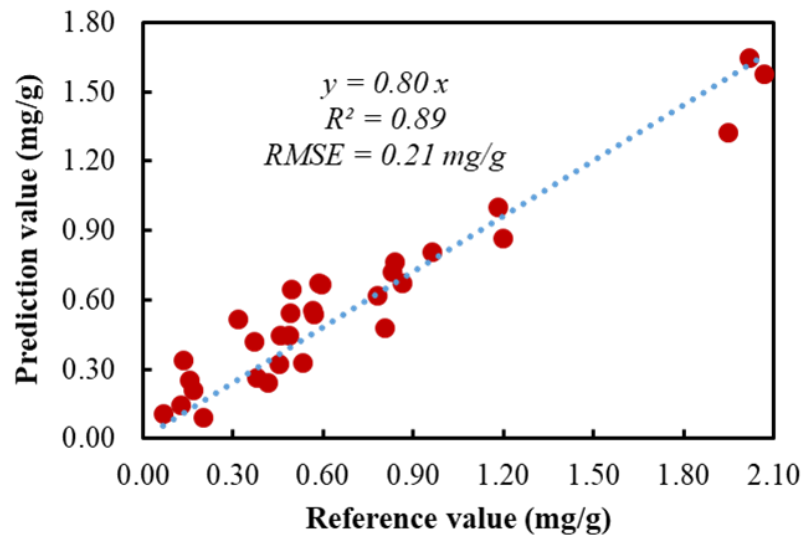


Figure 6. Prediction plot of anthocyanins in rice based on the spectrum of a single seed.

4. Conclusions

A novel technology to predict anthocyanins, a micro-component in rice for single seed, was reported in our research in a rapid and non-destructive technique. The model's performance to predict anthocyanins in seeds was evaluated by comparing the model's performance based on the powder sample. NIR hyperspectral imaging combined with PLSR resulted in a prediction performance (R^2) of 0.75–0.90, with SEP ranging from 0.26 to 0.16 mg/g. The SNV pre-processing method significantly improves the model performances, denoted by the highest R^2 of 0.85–0.95 and lowest SEP of 0.11–0.16 mg/g. This performance was comparable to the model developed based on the powder sample with R^2 and SEP of 0.92–0.95 and 0.09–0.15 mg/g, respectively. A chemical image clearly confirmed the anthocyanins content in the powder sample. The anthocyanins prediction based on seed samples showed the potential of NIR hyperspectral imaging to predict anthocyanins in a single seed of rice.

The HSI approach provides the potential for rapid analysis of a large numbers of seeds. Even though the single seed evaluation is not a substitute for the bulk analysis of anthocyanins, particularly in black rice, it can open up the possibility to develop an online chemical-based sorting machine for seeds. This machine has a potential application in the quality control of seed in the manufacturing industry as well as in the breeding program.

Author Contributions: Data curation, H.Z.A., M.A.F., M.P., and S.S.T.; formal analysis, H.Z.A., C.W., M.A.F., and M.-G.C.; software development, C.W. and M.P.; methodology and validation, R.E.M., W.-H.L., and K.-H.K.; and wrote the paper, H.Z.A. and B.-K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant from the Next-generation BioGreen 21 Program (No. PJ01311303), Rural Development Administration, Republic of Korea.

