

Article

Multispectral Wavebands Selection for the Detection of Potential Foreign Materials in Fresh-Cut Vegetables

Salma Sultana Tunny ¹, Hanim Z. Amanah ^{1,2}, Mohammad Akbar Faqeerzada ¹, Collins Wakholi ¹, Moon S. Kim ³, Insuck Baek ³ and Byoung-Kwan Cho ^{1,4,*}

- ¹ Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea; sstunny@cnu.ac.kr (S.S.T.); hanim_za@ugm.ac.id (H.Z.A.); akbar.faqeerzada@gmail.com (M.A.F.); wcoln@yahoo.com (C.W.)
- ² Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Gadjah Mada University, Yogyakarta 55281, Indonesia
- ³ Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, United States Department of Agriculture, Powder Mill Road, BARC-East, Bldg 303, Beltsville, MD 20705, USA; moon.kim@usda.gov (M.S.K.); insuck.baek@usda.gov (I.B.)
- ⁴ 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.: +82-42-821-6715

Abstract: Ensuring the quality of fresh-cut vegetables is the greatest challenge for the food industry and is equally as important to consumers (and their health). Several investigations have proven the necessity of advanced technology for detecting foreign materials (FMs) in fresh-cut vegetables. In this study, the possibility of using near infrared spectral analysis as a potential technique was investigated to identify various types of FMs in seven common fresh-cut vegetables by selecting important wavebands. Various waveband selection methods, such as the weighted regression coefficient (WRC), variable importance in projection (VIP), sequential feature selection (SFS), successive projection algorithm (SPA), and interval PLS (iPLS), were used to investigate the optimal multispectral wavebands to classify the FMs and vegetables. The application of selected wavebands was further tested using NIR imaging, and the results showed good potentiality by identifying 99 out of 107 FMs. The results indicate the high applicability of the multispectral NIR imaging technique to detect FMs in fresh-cut vegetables for industrial application.

Keywords: fresh-cut vegetables; foreign materials; near infrared spectroscopy; waveband selection



Citation: Tunny, S.S.; Amanah, H.Z.; Faqeerzada, M.A.; Wakholi, C.; Kim, M.S.; Baek, I.; Cho, B.-K. Multispectral Wavebands Selection for the Detection of Potential Foreign Materials in Fresh-Cut Vegetables. *Sensors* **2022**, *22*, 1775. <https://doi.org/10.3390/s22051775>

Academic Editor: Simone Borri

Received: 6 January 2022

Accepted: 22 February 2022

Published: 24 February 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

Fresh-cut fruits and vegetables refers to ready-to-eat or ready-to-cook plant products that have been slightly physically modified by slicing, cutting, peeling, or trimming and then packed as 100% edible products. The International Fresh-Cut Produce Association (IFPA) reported that since 1930, fresh-cut products have been available in retail supermarkets [1]. In the beginning, fresh-cut production companies were established for supplying products to hotels, restaurants, and catering services. However, this industry has become popular in the last few years because of the growing need for fresh fruit and vegetables for consumption [2].

Now, fresh-cut fruit and vegetables are among the most popular commodities in industrialized countries' food markets, remaining the most demanded plant-derived products to meet the needs of modern consumers. Moreover, in the last decade, because of technological advancements, this specialized industry has been successively growing and is better organized [3].

The main challenge of this industry is to ensure the freshness and quality of the fresh-cut products. Quality is dependent upon physical, chemical, and biological contamination. Several studies have been conducted regarding chemical and biological hazards

and changes in fresh-cut produce [4–7]. Fecal contamination and different types of foreign materials (FMs) are the major concerns of physical contamination [7]. In physical contamination, a wide range of foreign materials has been the main source of consumer complaints in the fresh-cut industry. A comprehensive study conducted as a part of the Food Standards Agency (UK) showed that around 20% of incidents concerning foreign materials were reported for vegetables and vegetable products, which is the highest among all food products. The study also found that plastics, glass, metal, wood and plant parts, and insects are the most frequent foreign materials found in vegetables [8].

The presence of FMs in food can cause health risks for the consumer and also affect the reputation of the manufacturer. It is mandatory to ensure the removal of all types of foreign bodies from foods. Several studies have been conducted to identify FMs in different types of food, including X-ray, near infrared spectroscopy, hyperspectral imaging, thermal imaging, ultrasonic, and terahertz methods [9]. Metal detectors, magnets, electrical impedance, and surface-penetrating radar are some of the commonly used methods of the different food processing industries. However, these methods are only suitable for specific types of FMs.

Though FMs are mostly found in vegetables and vegetable products, no study has been performed to evaluate the performance of the systems mentioned earlier for fresh-cut vegetables. Being concerned about the seriousness of foreign materials incidents in fresh-cut vegetables, fundamental research is needed to elucidate a suitable method to identify FMs in fresh-cut produce. A previous study found that the X-ray system is not ideal for detecting plastic fragments in food, and it is also costly [10]. Ultrasonic imaging is not appropriate in the air medium [11], and terahertz imaging is greatly attenuated in the water medium [12]. To use a thermal imaging system, a thermal contrast between the product and the background is needed [13], and no temperature interference from other surfaces is acceptable [11].

In contrast, NIR can penetrate the air spaces between food materials [12], and it is also cheaper, easier, and faster than the other techniques [14,15]. The main disadvantage of NIR spectra is that they often contain overlapping peaks caused by the overtone and combination of the fundamental vibration of the molecules, which draws on broad bands [16]. This makes the spectra challenging to interpret, and extracting chemical information is less straightforward. However, this problem can be avoided by using proper chemometric techniques to build a relationship between absorption values at certain wavelengths and reference values of the measured samples [17]. Another limitation of the NIR technique when using chemometrics is that the technique requires reference data during calibration [17]. However, to classify between two categorical groups, it is easy to use “0” and “1” as the reference values for calibration purposes, which can easily overcome the limitations of NIR.

Two different types of chemometric approaches are used for the classification of NIR spectra, namely the unsupervised and supervised methods [18]. PCA is one of the most common unsupervised techniques in which similar materials have a tendency to create a cluster in the multidimensional plot based on their principal component scores [19]. On the other hand, PLS-DA is a popular supervised chemometric method for analyzing high-dimensional data, where the variables are often correlated with one other [20]. In addition, the algorithm is computationally inexpensive, which is preferable for industrial use [21].

Therefore, an NIR imaging system with proper chemometric methods could be a solution for identifying FMs in fresh-cut vegetables. Nevertheless, before designing the spectral imaging system, it is essential to separately observe the spectral characteristics of potential FMs and vegetables and select critical wavebands to differentiate them. Furthermore, new applications appear in the food industry on a daily basis due to the introduction of noninvasive, chemical free, and fast-moving technologies [22]. Previously, the NIR imaging system was used to identify leaves, twigs, and stones in blueberries [23]; coins, glass balls, and rubber in dough, cheese, doughnuts, and meats [12]; leaves and stems in blueberries [24]; bone fragments in chicken breast fillets [25]; and polyethylene terephthalate (PET), polyethylene (PE), metal, insects, and bones in pork steaks [26]. Additionally, near infrared

spectroscopy was used to identify insects in various foods, such as chestnuts [27], rice [28], and olives [29]. However, each study considered very few selected materials as FMs.

