

Article

Study on Color Detection of Korla Fragrant Pears by Near-Infrared Spectroscopy Combined with PLSR

Yifan Xia ^{1,2,3}, Yang Liu ^{1,2,3,*}, Hong Zhang ^{2,3,4}, Jikai Che ^{1,2,3} and Qing Liang ^{1,2,3}

¹ College of Mechanical and Electronic Engineering, Tarim University, Alar 843300, China; 1075723227@stumail.taru.edu.cn (Y.X.); 10757232218@stumail.taru.edu.cn (J.C.); 10757232263@stumail.taru.edu.cn (Q.L.)

² Xinjiang Production and Construction Corps Key Laboratory of Utilization and Equipment of Special Agricultural and Forestry Products in Southern Xinjiang, Alar 843300, China; 120230116@taru.edu.cn

³ Modern Agricultural Engineering Key Laboratory at Universities of Education Department of Xinjiang Uygur Autonomous Region, Alar 843300, China

⁴ College of Water Resources and Architectural Engineering, Tarim University, Alar 843300, China

* Correspondence: hxtxylove@126.com

Abstract: The difficulty in controlling the quality of Korla pears is the main factor limiting their market value. The key to solving this problem is to detect the color of Korla pears quickly and accurately. This study employed near-infrared spectroscopy (NIRS) technology to measure the absorbance of Korla fragrant pears. The full-spectrum data were pre-processed using six methods: Savitzky–Golay convolution smoothing (SGCS), Savitzky–Golay convolution derivative (SGCD), multiplicative scatter correction (MSC), vector normalization (VN), min–max normalization (MMN), and standard normal variate transformation (SNV). The pre-processed spectral data were subjected to characteristic band extraction using the successive projections algorithm (SPA) and uninformative variable elimination (UVE) methods. Subsequently, detection models for the color indices L^* , a^* , and b^* of Korla fragrant pears were established using the partial least squares regression (PLSR) with full-spectrum and characteristic extracted spectral data. The optimal detection models were determined. The results indicated that pre-processing and characteristic extraction improved the accuracy of the PLSR model. The optimal detection model for the color index L^* was SGCD-UVE-PLSR (correlation coefficient (R) = 0.80, Root Mean Square Error (RMSE) = 1.19); for the color index a^* , it was VN-SPA-PLSR (R = 0.84 and RMSE = 1.28), and for the color index b^* , it was MSC-UVE-PLSR (R = 0.84 and RMSE = 1.25). This research provides a theoretical reference for developing color detection instruments for Korla fragrant pears.

Keywords: color; Korla fragrant pear; near-infrared spectroscopy; partial least squares regression; spectral pre-processing



Academic Editor: Isabel Lara

Received: 9 February 2025

Revised: 23 March 2025

Accepted: 24 March 2025

Published: 25 March 2025

Citation: Xia, Y.; Liu, Y.; Zhang, H.; Che, J.; Liang, Q. Study on Color Detection of Korla Fragrant Pears by Near-Infrared Spectroscopy Combined with PLSR. *Horticulturae* **2025**, *11*, 352. <https://doi.org/10.3390/horticulturae11040352>

Copyright: © 2025 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

The Korla fragrant pear is a specialty fruit from Xinjiang, China, renowned for its crisp exterior, tender interior, juiciness, minimal residue, and rich nutritional value [1]. In the pear quality evaluation system, relevant practitioners usually give priority to the physical and chemical indicators that directly affect the consumption experience, such as sugar degree, pulp texture, and flavor substance content, while the indicator of peel color is placed in a secondary position; modern consumers have also shifted from choosing fruits based on taste to choosing fruits based on appearance and color [2]. Not taking color into consideration can affect consumers' judgment of quality and purchasing decisions for fragrant pears [3].