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

References

1. Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14. [[CrossRef](#)] [[PubMed](#)]
2. Juliano, B.O.; Tũaño, A.P.P. Gross structure and composition of the rice grain. In *Rice*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 31–53, ISBN 9780128115084.
3. Zhang, M.W.; Guo, B.J.; Peng, Z.M. Genetic Effects on Grain Characteristics of indica Black Rice and their Uses on Indirect Selections for some Mineral Element Contents in Grains. *Genet. Resour. Crop Evol.* **2005**, *52*, 1121–1128. [[CrossRef](#)]
4. Min, B.; Gu, L.; McClung, A.M.; Bergman, C.J.; Chen, M.-H. Free and bound total phenolic concentrations, antioxidant capacities, and profiles of proanthocyanidins and anthocyanins in whole grain rice (*Oryza sativa* L.) of different bran colours. *Food Chem.* **2012**, *133*, 715–722. [[CrossRef](#)]
5. Tang, B.; He, Y.; Liu, J.; Zhang, J.; Li, J.; Zhou, J.; Ye, Y.; Wang, J.; Wang, X. Kinetic investigation into pH-dependent color of anthocyanin and its sensing performance. *Dye. Pigment.* **2019**, *170*, 107643. [[CrossRef](#)]
6. Zhao, S.; Park, C.H.; Yang, J.; Yeo, H.J.; Kim, T.J.; Kim, J.K.; Park, S.U. Molecular characterization of anthocyanin and betulinic acid biosynthesis in red and white mulberry fruits using high-throughput sequencing. *Food Chem.* **2019**, *279*, 364–372. [[CrossRef](#)] [[PubMed](#)]
7. Konczak, I.; Zhang, W. Anthocyanins—More Than Nature's Colours. *J. Biomed. Biotechnol.* **2004**, *2004*, 239–240. [[CrossRef](#)] [[PubMed](#)]
8. Kong, J.-M.; Chia, L.-S.; Goh, N.-K.; Chia, T.-F.; Brouillard, R. Analysis and biological activities of anthocyanins. *Phytochemistry* **2003**, *64*, 923–933. [[CrossRef](#)]
9. Yao, S.-L.; Xu, Y.; Zhang, Y.-Y.; Lu, Y.-H. Black rice and anthocyanins induce inhibition of cholesterol absorption in vitro. *Food Funct.* **2013**, *4*, 1602–1608. [[CrossRef](#)]
10. Somsana, P.; Wattana, P.; Suriharn, B.; Sanitchon, J. Stability and genotype by environment interactions for grain anthocyanin content of thai black glutinous upland rice (*Oryza sativa*). *Sabrao J. Breed. Genet.* **2013**, *45*, 523–532.
11. Lee, J.H. Identification and quantification of anthocyanins from the grains of black rice (*Oryza sativa* L.) varieties. *Food Sci. Biotechnol.* **2010**, *19*, 391–397. [[CrossRef](#)]
12. Hou, Z.; Qin, P.; Zhang, Y.; Cui, S.; Ren, G. Identification of anthocyanins isolated from black rice (*Oryza sativa* L.) and their degradation kinetics. *Food Res. Int.* **2013**, *50*, 691–697. [[CrossRef](#)]