This paper assesses the feasibility of detecting a wide range of FMs in fresh-cut vegetables based on NIR spectroscopy with multivariate analysis. The primary objective of this research is to establish the crucial wavebands that can separate the potential FMs from different types of fresh-cut vegetables and explore the limitations of NIR application in detecting FMs.

2. Materials and Methods

2.1. Samples

Seven types of vegetables, namely cabbage, carrot, radish, green onion, onion, potato, and zucchini, were used in this study. All of the vegetables were purchased fresh from the local market of Daejeon in the Republic of Korea, and then cut into small pieces, with the spectra measured on the same day. For each type of vegetable, 40 sample spectra were collected, i.e., a total of 280 sample spectra. Different types of plastics, wood, stone, bugs, paper, tissue, nails, rubber, etc., were considered as FMs. Figure 1 shows the representative FMs, and a list of potential FMs is provided in Table 1. A total of 320 spectra were collected from all types of FMs.



Figure 1. Photograph of representative FMs.

2.2. NIR Spectroscopy

The spectra of all the vegetables and FMs were acquired using an FT-NIR spectrometer (Antaris II FT-NIR analyzer, Thermo Scientific CO., Waltham, MA, USA), which was set with an InGaAs detector. The sampling window consisted of chemically resistant sapphire. Each piece of vegetable or FM was put in the middle of a custom-made sample holder with

a scanning glass (low OH quartz) window designated for solid samples, which helped to ensure the reproducibility of the diffuse reflectance spectra. The holder had a black lid to restrict the outside light effect. The absorbance spectra of each piece were collected using a total of 32 successive scans in the wave range of 1000–2500 nm ($4000\text{--}10,000\text{ cm}^{-1}$) at an interval of 4 cm^{-1} spectral resolution. Large pieces of FMs and vegetables were cut into small pieces to fit the sampling area properly. Then, the spectra of each piece of vegetables and FMs were obtained separately.

Table 1. List of potential foreign materials used in this study.

Name of FMs	Type	Source
Paper	Different color	Printing paper, books, dairy, packaging paper, packaging boxes, sticky notes, etc.
Plastics	ABS LDPE HDPE PET PS PC PP Nylon	Various processed food packages, bottles, industrial waste plastics, laboratory waste plastics, etc.
Rubber	Different color	Nondestructive biosensing lab, Chungnam National University
Tissue	Different source	Nondestructive biosensing lab, Chungnam National University
Threads	Different color	Nondestructive biosensing lab, Chungnam National University
Stone	Different size and place	Daejeon, South Korea
Wood	Processed Raw	Toothpick, earbuds Plant stem
Cigarettes	Different brand	Different person
Insects	Mosquito Bee Fly Ant Spider Grasshopper Butterfly Others	Environment
Human nail	Various size	Different person
Metal	Nuts, bolts, wires, springs, foils, etc.	Nondestructive biosensing lab, Chungnam National University

2.3. NIR Imaging System

A laboratory-based line scan NIR imaging system was used in this study to acquire the imaging data in the wave range of 900–2500 nm. The system consisted of an imaging spectrograph (SWIR, Headwall Photonics, Fitchburg, MA, USA), an MCT (mercury cadmium telluride) detector (Xeva-2.5-320, Xenics, Belgium) with a camera of 320×256 pixel resolution, a 25 mm $f/1.4$ (OB-SWIR25/1.4—P/N C0808.010) lens, and six 100 W tungsten-

halogen light sources. The sample was moved at 4.73 mm/scan on the way to the camera field of view by a translation stage controlled by a DC motor.

2.4. Spectral Data Preprocessing and Multivariate Analysis

Because of the scattering of the light, changes in particle size, and instrumental drift, the FT-NIR spectra may have had noise. Therefore, the raw data were required to undergo preprocessing methods to attain precise chemical information from the spectral peaks and valleys [30]. In this study, all spectral data were treated with seven preprocessing methods: three data normalization methods (maximum, minimum, and range), two derivatives (Savitzky–Golay first and second), multiplicative scatter correction (MSC), and standard normal variate (SNV).

To understand the characteristics of the spectral data and the extent of data diffusion, principal component analysis (PCA) could be easily applied. The PC1 (first principal component) explains the most significant variation in the data. Then, a PC2 (second principal component), which is orthogonal to the first PC, defines the remaining variation, which PC1 does not describe, and so on. The most significant characteristics of the spectral data can be observed on a graphical plot using this tool [31].

Afterward, a multivariate analytical model of PLS-DA was developed to discriminate the vegetables from the different types of FMs. PLS-DA, typically applied for model classification, is the modified form of partial least square regression (PLS-R). It is expressed as:

$$Y = X \times \beta + e \quad (1)$$

where X is an $m \times n$ matrix that contains the predictor variables of each class, β is the vector of the regression coefficient, e is the error term, and Y is the response vector comprised of a synthetic value expressing class, as provided below:

$$Y = \begin{cases} 0 & \text{= sample aligned to vegetable class} \\ 1 & \text{= sample aligned to foreign materials class} \end{cases}$$

In order to precisely distinguish the vegetables and FMs, a baseline was chosen as ± 0.5 with respect to each class. Then, the X and Y values were decomposed into latent variables (LVs) to establish a linear connection between the response and predictor variables.

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

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

where T and U are score matrices, and whereas P and Q are loading matrices. E_x and E_y are the error matrices of X and Y , respectively.

2.5. Overview of Multispectral Waveband Selection Methods

The main objective of an important variable selection is to extract the most valuable wavebands from all of the spectral data, which can discriminate the FMs from vegetables, while removing the redundant bands. This will make the model more suitable for industrial applications by significantly reducing the computational time. In this study, five optimal variable selection methods were used, namely weighted regression coefficient (WRC), variable importance in projection (VIP), sequential feature selection (SFS), successive projection algorithm (SPA), and interval PLS (iPLS).

2.5.1. Waveband Selection Based on WRC

The weighted regression coefficient (weighted β -coefficient) shows the connections between the predictors and the corresponding response variables by measuring their weight. The WRC was obtained from the PLS-DA model by dividing each spectrum by its standard deviation to adjust the data on the same scale. The higher positive or negative values of the β -coefficient represent the essential variables that contribute the most to build

the model. It is a widespread method for selecting wavelengths and used successfully by many researchers [32–34]. However, the absorbance value from the original sample spectra needs to consider the corresponding WRC value to select the variables. Variables with a higher β -coefficient but lower (or no) peak in the sample spectra may not contribute to the model prediction [35].

2.5.2. Waveband Selection Based on VIP

Variable importance in projection, as described in [36], was also applied as an optimal variable selection method to build a classification model to differentiate between vegetables and FMs. The general equation to calculate the VIP score of variable j is given below:

$$\text{VIP}_j = \sqrt{\frac{\sum_{i=1}^I W_{ji}^2 \text{SSY}_i J}{\text{SSY}_{\text{total}} I}} \quad (4)$$

where W_{ji} is the weight value for the element i and the variable j , SSY_i is the sum of the squares of the measured variable for the i th element, J is the total number of variables, $\text{SSY}_{\text{total}}$ is the overall sum of squares measured for the response variable, and I is the total quantity of the elements.