Color is one of the crucial indicators for assessing the ripeness of Korla fragrant pears in the ripening process; when facing the sun, their color changes from green to red and yellow with an increase in maturity [4]. The a^* value in the color data represents a change in the red and green colors, and the b^* value represents a change in blue and yellow. The values of a^* and b^* in the sample color data can be used to judge the maturity of the sample; moreover, pears with a good color have higher market value [5]. Therefore, color detection is an essential step in the grading process of fragrant pears, helping producers and retailers ensure product consistency and high quality, thereby enhancing consumer satisfaction and the economic benefits of the fragrant pear industry. Traditionally, colorimeters are commonly used for fruit color detection, ensuring accuracy in color measurement. However, the results from colorimeters are easily influenced by environmental factors, such as variations in natural lighting and the duration of artificial lighting, which can reduce detection accuracy [4]. Additionally, colorimeters have limitations in multi-indicator online detection [6]. Therefore, researching precise and rapid color detection technologies for Korla fragrant pears is of great significance for improving the online detection level of their color.

Standard methods for detecting fruit color indices include machine vision [7], hyperspectral imaging [8], and NIRS [9]. Among them, the drawback of machine vision systems is that variations in lighting conditions can affect image quality, thereby affecting the accuracy of color detection. Its versatility and adaptability may also be limited when facing diverse detection requirements. Hyperspectral imaging technology generates large data volumes, requiring more computational resources and complex data processing and analysis algorithms. The large size and weight of the equipment further restrict its application scenarios [10]. Near-infrared spectroscopy (NIRS), characterized by its capability for multi-component non-destructive analysis, contains rich vibrational absorption information from chemical bonds such as O-H, C-H, and N-H, making it applicable to quality measurements of pears. Among them, the quality prediction of pears was performed by Nicacia et al., who used NIR spectroscopy to predict soluble solids and pulp hardness of pears [11]. Liu et al. employed near-infrared spectroscopy to achieve an accurate prediction of pear surface color [12]. These studies have shown that NIRS is applicable in pear color detection. Near-infrared spectroscopy has been applied to predict the hardness and soluble solid content of Korla fragrant pears [13], while color, as a key factor in quality grading and consumer decision-making, is equally important. However, no studies have reported the detection of Korla fragrant pear color based on near-infrared spectroscopy. A commonly used modeling method for NIRS in fruit color detection is partial least squares (PLS). For example, Kusumiyati et al. [14] used NIRS to detect the color of two cucumber varieties, and the PLS method was established for the color indices L^* , a^* , and b^* . The results indicated that NIRS provided an accuracy comparable to conventional methods for cucumber color detection. Chen et al. [15] established a PLS method for the surface color indices L^* , a^* , and b^* of grapes based on NIRS and demonstrated the feasibility of using NIRS for grape color detection. Mancini et al. [16] developed a PLS method for strawberry quality based on NIRS and showed good results for soluble solid detection in strawberries. Zeng et al. [17] used NIRS to establish a PLSR model capable of accurately detecting apples' color and other quality indices. These studies indicate that combining NIRS with the PLS method can establish effective fruit quality detection models, making it feasible to use the PLS method to detect the color of Korla fragrant pears. However, raw spectral data often contain irrelevant noise and interference signals. Researchers commonly use pre-processing methods to optimize data quality and reduce noise. Liu et al. [18] used the PLS method to develop a Kiwi quality detection model based on NIRS and discovered that pre-processing methods like Savitzky–Golay convolution smoothing (SGCS), stan-

standard normal variate transformation (SNV), and multiplicative scatter correction (MSC) effectively improved model accuracy. Ping et al. [19] applied six common pre-processing methods to pre-process the spectral data to build a spectral prediction model for grape indicators, and the pre-processed model obtained the best results. Toktam et al. [20] studied the impact of different pre-processing methods on the accuracy of a PLSR model for predicting pistachio kernel indices and found that pre-processing significantly enhanced model accuracy. Nonetheless, Xie et al. [21] discovered that although noise interference was removed through pre-processing, redundant information remained in the spectral data. To reduce data dimensionality and improve the efficiency and accuracy of PLSR models, the SPA and UVE were used to extract characteristic bands from NIRS data, effectively reducing the impact of redundant spectral information on the model, accelerating computation speed, and improving detection accuracy. Chen et al. [22] improved the detection capability and computation speed of PLSR models for fragrant pears by combining pre-processing with characteristic band extraction using the SPA, thereby enhancing model robustness. Therefore, combining pre-processing with characteristic band extraction in NIRS can effectively improve the accuracy of PLSR-based fruit quality detection models. However, there are few reports on using NIRS with pre-processing and characteristic band extraction methods to establish PLSR models for detecting the color indices L^* , a^* , and b^* of Korla fragrant pears.