13. Bordiga, M.; Gomez-Alonso, S.; Locatelli, M.; Travaglia, F.; Coisson, J.D.; Hermosin-Gutierrez, I.; Arlorio, M. Phenolics characterization and antioxidant activity of six different pigmented *Oryza sativa* L. cultivars grown in Piedmont (Italy). *Food Res. Int.* **2014**, *65*, 282–290. [[CrossRef](#)]
14. Zhang, M.W.; Zhang, R.F.; Zhang, F.X.; Liu, R.H. Phenolic Profiles and Antioxidant Activity of Black Rice Bran of Different Commercially Available Varieties. *J. Agric. Food Chem.* **2010**, *58*, 7580–7587. [[CrossRef](#)] [[PubMed](#)]
15. Yoshimura, Y.; Zaima, N.; Moriyama, T.; Kawamura, Y. Different Localization Patterns of Anthocyanin Species in the Pericarp of Black Rice Revealed by Imaging Mass Spectrometry. *PLoS ONE* **2012**, *7*, e31285. [[CrossRef](#)]
16. Shao, Y.; Xu, F.; Sun, X.; Bao, J.; Beta, T. Phenolic acids, anthocyanins, and antioxidant capacity in rice (*Oryza sativa* L.) grains at four stages of development after flowering. *Food Chem.* **2014**, *143*, 90–96. [[CrossRef](#)] [[PubMed](#)]
17. López-Fernández, O.; Domínguez, R.; Pateiro, M.; Munekata, P.E.S.; Rocchetti, G.; Lorenzo, J.M. Determination of Polyphenols Using Liquid Chromatography–Tandem Mass Spectrometry Technique (LC–MS/MS): A Review. *Antioxidants* **2020**, *9*, 479. [[CrossRef](#)]
18. Kim, Y.; Ahn, H.; Lee, E.; Kim, H. Development of Prediction Model by NIRS for Anthocyanin Contents in Black Colored Soybean. *Korean J. Crop Sci.* **2008**, *53*, 15–20.
19. Xiaowei, H.; Xiaobo, Z.; Jiewen, Z.; Jiyong, S.; Xiaolei, Z.; Holmes, M. Measurement of total anthocyanins content in flowering tea using near infrared spectroscopy combined with ant colony optimization models. *Food Chem.* **2014**, *164*, 536–543. [[CrossRef](#)]
20. Pullanagari, R.R.; Li, M. Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *J. Food Eng.* **2021**, *289*, 110177. [[CrossRef](#)]
21. Amanah, H.Z.; Joshi, R.; Masithoh, R.E.; Choung, M.-G.; Kim, K.-H.; Kim, G.; Cho, B.-K. Non-destructive measurement of anthocyanin in intact soybean seed using Fourier Transform Near-Infrared (FT-NIR) and Fourier Transform Infrared (FT-IR) spectroscopy. *Infrared Phys. Technol.* **2020**, *111*, 103477. [[CrossRef](#)]
22. Wang, Q.; Liu, Y.; Gao, X.; Xie, A.; Yu, H. Potential of hyperspectral imaging for non-destructive determination of chlorogenic acid content in Flos Lonicerae. *J. Food Meas. Charact.* **2019**, *13*, 2603–2612. [[CrossRef](#)]
23. Liang, K.; Liu, Q.X.; Xu, J.H.; Wang, Y.Q.; Okinda, C.S.; Shena, M.X. Determination and Visualization of Different Levels of Deoxynivalenol in Bulk Wheat Kernels by Hyperspectral Imaging. *J. Appl. Spectrosc.* **2018**, *85*, 953–961. [[CrossRef](#)]
24. Weng, S.; Yu, S.; Dong, R.; Pan, F.; Liang, D. Non-destructive detection of storage time of strawberries using visible/near-infrared hyperspectral imaging. *Int. J. Food Prop.* **2020**, *23*, 269–281. [[CrossRef](#)]
25. Liu, C.; Huang, W.; Yang, G.; Wang, Q.; Li, J.; Chen, L. Determination of starch content in single kernel using near-infrared hyperspectral images from two sides of corn seeds. *Infrared Phys. Technol.* **2020**, *110*, 103462. [[CrossRef](#)]
26. Caporaso, N.; Whitworth, M.B.; Fisk, I.D. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chem.* **2018**, *240*, 32–42. [[CrossRef](#)]
27. 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](#)]
28. Tian, X.; Aheto, J.H.; Bai, J.; Dai, C.; Ren, Y.; Chang, X. Quantitative analysis and visualization of moisture and anthocyanins content in purple sweet potato by Vis–NIR hyperspectral imaging. *J. Food Process. Preserv.* **2020**. [[CrossRef](#)]
29. 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](#)]
30. Zhang, C.; Wu, W.; Zhou, L.; Cheng, H.; Ye, X.; He, Y. Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. *Food Chem.* **2020**, *319*, 126536. [[CrossRef](#)]
31. Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P.A.; Massart, D.L. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *J. Pharm. Biomed. Anal.* **1999**, *21*, 115–132. [[CrossRef](#)]
32. Osborne, B.G.; Fearn, T.; Hindle, P.H. Practical NIR spectroscopy with applications in food and beverage analysis. *Longman Sci. Tech.* **1993**, *4*, 12.
33. Kresta, J.V.; Marlin, T.E.; MacGregor, J.F. Development of inferential process models using PLS. *Comput. Chem. Eng.* **1994**, *18*, 597–611. [[CrossRef](#)]
34. Choi, Y.M.; Yoon, H.; Lee, S.; Ko, H.C.; Shin, M.J.; Lee, M.C.; Hur, O.S.; Ro, N.Y.; Desta, K.T. Isoflavones, anthocyanins, phenolic content, and antioxidant activities of black soybeans (*Glycine max* (L.) Merrill) as affected by seed weight. *Sci. Rep.* **2020**, *10*, 1–13. [[CrossRef](#)]
35. Masithoh, R.E.; Lohumi, S.; Yoon, W.-S.; Amanah, H.Z.; Cho, B.-K. Development of multi-product calibration models of various root and tuber powders by fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents. *Heliyon* **2020**, *6*, e05099. [[CrossRef](#)]
36. Aenugu, H.P.R.; Sathis Kumar, D.; Srisudharson; Parthiban, N.; Ghosh, S.S.; Banji, D. Near infra red spectroscopy—An overview. *Int. J. ChemTech Res.* **2011**, *3*, 825–836.
37. Jing, P.; Bomser, J.A.; Schwartz, S.J.; He, J.; Magnuson, B.A.; Giusti, M.M. Structure–function relationships of anthocyanins from various anthocyanin-rich extracts on the inhibition of colon cancer cell growth. *J. Agric. Food Chem.* **2008**, *56*, 9391–9398. [[CrossRef](#)]
38. Park, Y.S.; Kim, S.-J.; Chang, H.-I. Isolation of Anthocyanin from Black Rice (Heugjinjubyeo) and Screening of its Antioxidant Activities. *Kor. J. Microbiol. Biotechnol.* **2008**, *36*, 55–60.