2.5.3. Waveband Selection Based on SFS

Sequential feature selection is a method that selects features one-by-one in an iterative way by checking the best model fit. It mainly has two components: the criterion function and a sequential search algorithm. The algorithm starts with an empty set and then sequentially adds features, one-by-one, that significantly describe the model based on the criterion function [37]. The criterion function can be defined in two ways: for the classification model, the criterion function is a misclassification rate, and for the regression model, it is the mean squared error. The searching algorithm continues the same process until the required number of variables is selected.

2.5.4. Waveband Selection Based on SPA

The successive projection algorithm was used to select the optimum number of variables by minimizing the redundancy information produced by the spectroscopy to improve the collinearity problems among the variables [38]. The algorithm starts with a single wavelength or variable, then adds a new variable with the maximum projection, and repeats the process until the desired number of variables is selected. Thus, the algorithm produces a subset of variables with the least linear relationship between them from all the variables of the training set and applies this set to the cross-validation data set to evaluate the performance. This algorithm is extensively used to select the optimum number of variables in the multivariate quantitative and qualitative analysis [34].

2.5.5. Waveband Selection Based on iPLS-DA

iPLS-DA is a modified version of the iPLS algorithm described in [39], which identifies the critical variables from the entire spectrum region by dividing them into equal subintervals and then finding the minimum misclassification rate. The model can be developed in two selection modes: forward and backward. In the forward selection mode, the intervals are added in an iterative way until the user-defined conditions are fulfilled. On the other hand, the intervals are removed iteratively in the backward selection mode. However, the backward selection mode typically selects more variables than the forward one [40]; therefore, the forward selection model was used in this study.

All preprocessing methods, variable selection algorithms, and PLS-DA model present in this study were executed using MATLAB R2020b (MathWorks, Natick, MA, USA) computer software.

2.6. Model Performance Assessment

Sensitivity, specificity, and accuracy were calculated to assess the performance of the model developed with the selected variables by using the following equations adapted from Faqeerzada et al., 2020 [41].

$$\text{Sensitivity}(\%) = \frac{T_p}{T_p + F_n} \times 100 \quad (5)$$

$$\text{Specificity}(\%) = \frac{T_n}{T_n + F_p} \times 100 \quad (6)$$

$$\text{Accuracy}(\%) = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \quad (7)$$

where T_p = True positive (sum of accurately detected vegetable samples), F_n = False negative (sum of the vegetable samples detected as FMs), T_n = True negative (sum of the accurately categorized FMs), and F_p = False positive (sum of the FMs grouped as vegetables).

3. Results and Discussion

3.1. Spectral Data Interpretation

The averages of the raw spectral data of the fresh-cut vegetables and all FMs were plotted for the full NIR wavelength (1000–2500 nm) region. In order to make these spectral features easy to comprehend, spectra were plotted in three sections: biological FMs spectra are displayed in Figure 2a, non-biological FMs spectra are shown in Figure 2b, and fresh-cut vegetable spectra are shown in Figure 2c. The spectra in the wavelength regions of 1400–1450 nm and 1900–1950 nm show higher intensity for vegetables than the FMs due to the presence of excessive water content, which can differentiate the fresh-cut produce from foreign bodies. The absorption bands of vegetables found at 1190 nm, 1450 nm, and 1940 nm were assigned to the combinations of the O–H first overtone and O–H bending band, the first overtone of O–H stretching band, and the combination of O–H stretching band and O–H bending band, individually [42].

However, it was observed that almost all of the FMs and vegetables had a peak near 1700–1800 nm, with different intensities. Several studies have found that the absorption band detected close to 1700–1800 nm due to the first overtone of C–H stretch proves the presence of cellulose content in vegetables, bugs, papers, and woods [43–45]. In addition to this, plastics (LDPE, HDPE, PP, ABS, and PET) also have peaks over 1720–1770 nm due to the first overtone of the asymmetric stretching of a methyl group [46]. As the main purpose of this study was to classify the FMs in fresh-cut vegetables, the bands with no similarity or conflict with vegetable spectra were not discussed.

3.2. Principal Component Analysis (PCA)

Spectral datasets typically contain many more variables than the number of samples, which can make a model overly confident when multiple linear regression techniques are used [47]. As an unsupervised approach, the principal components in a PCA model can serve as an indicator for model reliability by pointing out the direction of maximum variance [48]. The score plot of the PCA model is a subspace projection of data that is used to analyze the relationships between the observations.

The obtained scatter plot of the PC scores after applying the mean normalization preprocessing method, displayed in Figure 3a,b, shows significant variation between the vegetables and FMs. The first three PCs demonstrated a cumulative total variation of 96.51% in the classification of FMs and vegetables. As demonstrated in the two-dimensional projection (Figure 3b), PC1 and PC2 were the most important components for FM identification, contributing a total variation of 92.9%.

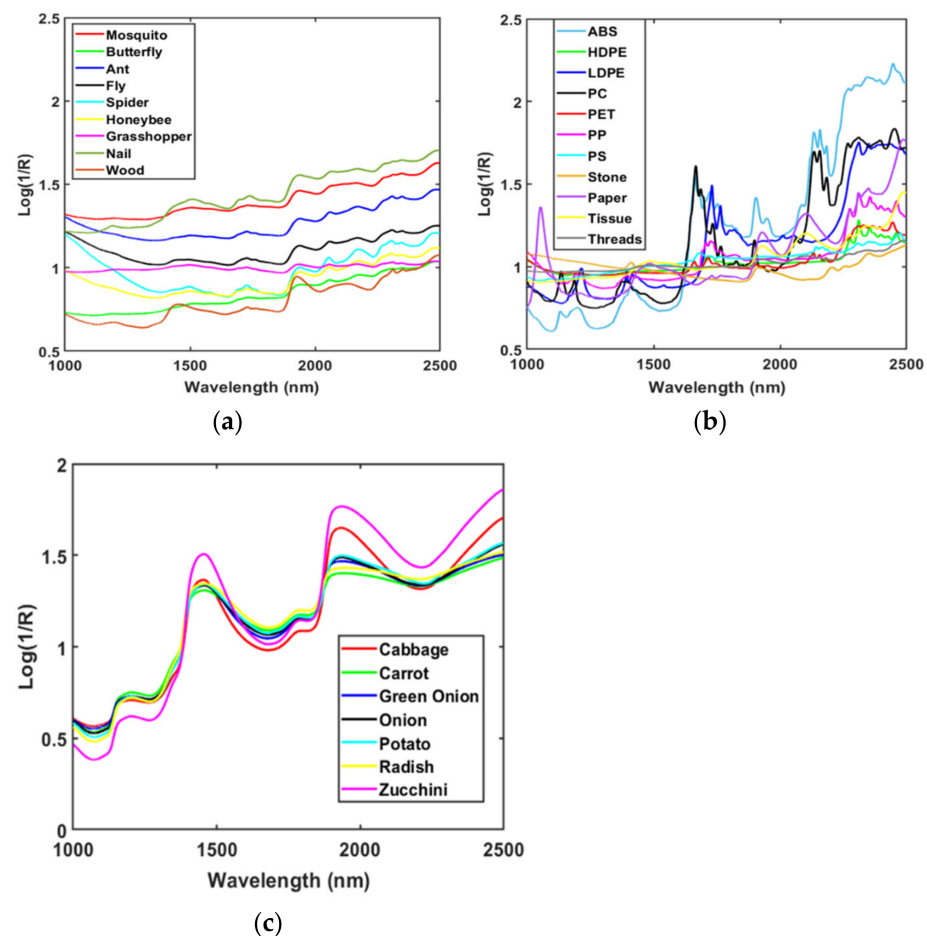


Figure 2. Raw FT-NIR spectral plots of (a) biological FMs, (b) non-biological FMs, and (c) fresh-cut vegetables.