This study measured the absorbance of fragrant pears within the band range of 900–1800 nm using NIRS. The spectral data were pre-processed using six methods: SGCS, SGCD, MSC, SNV, VN, and MMN. After pre-processing, characteristic bands were selected using the SPA and UVE characteristic extraction methods. Detection models for the color of Korla fragrant pears were then established, enabling rapid and precise detection of their color.

2. Materials and Methods

2.1. Experimental Samples

The experimental samples for this study were collected from a Korla fragrant pear orchard in Alar City, First Division of the Xinjiang Production and Construction Corps, China. The pears were harvested on 1 October 2023 and 8 October 2023, with 110 pears picked each time. Fruits were selected from three canopy positions (upper, middle, and lower) and both sun-exposed and shaded sides of the trees to comprehensively capture variations in maturity and microenvironmental conditions. Each pear weighed approximately 115 ± 10 g, and 220 uniformly shaped and pest-free pears were selected. The skins of the harvested pears were washed, cleaned, and dried before conducting absorbance and color index detection experiments.

2.2. Spectral Data Measurement

Near-infrared spectral measurements of Korla fragrant pears were conducted at the Key Laboratory of Modern Agricultural Engineering, Tarim University, Xinjiang. The spectral detection equipment used was a near-infrared spectrometer (NIRmagic3500, Beijing Weichuang Yingtu Technology Co., Ltd., Beijing, China) with a band range of 900–1800 nm and spectral resolution <16 nm, allowing 900 spectral data points to be collected per measurement. Before measuring the spectral data, the equipment was preheated and calibrated with a reference for 30 min to minimize the impact of the baseline drift. To ensure the accuracy of the absorbance data, the test samples were in close contact with the probe of the spectrometer, maintaining a stable and stationary position. For each sample, measurements were taken at four evenly distributed points along the equator of the pear, spaced 90° apart. The spectral data were measured three times for each point and averaged. The average spectral data from these four points were used as the measurement result for each pear.

2.3. Color Measurement

The color of the samples was measured using a precision colorimeter (SC-10, Shenzhen 3nh Technology Co., Ltd., Shenzhen, China). The color indices were represented by L^* , a^* , and b^* . Figure 1 shows the CIELAB color space map of Korla fragrant pears, where L^* indicates the brightness, with higher values representing a brighter surface; a^* indicates the red–green difference, with positive values for red, negative values for green, and higher absolute values indicating a deeper red or green color; and b^* indicates the yellow–blue difference, with positive values for yellow, negative values for blue, and higher absolute values indicating a deeper yellow or blue color. Each channel value of the Lab model can accurately represent color, and it is a more linear color space than RGB in perception. The nonlinear relationship of Lab means that the same number of changes in color space can produce approximately the same visual changes, and it is the system closest to human vision, such that it can replace human language description and truly reflect the color of the measured object. It is the most commonly used color expression model [23,24]. The measurement positions of the spectrometer and colorimeter were kept consistent by using marked points on the Korla fragrant pear samples, and the average values of L^* , a^* , and b^* from the four points were used as the color data for each pear.