39. Huang, L.; Zhou, Y.; Meng, L.; Wu, D.; He, Y. Comparison of different CCD detectors and chemometrics for predicting total anthocyanin content and antioxidant activity of mulberry fruit using visible and near infrared hyperspectral imaging technique. *Food Chem.* **2017**, *224*, 1–10. [[CrossRef](#)]
40. Chen, S.; Zhang, F.; Ning, J.; Liu, X.; Zhang, Z.; Yang, S. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chem.* **2015**, *172*, 788–793. [[CrossRef](#)]
41. Mariani, N.C.T.; De Almeida Teixeira, G.H.; De Lima, K.M.G.; Morgenstern, T.B.; Nardini, V.; Júnior, L.C.C. NIRS and iSPA-PLS for predicting total anthocyanin content in jaboticaba fruit. *Food Chem.* **2015**, *174*, 643–648. [[CrossRef](#)]
42. Verardo, V.; Cevoli, C.; Pasini, F.; Gómez-Caravaca, A.M.; Marconi, E.; Fabbri, A.; Caboni, M.F. Analysis of oligomer proanthocyanidins in different barley genotypes using high-performance liquid chromatography-fluorescence detection-mass spectrometry and near-infrared methodologies. *J. Agric. Food Chem.* **2015**, *63*, 4130–4137. [[CrossRef](#)] [[PubMed](#)]
43. Faqeerzada, M.A.; Lohumi, S.; Kim, G.; Joshi, R.; Lee, H.; Kim, M.S.; Cho, B.K. Hyperspectral shortwave infrared image analysis for detection of adulterants in almond powder with one-class classification method. *Sensors* **2020**, *20*, 5855. [[CrossRef](#)] [[PubMed](#)]
44. Zhao, X.; Wang, W.; Ni, X.; Chu, X.; Li, Y.F.; Sun, C. Evaluation of near-infrared hyperspectral imaging for detection of peanut and walnut powders in whole wheat flour. *Appl. Sci.* **2018**, *8*, 1076. [[CrossRef](#)]