The PC loading plot was developed to define the corresponding weights of the variables for the first three PCs (PC1, PC2, and PC3), as shown in Figure 3c. These loading plots showed some critical information about wavelengths and had similarities with the peaks described in the previous section. Three peaks positioned at 1450 nm, 1730 nm, and 1920 nm may have a maximum contribution to discriminate the FMs from vegetables. The disparity between vegetables and FMs using PCA and the existence of spectral peaks at the specified wavelength established the base for the following study to construct the classification model using PLS-DA.

3.3. PLS-DA Model

Partial least square discriminant analysis (PLS-DA) was accomplished as a supervised learning model where two known classes were defined. The total data set was randomly shuffled and divided into two data sets where 70% of the data was used in the calibration model and the remaining 30% of data was used in model validation. The classification parameters to identify FMs in fresh-cut vegetables were acquired from the PLS-DA model after employing the preprocessing methods. A baseline shift and non-linear effects can be effectively avoided by using appropriate preprocessing to enhance the classification models [49]. Hence, the effect of different preprocessing methods on the accuracy of classification models was investigated and is summarized in Table 2.

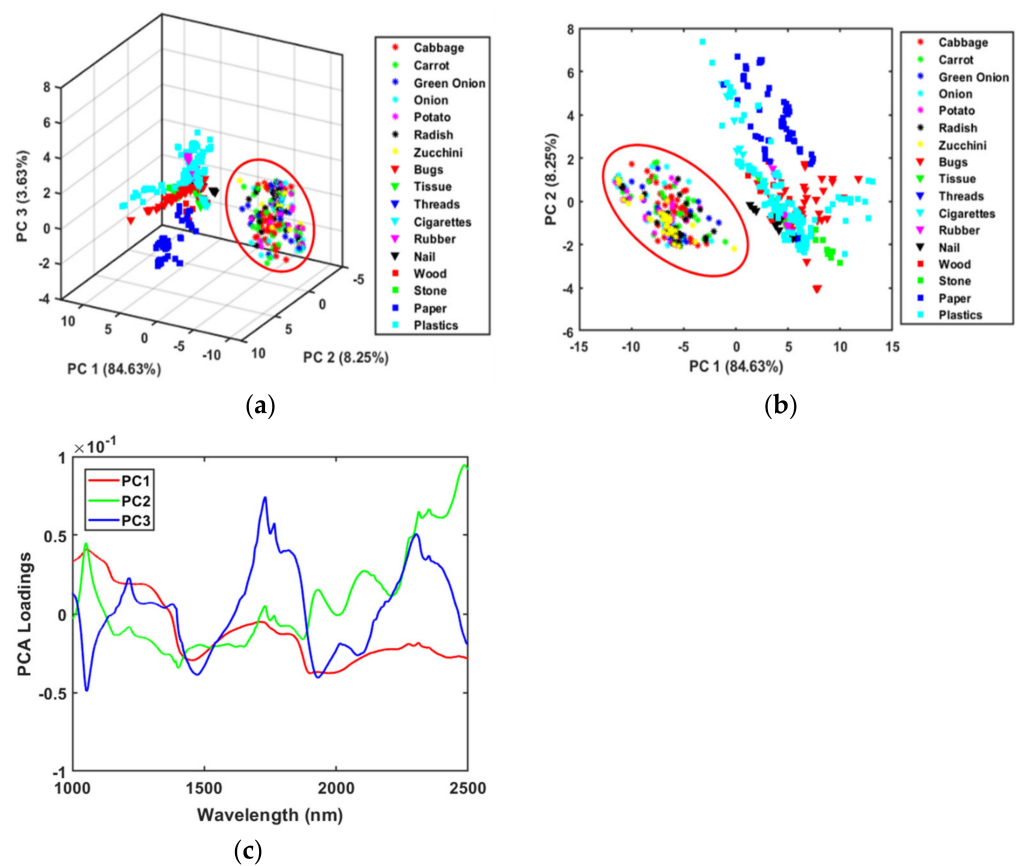


Figure 3. (a) First three PC score plot, (b) first two PC score plot, and (c) PCA loadings plot of mean normalized spectra.

Table 2. Summary of PLS-DA model on the identification of FMs in vegetables using different preprocessing methods.

		Total Number of Samples	Calibration (420 Samples)			Validation (180 Samples)			LVs
			Correctly Classified Fresh-Cuts	Correctly Classified FMs	Accuracy (%)	Correctly Classified Fresh-Cuts	Correctly Classified FMs	Accuracy (%)	
Normalization	Mean ¹	600	196	224	100	84	96	100	4
	Max ²	600	196	224	100	84	96	100	5
	Range ³	600	196	224	100	84	96	100	6
	MSC ⁴	600	196	224	100	84	96	100	9
	SNV ⁵	600	196	224	100	84	96	100	6
Derivatives	SG1 ⁶	600	196	224	100	84	96	100	4
	SG2 ⁷	600	196	224	100	84	95	99.4	6
	Raw	600	196	224	100	84	96	100	5

¹ Mean normalization; ² maximum normalization; ³ range normalization; ⁴ multiplicative scatter correction; ⁵ standard normal variance; ⁶ Savitzky–Golay first derivative; ⁷ Savitzky–Golay second derivative.

The yielded PLS-DA models demonstrated maximum accuracy (100%) in calibration data sets, including all the preprocessing methods; in validation data sets, only the Savitzky–Golay second derivative (99.4%) had slightly lower classification accuracy. The raw data, i.e., data without any preprocessing, clearly classified the vegetables and FMs with a reference baseline of 0.5, both in the calibration and validation data, respectively (Figure 4). Sugiyama [24] was able to detect the leaves and stems from blueberries with 100% accuracy using NIR spectra with discriminant analysis. Diaz [26] clearly distinguished the PET, PE, metal, insects, and bone from pork steaks in the 1100–2500 nm wavelength

region. The high classification accuracy indicated that there are apparent differences in chemical composition between the two groups, and FT-NIR spectroscopy could correctly and reliably identify the FMs in fresh-cut vegetables.