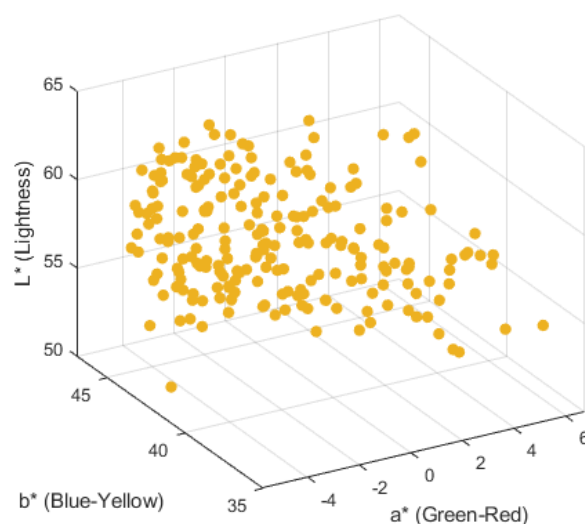


Figure 1. CIELAB color space map of Korla fragrant pears.

2.4. Spectral Pre-Processing Methods

Spectral data often have a low signal-to-noise ratio and include noise and interference unrelated to the detection information, such as background noise, scattering, baseline drift, and electrical noise [25]. Therefore, this study employs six pre-processing methods: Savitzky–Golay convolution smoothing (SGCS), Savitzky–Golay convolution derivative (SGCD), multiplicative scatter correction (MSC), standard normal variate transformation (SNV), vector normalization (VN), and min–max normalization (MMN). Different spectral pre-processing methods have varied effects. The SGCD method can eliminate the influence of baseline drift and background interference, thereby enhancing spectral characteristics. SGCS pre-processing helps eliminate high-frequency noise, smoothen the spectral data, improve the signal-to-noise ratio, and retain important information. SNV is primarily used to eliminate the impact of differences in solid particle size, surface scattering, and optical path variations on reflectance spectra. MSC and SNV have similar pre-processing effects; MSC is mainly used to eliminate scattering caused by uneven particle distribution and particle size differences. VN adjusts the intensity ratio of the spectrum to the same level, facilitating comparisons between datasets. The MMN converts characteristic values proportionally to

the (0, 1) interval by calculating the minimum and maximum values of the characteristics, thus eliminating the influence of differences in dimensions or numerical ranges between different characteristics. Consequently, using spectral pre-treatment methods can improve the accuracy of detection models.

2.5. Characteristic Band Extraction Methods

Near-infrared spectral data contain rich chemical structure information, and there is a high degree of multicollinearity among the spectral data. Using full-spectrum data for modeling often includes a large amount of redundant information, resulting in high computational complexity and slow model operation speed. To simplify the detection model's complexity and enhance its computational speed, this study employs the successive projections algorithm (SPA) and uninformative variable elimination (UVE) methods to extract the characteristic bands of the spectra. The SPA is a forward variable selection method that selects bands by projecting them onto the unselected band space and choosing the band with the maximum projection vector as the characteristic band [26]. This method eliminates multicollinearity among the collinear spectral data, simplifying the model structure and increasing its computational speed. UVE, on the other hand, uses PLSR regression coefficients as a measure of importance for band selection. The basic principle of UVE involves introducing noise, spectral, and reference measurement matrices to form a new matrix and using this matrix to determine a threshold h_{max} to judge whether a variable is a noise [27]. UVE can significantly avoid extracting characteristic bands that are insensitive to the measured components.

2.6. Modeling Methods

This study used PLSR to establish detection models for the color indices of fragrant pears, where the input was absorbance and the output was the color indices L^* , a^* , and b^* . PLSR is a multivariate statistical analysis method that is currently the most widely used multivariate linear regression method in visible and NIRS analysis [28]. It is particularly suitable for constructing detection models when multicollinearity exists among the detection variables. This method combines the characteristics of multiple linear regression, canonical correlation analysis, and principal component analysis. It aims to maximize the covariance between the detection and response variables by projecting them onto a new lower-dimensional variable space. PLSR achieves this by performing regression analysis on the linear combinations of the detection and response variables and identifying the linear combinations of detection variables that best explain the variability in the response variables. These new composite variables, known as latent variables (LVs), are weighted sums of the original detection variables and are orthogonal to each other. Compared to traditional multiple linear regression, PLSR is more robust in handling high-dimensional data and multicollinearity issues. It can process complex datasets while reducing the data dimensions and retaining the most relevant information. For spectral data processing, we used the MATLAB 2019b software (MathWorks, Torrance, CA, USA).

2.7. Model Evaluation

The total number of fragrant pear samples was 220. The sample data were randomly divided into a training set and a test set with a 7:3 ratio. The detection performance of the models was evaluated using the correlation coefficient (R) and Root Mean Square Error

(RMSE). Typically, a high-accuracy model should have high R-values and low values for the RMSE. The formulas for calculating these indicators are as follows:

$$R = \frac{\sqrt{\sum_{j=1}^N (M_j - T_j)^2}}{\sqrt{\sum_{j=1}^N (M_j - \bar{T}_j)^2}} \quad (1)$$

$$\text{RMSE} = \sqrt{\sum_j \frac{(M_j - T_j)^2}{N}} \quad (2)$$

where M_j and T_j are the predicted value and measured value of the data j , respectively, $\bar{T}_j$ is the average value of the measured color value, and N is the total number of data points.

The Residual Prediction Deviation (RPD) and Range Error Ratio (RER) serve as widely adopted model performance evaluation metrics in spectral analysis and chemometrics, providing a holistic assessment of predictive robustness and practical applicability across analytical scenarios. An $\text{RPD} > 2.5$ indicates excellent model effectiveness with high predictive accuracy, $2.0 < \text{RPD} \leq 2.5$ suggests acceptable model performance for preliminary screening, and an $\text{RPD} < 1.4$ denotes inadequate model reliability, rendering it unsuitable for analytical purposes [29]. When $\text{RER} > 4$, the calibration model is suitable for sample screening; an $\text{RER} > 10$ indicates suitability for quality control applications, while an $\text{RER} > 15$ demonstrates applicability for precise quantitative analysis [30]. The formulas for calculating these indicators are as follows:

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \quad (3)$$

$$\text{RER} = \frac{y_{\max} - y_{\min}}{\text{RMSE}} \quad (4)$$

where SD denotes the standard deviation; $y_{\max}$ represents the maximum of the measured values; and $y_{\min}$ denotes the minimum of the measured values.

3. Results and Analysis

3.1. Analysis of Full-Spectrum Data After Pre-Processing

The original near-infrared spectra of 220 Korla fragrant pear samples showed some differences (Figure 2a). However, the overall trends among the spectra were similar, with distinct characteristic peaks at 970 nm, 1190 nm, and 1460 nm. After SGCS processing, the random noise was largely eliminated, resulting in smoother spectra. Although some spectral details were removed, the overall spectral characteristics were not weakened (Figure 2b). The original spectra processed with the SGCD method retained the characteristic peaks of the original spectra and exhibited new characteristic peaks around 1690 nm (Figure 2c). The original spectra processed with MSC (Figure 2d) and SNV (Figure 2e) pre-processing methods eliminated scattering effects caused by uneven sample tissue distribution, making the characteristic peak at 1710 nm more prominent. The spectra showed closer aggregation after VN, enhancing the spectral trend (Figure 2f). The spectra processed with the MMN method showed close aggregation in the concentrated data region. Still, the maximum and minimum values were unstable, leading to poorer aggregation in the spectrum beyond 1550 nm (Figure 2g).