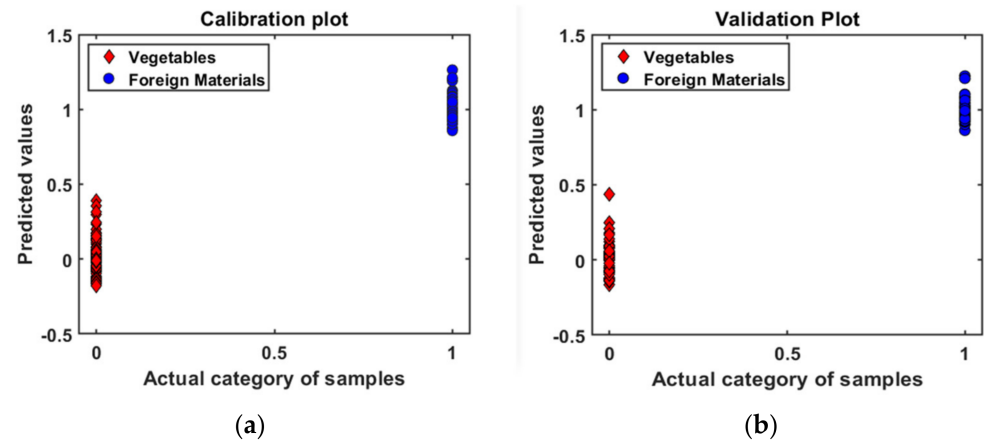


Figure 4. PLS-DA classification plots of raw data: (a) calibration plot and (b) validation plot.

3.4. Selection of Important Wavebands

For the real-time application of a model, it is crucial to reduce the number of variables to run the model faster with better performance and reduce the associated cost. In addition, it is also helpful to develop a multispectral, low-cost imaging system that can ease the data collection and processing in a short time. Therefore, five extensively used variable selection algorithms, namely WRC, VIP, SFS, SPA, and iPLS, were used to select the most important variables.

Firstly, the weighted regression coefficient obtained from the PLS-DA model (Figure 5) was used to explore the connection strength between predictor and criterion variables. The peaks and valleys of the WRC represented the most valuable bands that contributed to classifying the FMs and vegetables. From Figure 5, five peaks and valleys with large weight values corresponding to specific wavelengths were selected, and the bands were 1150 nm, 1400 nm, 1731 nm, 1880 nm, and 1920 nm.

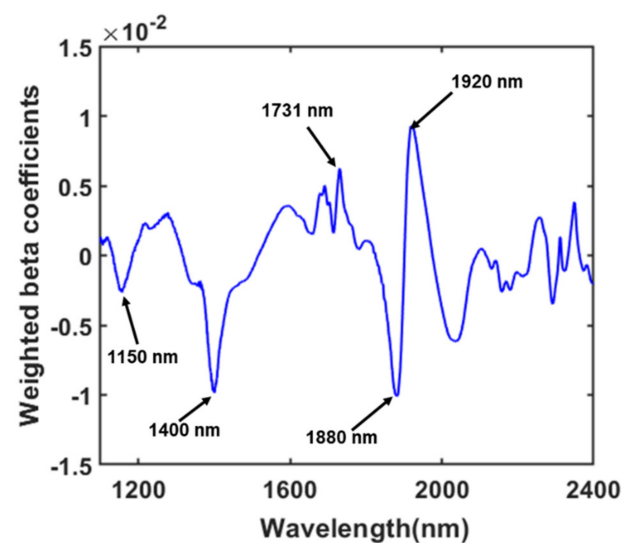


Figure 5. The beta coefficient curve obtained from the PLS-DA model.

The peaks around 1156 nm reflect the second overtone of the CH_3 asymmetrical stretch associated with different plastics [46,50]. The wavelength at 1731 nm corresponding to the first overtone of CH stretching of cellulose is the main component of woods, papers, and

bugs [43,44]. The absorption bands of foods containing 70–90% water are almost identical to those of pure water. Therefore, for the quantitative measurement of water content in foods, substantial FT-NIR absorption band ranges of approximately 1400–1450 nm and 1900–1950 nm were commonly used [51]. In this study, three strong peaks were found in the regression plot; among them, 1400 nm and 1920 nm corresponded to the moisture content of vegetables.

Secondly, the VIP, one of the well-known methods in multivariate analysis, was used to select the optimum wavebands. The number of predictors and the accuracy of the PLS-VIP model is usually determined by the cut-off value. The most excellent combination of model accuracy and number of variables was found at 1.4, which was chosen by exploring a range of values ranging from 0.8 to 1.5. After applying VIP analysis, the number of predictors drastically reduced from 1557 to 10, which was only 0.64% of the original FT-NIR spectral data. The selected wavelengths were in the 1881–1901 nm region, corresponding to the O-H structure.

Thirdly, the sequential feature selection, a wrapper method, was used to evaluate the selected features each time by calculating the error rate. This study calculated the error rate using a randomly selected 20-fold cross-validation method. The first selected wavelength was 1920 nm, and the error rate was 0.02%. After adding the second feature (1135 nm) to the dataset, the error rate became 0.004%, and finally, the misclassification rate became 0% when the third feature (1450 nm) was added to the dataset.

Fourthly, to select the important variables by SPA, the minimum and the maximum number of variables was selected (2 and 10, respectively), as the target was to select as minimal a wavelength as possible. Finally, the algorithm selected only four wavelengths (1300, 1402, 1925, and 2114 nm) without misclassification. The wavelength 1300 nm and 2114 nm are related to the CH bond in the lipid and carbohydrate, respectively [52,53].

Lastly, the forward selection interval PLS (iPLS) algorithm selected three wavelengths (1094, 1343, 1958 nm) using three intervals, with an interval width size of 1. The maximum number of the latent variable was five, and the model used cross-validation to evaluate the selected variables. The absorption band near 1094, 1343, and 1958 nm in the near-infrared region corresponds to the combination of CH, the first overtone, and the combination of the OH bond, respectively [34,52,54].

However, each variable selection algorithm works based on different principles. Hence, the different variable selection approaches have shown considerable variation among the selected variables. Figure 6 and Table 3 show the wavelengths selected by each variable selection method. It can be seen that all the variable selection methods selected at least one variable in the range of 1900–1925 nm, except iPLS. In addition to that, WRC, SFS, and SPA selected a single wavelength in the 1400–1450 nm region. Several studies have proved that these two wavelength regions resemble the combination of the O-H bands associated with the water [35,51] present in the vegetables. However, both vegetables and FMs had some common peaks near the rest of the selected wavebands, but they had a clear difference in intensity. Therefore, it can be concluded that the models developed using these selected wavelengths are more robust and effective for the identification of FMs.

3.5. Model Performance Predictions Using Selected Wavebands

In order to prove the application of the selected wavebands, a new PLS-DA model was developed for each optimal variable selection method using raw spectra (without any preprocessing). Table 4 shows the performance of the new PLS-DA models developed using the selected variables. The results indicate that the new models (except VIP-PLS-DA) were able to differentiate the vegetables from the FMs with 100% accuracy (as with the original PLS-DA model (using all variables)). However, the VIP-PLS-DA model detected 93% of the FMs, while having the maximum number of variables (10) among the five variable selection methods. This could be because all the variables selected by the VIP method were in the range of 1888–1901 nm, which was mainly responsible for the moisture content of vegetables. Hence, it is necessary to select the bands from the spectral range that could

represent almost all of the FMs, as well as the vegetables. However, the models developed based on FT-NIR spectroscopy are only suitable for selecting the wavebands, and are not for real-time application. Hence, NIR imaging was used to verify the selected wavebands for identifying the FMs from fresh-cut vegetables in the following analysis.