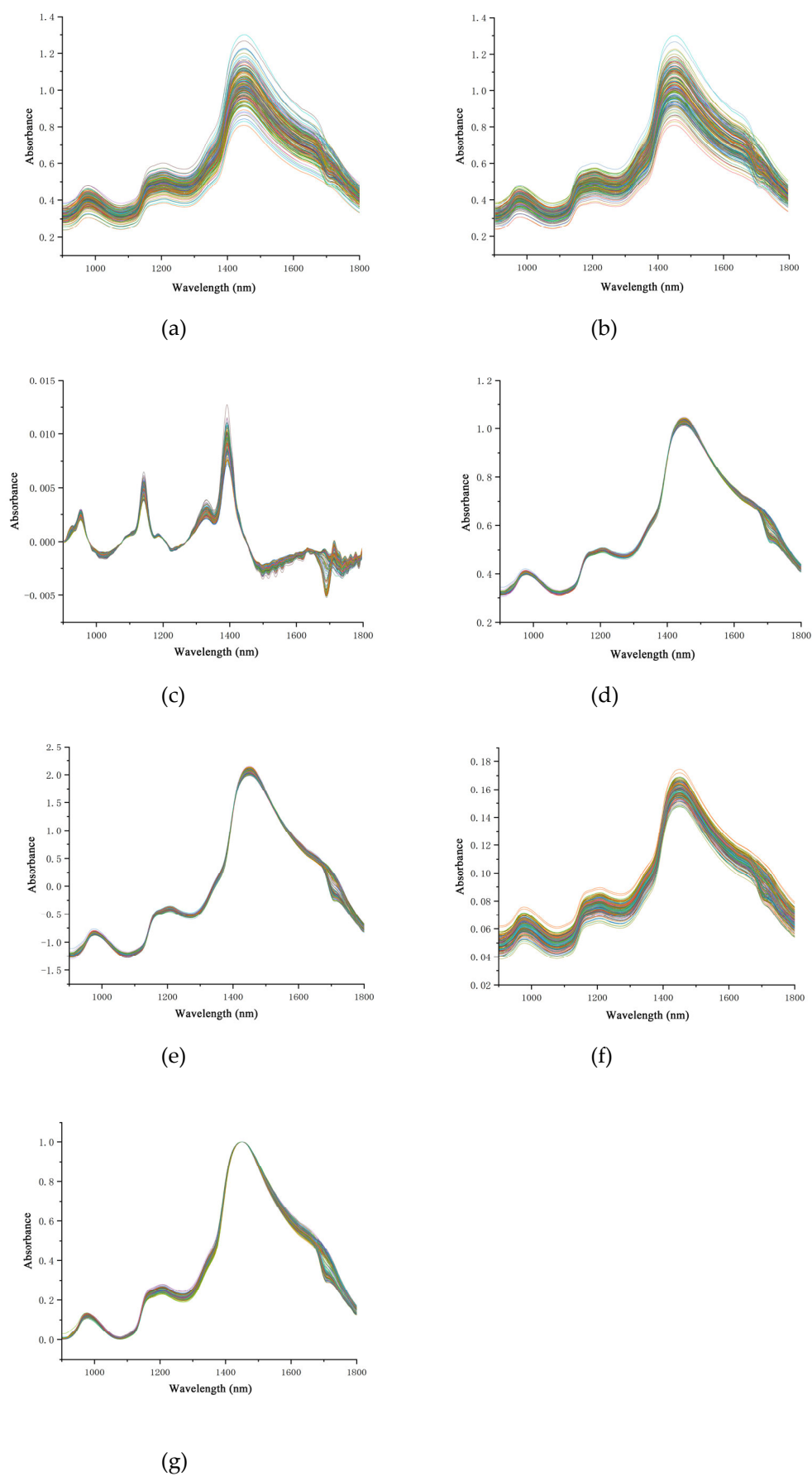


Figure 2. NIRS under different pre-processing methods: (a) original spectrum, (b) SGCS, (c) SGCD, (d) MSC, (e) SNV, (f) VN, and (g) MMN.

3.2. Detection of the Color Index of Korla Fragrant Pears

As shown in Table 1, the minimum, maximum, and mean $\pm$ standard deviation values of color parameters L^* , a^* , and b^* are presented for 220 Korla fragrant pear samples. The L^* values ranged from 51.1 to 63.71 (mean: 58.12 ± 2.32), a^* values spanned a range of -6.21 to 7.33 (mean: -1.24 ± 2.69), and b^* values varied between 33.57 and 48.30 (mean: 42.45 ± 2.38). The significant variation in Korla fragrant pear color demonstrates a strong representativeness of the dataset, providing a reliable foundation for correlation analysis between near-infrared spectroscopy data and color parameters.

Table 1. Color characteristics of Korla fragrant pears.

Indicator	Maximum	Minimum	Average	Standard Deviation
L^*	63.71	51.1	58.12	2.32
a^*	7.33	-6.21	-1.24	2.69
b^*	48.30	33.57	42.45	2.38

3.3. Color Index Detection of Pears After Pre-Processing

The results of detecting the color indices of the pears after six pre-processing methods are shown in Table 2. Compared to PLSR models based on original spectral data, those established after applying six pre-processing methods demonstrated improved detection accuracy. For the color index L^* , the established detection model based on the original spectral data had an R and RMSE of 0.68 and 1.58, respectively. All pre-processing methods improved the detection accuracy of the model, with the VN pre-processing method yielding the highest R and the lowest RMSE, 0.70 and 1.41, respectively. For the color index a^* , the established detection model based on the original spectral data had an R and RMSE of 0.69 and 1.70, respectively. All pre-processing methods improved the detection accuracy of the model, with the SGCS pre-processing method yielding the highest R and the lowest RMSE, 0.74 and 1.48, respectively. For the color index b^* , the established detection model based on the original spectral data had an R and RMSE of 0.71 and 1.76, respectively. All pre-processing methods, except SNV, improved the detection accuracy of the model, with the SGCD pre-processing method yielding the highest R and the lowest RMSE of 0.75 and 1.51, respectively.

Table 2. Test results of the color indices L^* , a^* , and b^* under six pre-processing methods.

Test Index	Pre-Processing Method	Test Set	
		R	RMSE
L^*	No pre-processing	0.68	1.58
	SGCD	0.70	1.45
	SGCS	0.68	1.42
	VN	0.70	1.40
	MMN	0.68	1.54
	MSC	0.69	1.41
	SNV	0.68	1.51
a^*	No pre-processing	0.69	1.70
	SGCD	0.71	1.62
	SGCS	0.74	1.48
	VN	0.73	1.54
	MMN	0.72	1.48
	MSC	0.75	1.59
	SNV	0.71	1.51

Table 2. Cont.

Test Index	Pre-Processing Method	Test Set	
		R	RMSE
b*	No pre-processing	0.71	1.76
	SGCD	0.75	1.51
	SGCS	0.72	1.58
	VN	0.73	1.50
	MMN	0.71	1.69
	MSC	0.73	1.62
	SNV	0.69	1.78

3.4. Detection of Pear Color Indices After Characteristic Band Extraction

After pre-processing the Korla fragrant pear spectral data using six methods—SGCS, SGCD, MSC, SNV, VN, and MMN, the accuracy of the pear color detection model improved, but the results were still not ideal. This may be due to redundant information in the spectral data, which includes irrelevant variables that affect the model's detection accuracy. To further improve the model's accuracy, this study combined the six pre-processing methods with UVE and SPA characteristic band extraction methods.

3.4.1. Detection of Pear Color Indices Based on UVE

After UVE characteristic band extraction, 52 characteristic bands were extracted for the color index L^* , reducing the data volume by more than 94.3%. For the color index a^* , 55 characteristic bands were extracted, reducing the data volume by 93.9%. For the color index b^* , 52 characteristic bands were extracted, reducing the data volume by more than 94.3%. As shown in Table 3, the SGCD-UVE-PLSR model had the highest R-value and the lowest RMSE value for detecting the color index L^* , which were 0.80 and 1.19, respectively. For detecting the color index a^* , the SNV-UVE-PLSR model had the highest R-value and a relatively low RMSE value, 0.83 and 1.51, respectively. For the detection of the color index b^* , the MSC-UVE-PLSR model had the highest R-value and a relatively low RMSE value, 0.84 and 1.25, respectively. In summary, after UVE characteristic band extraction, the detection accuracy of the PLSR model was effectively improved. The best detection model for the color index L^* was SGCD-UVE-PLSR, the best detection model for the color index a^* was SNV-UVE-PLSR, and the best detection model for the color index b^* was MSC-UVE-PLSR.