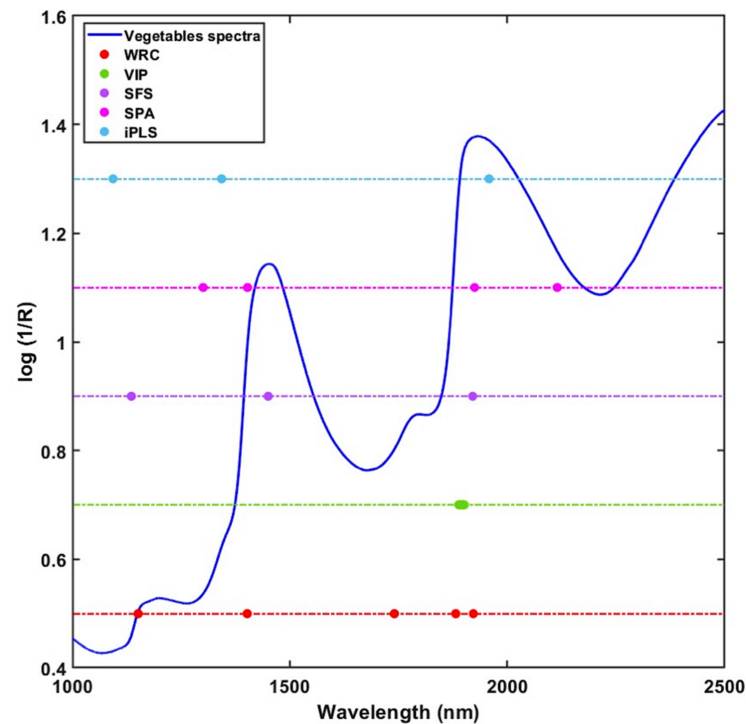


Figure 6. Graphical representation of the selected wavebands using the five variable selection methods.

Table 3. The selected wavelengths for identifying the FMs in fresh-cut vegetables using different variable selection methods.

Variable Selection Method	Selected Variable Number	Selected Wavelengths (nm)
WRC	5	1150, 1401, 1731, 1880, and 1920
VIP	10	1888–1901
SFS	3	1136, 1450, and 1921
SPA	4	1300, 1402, 1925, and 2114
iPLS	3	1094, 1343, and 1958

Table 4. The results of the test data set for identifying the FMs in fresh-cut vegetables using the new PLS-DA models with selected wavebands.

Model	Samples Used	No. of Correctly Detected Vegetables	Sensitivity (%)	No. of Correctly Detected FMs	Specificity (%)	Accuracy (%)
WRC-PLS-DA	300	200	100	100	100	100
VIP-PLS-DA	300	200	100	100	93	97.67
SFS-PLS-DA	300	200	100	100	100	100
SPA-PLS-DA	300	200	100	100	100	100
iPLS-DA	300	200	100	100	100	100

3.6. Testing the Selected Wavebands Using NIR Imaging

Though the selected wavebands performed well in the NIR spectral data, the main challenge was to find out how the bands performed in multispectral imaging techniques. Therefore, the selected bands were tested using the NIR images of the vegetables and

FMs (Figure 7). However, the wavebands selected using individual techniques were not enough to detect the FMs accurately by using the image. Hence, different combinations of the selected bands were explored to find out the best combination. Table 5 represents the number of correctly detected FMs in seven fresh-cut vegetables using three different combinations of bands from the selected wavelengths. The best accuracy was found using the combination of six bands (1150, 1400, 1450, 1731, 1880, and 1920), as shown in Table 5. These six wavebands represented almost all of the variables selected by the different wavelength selection techniques. Overall, 99 FMs were detected accurately among the 107 FMs in all of the vegetables. Table 5 also shows that the classification accuracy using five bands (1150, 1400, 1731, 1880, and 1920) was only 74.8%, whereas the accuracy became 92.5% after using six bands, and it did not improve after adding one more, i.e., seven bands (1150, 1400, 1450, 1731, 1880, 1920, and 2114).

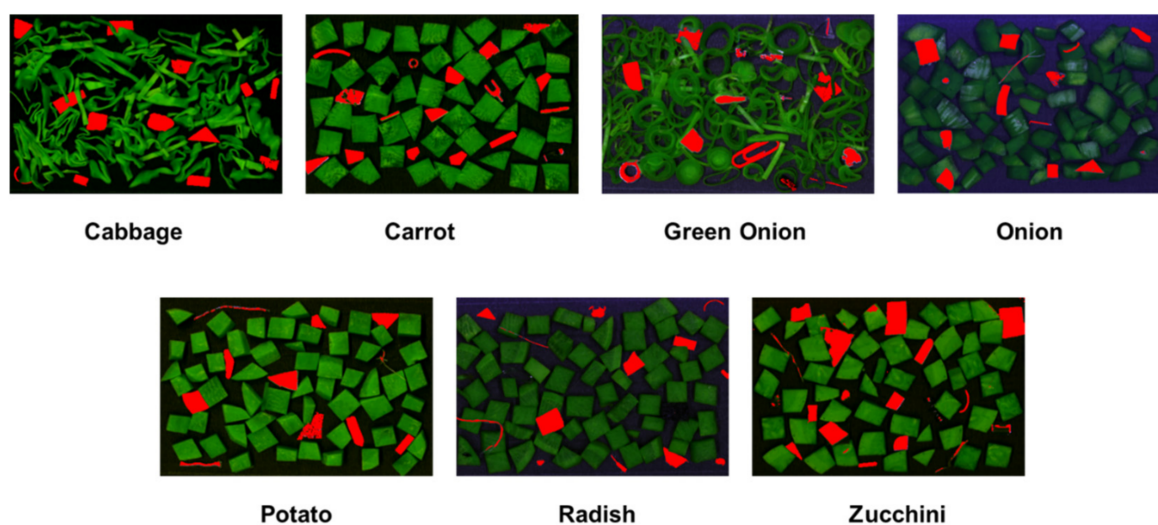


Figure 7. FMs detection images of fresh-cut vegetables using six wavebands.

Table 5. FMs detection accuracy in fresh-cut vegetables using the selected variables in NIR imaging.

Vegetables	Total No. of FMs	5 Bands		6 Bands		7 Bands		False Positive (6 Bands)
		No. of Correctly Identified FMs	Total Accuracy (%)	No. of Correctly Identified FMs	Total Accuracy (%)	No. of Correctly Identified FMs	Total Accuracy (%)	
Cabbage	13	12		13		13		
Carrot	18	15		17		17		
Green onion	13	9	74.77	13	92.5	13	92.5	3
Onion	16	10		14		14		
Potato	12	7		11		11		
Radish	16	12		13		13		
Zucchini	19	15		18		18		

However, the classification accuracy (92.5%) using the NIR imaging of selected bands was lower than the FT-NIR spectroscopy because the spectra were measured separately for the vegetables and FMs using FT-NIR spectroscopy, while for NIR imaging, the data were collected simultaneously by mixing the FMs with the vegetables. In addition, the spectral resolution of FT-NIR spectroscopy (0.6 nm at 1250 nm) was much higher than in the NIR imaging system (5.876 nm), and the instrumental set-up of FT-NIR spectroscopy was very compacted, which reduced the environmental effect and instrumental noise. In contrast, the noise produced from the instrument and the environment might affect the NIR images.

Besides, the FMs, which had similar intensity with the background, were removed during the background removal process.

4. Conclusions

The results obtained from the PLS-DA models confirm that NIR spectroscopy has good potential for distinguishing FMs from fresh-cut vegetables with almost 100% accuracy. In this work, seven types of vegetables and various FMs were used to explore the waveband regions that would successfully separate all types of FMs from different varieties of vegetables. The effective wavebands were selected using WRC, VIP, SFS, SPA, and iPLS algorithms to identify the FMs accurately. The models developed using the selected wavebands, which were around 99% less than the total variables, on average, were almost similar to the original models. Therefore, the selected wavebands were used to test their effectiveness for real-time application using NIR imaging, and finally, six bands were selected after trying different combinations of bands. The test results showed an FMs detection accuracy of 92.5% for all seven fresh-cut vegetables. However, in the future, more research should be conducted to increase accuracy by increasing the spectral resolution of the NIR camera and by selecting a suitable background that can easily differentiate all kinds of FMs from vegetables. Overall, the results demonstrated good potential for designing an NIR multispectral imaging system to distinguish FMs from different types of fresh-cut vegetables.

Author Contributions: Data curation, S.S.T.; formal analysis, S.S.T., H.Z.A., M.A.F. and C.W.; investigation, S.S.T.; methodology, S.S.T. and H.Z.A.; software, S.S.T.; validation, B.-K.C.; writing—original draft, S.S.T.; conceptualization, M.S.K., I.B. and B.-K.C.; writing—review and editing, B.-K.C.; supervision, B.-K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Chungnam National University (2021–2022).

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

References

1. Rojas-Graü, M.A.; Garner, E.; Martín-Belloso, O. The fresh-cut fruit and vegetables industry. Current situation and market trends. In *Advances in Fresh-Cut Fruits and Vegetables Processing*; CRC Press: Boca Raton, FL, USA, 2011; pp. 1–11.
2. Mayen, C.; Marshall, M.I. *Opportunities in the Fresh-Cut Fruit Sector for Indiana Melon Growers*; Department of Agriculture Economics, Purdue University: West Lafayette, IN, USA, 2003; pp. 1–7.
3. Pareek, S. *Fresh-Cut Fruits and Vegetables: Technology, Physiology, and Safety*; CRC Press, Taylor and Francis: Boca Raton, FL, USA, 2016; ISBN 1315353091.
4. Francis, G.A.; Gallone, A.; Nychas, G.J.; Sofos, J.N.; Colelli, G.; Amodio, M.L.; Spano, G. Factors Affecting Quality and Safety of Fresh-Cut Produce. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 595–610. [[CrossRef](#)]
5. Castro-Ibáñez, I.; Gil, M.I.; Allende, A. Ready-to-eat vegetables: Current problems and potential solutions to reduce microbial risk in the production chain. *LWT Food Sci. Technol.* **2017**, *85*, 284–292. [[CrossRef](#)]
6. Mo, C.; Kim, G.; Kim, M.S.; Lim, J.; Cho, H.; Barnaby, J.Y.; Cho, B.K. Fluorescence hyperspectral imaging technique for foreign substance detection on fresh-cut lettuce. *J. Sci. Food Agric.* **2017**, *97*, 3985–3993. [[CrossRef](#)] [[PubMed](#)]
7. Feng, Y.Z.; Sun, D.W. Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 1039–1058. [[CrossRef](#)] [[PubMed](#)]
8. Edwards, M.C.; Stringer, M.F. Observations on patterns in foreign material investigations. *Food Control* **2007**, *18*, 773–782. [[CrossRef](#)]
9. Mohd Khairi, M.T.; Ibrahim, S.; Md Yunus, M.A.; Faramarzi, M. Noninvasive techniques for detection of foreign bodies in food: A review. *J. Food Process. Eng.* **2018**, *41*, e12808. [[CrossRef](#)]
10. Haff, R.P.; Toyofuku, N. X-ray detection of defects and contaminants in the food industry. *Sens. Instrum. Food Qual. Saf.* **2008**, *2*, 262–273. [[CrossRef](#)]
11. Jha, S.N.; Narsaiah, K.; Basediya, A.L.; Sharma, R.; Jaiswal, P.; Kumar, R.; Bhardwaj, R. Measurement techniques and application of electrical properties for nondestructive quality evaluation of foods—A review. *J. Food Sci. Technol.* **2011**, *48*, 387–411. [[CrossRef](#)]
12. Pallav, P.; Diamond, G.G.; Hutchins, D.A.; Green, R.J.; Gan, T.H. A Near-Infrared (NIR) Technique for Imaging Food Materials. *J. Food Sci.* **2009**, *74*, E23–E33. [[CrossRef](#)]
13. Chen, Q.; Zhang, C.; Zhao, J.; Ouyang, Q. Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety. *TrAC Trends Anal. Chem.* **2013**, *52*, 261–274. [[CrossRef](#)]
14. Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **2007**, *44*, 683–700. [[CrossRef](#)] [[PubMed](#)]

15. Tsuchikawa, S. A Review of Recent Near Infrared Research for Wood and Paper. *Appl. Spectrosc. Rev.* **2007**, *42*, 43–71. [[CrossRef](#)]
16. de la Guardia, M.; Armenta, S. Avoiding Sample Treatments. *Compr. Anal. Chem.* **2011**, *57*, 59–86. [[CrossRef](#)]
17. Manley, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* **2014**, *43*, 8200–8214. [[CrossRef](#)] [[PubMed](#)]
18. Blanco, M.; Villarroya, I. NIR spectroscopy: A rapid-response analytical tool. *TrAC Trends Anal. Chem.* **2002**, *21*, 240–250. [[CrossRef](#)]
19. Downey, G. Authentication of Food and Food Ingredients by near Infrared Spectroscopy. *J. Near Infrared Spectrosc.* **1996**, *4*, 47–61. [[CrossRef](#)]
20. Lee, L.C.; Liang, C.Y.; Jemain, A.A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. *Analyst* **2018**, *143*, 3526–3539. [[CrossRef](#)]
21. Mendez, K.M.; Reinke, S.N.; Broadhurst, D.I. A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics* **2019**, *15*, 1–15. [[CrossRef](#)]
22. Yasmin, J.; Ahmed, M.R.; Lohumi, S.; Wakholi, C.; Kim, M.S.; Cho, B.K. Classification method for viability screening of naturally aged watermelon seeds using FT-NIR spectroscopy. *Sensors* **2019**, *19*, 1190. [[CrossRef](#)]
23. Tsuta, M.; Takao, T.; Sugiyama, J.; Wada, Y.; Sagara, Y. Foreign Substance Detection in Blueberry Fruits by Spectral Imaging. *Food Sci. Technol. Res.* **2006**, *12*, 96–100. [[CrossRef](#)]
24. Sugiyama, T.; Sugiyama, J.; Tsuta, M.; Fujita, K.; Shibata, M.; Kokawa, M.; Araki, T.; Nabetani, H.; Sagara, Y. NIR spectral imaging with discriminant analysis for detecting foreign materials among blueberries. *J. Food Eng.* **2010**, *101*, 244–252. [[CrossRef](#)]
25. Yoon, S.C.; Lawrence, K.C.; Smith, D.P.; Park, B.; Windham, W.R. Bone Fragment Detection in Chicken Breast Fillets Using Transmittance Image Enhancement. *Trans. ASABE* **2007**, *51*, 331–339. [[CrossRef](#)]
26. Díaz, R.; Cervera, L.; Fenollosa, S.; Ávila, C.; Belenguer, J. Hyperspectral system for the detection of foreign bodies in meat products. *Procedia Eng.* **2011**, *25*, 313–316. [[CrossRef](#)]
27. Moschetti, R.; Haff, R.P.; Saranwong, S.; Monarca, D.; Cecchini, M.; Massantini, R. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. *Postharvest Biol. Technol.* **2014**, *87*, 88–94. [[CrossRef](#)]
28. Biancolillo, A.; Firmani, P.; Bucci, R.; Magri, A.; Marini, F. Determination of insect infestation on stored rice by near infrared (NIR) spectroscopy. *Microchem. J.* **2019**, *145*, 252–258. [[CrossRef](#)]
29. Moschetti, R.; Haff, R.P.; Stella, E.; Contini, M.; Monarca, D.; Cecchini, M.; Massantini, R. Feasibility of NIR spectroscopy to detect olive fruit infested by *Bactrocera oleae*. *Postharvest Biol. Technol.* **2015**, *99*, 58–62. [[CrossRef](#)]
30. Lohumi, S.; Lee, S.; Lee, W.H.; Kim, M.S.; Mo, C.; Bae, H.; Cho, B.K. Detection of starch adulteration in onion powder by FT-NIR and FT-IR spectroscopy. *J. Agric. Food Chem.* **2014**, *62*, 9246–9251. [[CrossRef](#)]
31. Benar, P.; Gonçalves, A.R.; Mandelli, D.; Ferreira, M.M.C.; Schuchardt, U. Principal Component Analysis of the Hydroxymethylation of Sugarcane Lignin: A Time-Depending Study by FTIR. *J. Wood Chem. Technol.* **1999**, *19*, 151–165. [[CrossRef](#)]
32. Kamruzzaman, M.; Elmasry, G.; Sun, D.W.; Allen, P. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innov. Food Sci. Emerg. Technol.* **2012**, *16*, 218–226. [[CrossRef](#)]
33. Elmasry, G.; Sun, D.W.; Allen, P. Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *J. Food Eng.* **2012**, *110*, 127–140. [[CrossRef](#)]
34. Rahman, A.; Faqeerzada, M.A.; Cho, B.K. Hyperspectral imaging for predicting the allicin and soluble solid content of garlic with variable selection algorithms and chemometric models. *J. Sci. Food Agric.* **2018**, *98*, 4715–4725. [[CrossRef](#)]
35. Kamruzzaman, M.; Kalita, D.; Ahmed, M.T.; ElMasry, G.; Makino, Y. Effect of variable selection algorithms on model performance for predicting moisture content in biological materials using spectral data. *Anal. Chim. Acta* **2021**, 339390. [[CrossRef](#)]
36. Lohumi, S.; Lee, S.; Cho, B.K. Optimal variable selection for Fourier transform infrared spectroscopic analysis of starch-adulterated garlic powder. *Sensors Actuators B Chem.* **2015**, *216*, 622–628. [[CrossRef](#)]
37. Al Ghayab, H.R.; Li, Y.; Abdulla, S.; Diykh, M.; Wan, X. Classification of epileptic EEG signals based on simple random sampling and sequential feature selection. *Brain Inform.* **2016**, *3*, 85–91. [[CrossRef](#)] [[PubMed](#)]
38. Hu, F.; Zhou, M.; Yan, P.; Li, D.; Lai, W.; Zhu, S.; Wang, Y. Selection of characteristic wavelengths using SPA for laser induced fluorescence spectroscopy of mine water inrush. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *219*, 367–374. [[CrossRef](#)] [[PubMed](#)]
39. Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 413–419. [[CrossRef](#)]
40. Ulrici, A.; Serranti, S.; Ferrari, C.; Cesare, D.; Foca, G.; Bonifazi, G. Efficient chemometric strategies for PET-PLA discrimination in recycling plants using hyperspectral imaging. *Chemom. Intell. Lab. Syst.* **2013**, *122*, 31–39. [[CrossRef](#)]
41. 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](#)]
42. Luck, W. Structure of water and aqueous solutions. In *Proceedings of the International Symposium Marburg*; Luck, W.A.P., Ed.; VCH: Berlin, Germany, 1974; pp. 248–284.
43. Aenugu, H.P.; Kumar, D.S.; Srisudharson, N.P.; Ghosh, S.S.; Banji, D. Near infra red spectroscopy—An overview. *Int. J. ChemTech Res.* **2011**, *3*, 825–836.

44. Schwanninger, M.; Rodrigues, J.C.; Fackler, K. A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* **2011**, *19*, 287–308. [[CrossRef](#)]
45. Mandrile, L.; Fusaro, I.; Amato, G.; Marchis, D.; Martra, G.; Rossi, A.M. Detection of insect's meal in compound feed by Near Infrared spectral imaging. *Food Chem.* **2018**, *267*, 240–245. [[CrossRef](#)] [[PubMed](#)]
46. Alassali, A.; Fiore, S.; Kuchta, K. Assessment of plastic waste materials degradation through near infrared spectroscopy. *Waste Manag.* **2018**, *82*, 71–81. [[CrossRef](#)] [[PubMed](#)]
47. Breton, R.G.; Lloyd, G.R. Partial least squares discriminant analysis: Taking the magic away. *J. Chemom.* **2014**, *28*, 213–225. [[CrossRef](#)]
48. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [[CrossRef](#)]
49. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
50. Zheng, Y.; Bai, J.; Xu, J.; Li, X.; Zhang, Y. A discrimination model in waste plastics sorting using NIR hyperspectral imaging system. *Waste Manag.* **2018**, *72*, 87–98. [[CrossRef](#)]
51. Büning-Pfaue, H. Analysis of water in food by near infrared spectroscopy. *Food Chem.* **2003**, *82*, 107–115. [[CrossRef](#)]
52. Johnson, J.B. An overview of near-infrared spectroscopy (NIRS) for the detection of insect pests in stored grains. *J. Stored Prod. Res.* **2020**, *86*, 101558. [[CrossRef](#)]
53. Gergely, S.; Salgó, A. Changes in carbohydrate content during wheat maturation—What is measured by near infrared spectroscopy? *J. Near Infrared Spectrosc.* **2005**, *13*, 9–17. [[CrossRef](#)]
54. Cozzolino, D.; Murray, I. Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT Food Sci. Technol.* **2004**, *37*, 447–452. [[CrossRef](#)]