Table 3. Test results of fragrant pear color indices based on the UVE method.

Test Index	Pre-Processing Method	Test Set	
		R	RMSE
L^*	SGCS	0.73	1.54
	SGCD	0.80	1.19
	VN	0.73	1.49
	MMN	0.76	1.35
	MSC	0.75	1.41
	SNV	0.78	1.32
a^*	SGCS	0.79	1.69
	SGCD	0.81	1.51
	VN	0.79	1.58
	MMN	0.83	1.56
	MSC	0.82	1.52
	SNV	0.83	1.51

Table 3. Cont.

Test Index	Pre-Processing Method	Test Set	
		R	RMSE
b*	SGCS	0.83	1.42
	SGCD	0.84	1.28
	VN	0.81	1.31
	MMN	0.84	1.34
	MSC	0.84	1.25
	SNV	0.81	1.42

3.4.2. Detection of Pear Color Indices Based on the SPA

After SPA characteristic band extraction, 50 characteristic bands were extracted for the color index L^* , reducing the data volume by more than 94.5%. For the color index a^* , 49 characteristic bands were extracted, reducing the data volume by 94.6%. For the color index b^* , 50 characteristic bands were extracted, reducing the data volume by more than 94.5%. As shown in Table 4, the MMN-SPA-PLSR model had the highest R-value and the lowest RMSE value for detecting the color index L^* , which were 0.78 and 1.34, respectively. To detect the color index a^* , the VN-SPA-PLSR model had the highest R-value and the lowest RMSE value, which were 0.84 and 1.28, respectively. For the detection of the color index b^* , the SGCD-SPA-PLSR model had the highest R-value and a relatively low RMSE value, which were 0.83 and 1.28, respectively. In summary, after SPA characteristic band extraction, the detection accuracy of the PLSR model improved effectively. The best detection model for the color index L^* was MMN-SPA-PLSR, the best detection model for the color index a^* was VN-SPA-PLSR, and the best detection model for the color index b^* was SGCD-SPA-PLSR.

Table 4. Test results of fragrant pear color index based on the SPA method.

Test Index	Pre-Processing Method	Test Set	
		R	RMSE
L^*	SGCS	0.74	1.42
	SGCD	0.72	1.34
	VN	0.78	1.39
	MMN	0.78	1.34
	MSC	0.74	1.31
	SNV	0.73	1.37
a^*	SGCS	0.80	1.59
	SGCD	0.79	1.65
	VN	0.84	1.28
	MMN	0.81	1.52
	MSC	0.80	1.49
	SNV	0.83	1.46
b^*	SGCS	0.84	1.42
	SGCD	0.83	1.28
	VN	0.81	1.29
	MMN	0.79	1.37
	MSC	0.83	1.36
	SNV	0.82	1.29

3.4.3. Optimal Detection Models

Combining the six pre-processing methods with UVE and the SPA effectively improved the detection accuracy of the PLSR model. The optimal detection results for the pear color indices are shown in Table 5. For the detection of the color index L^* , the SGCD-UVE-PLSR model had the highest R-value and the lowest RMSE value, which were 0.80 and 1.19, respectively, along with an RPD of 1.95 and an RER of 10.60, which indicate that the model exhibits good detection accuracy and practical applicability. To detect the color index a^* , the VN-SPA-PLSR model had the highest R-value and the lowest RMSE value, which were 0.84 and 1.28, respectively, along with an RPD of 2.10 and an RER of 10.58, which demonstrate that the model achieves strong detection accuracy and practical applicability. For the detection of the color index b^* , the MSC-UVE-PLSR model had the highest R-value and the lowest RMSE value, which were 0.84 and 1.25, respectively, along with an RPD value of 1.90 and an RER value of 11.78, which validate the model's robust detection accuracy.

Table 5. Optimal detection results of fragrant pear color indices.

Test Index	Pre-Processing Method	Characteristic Extract Extraction	Training Set		Test Set		RPD	RER
			R	RMSE	R	RMSE		
L^*	MMN	SPA	0.78	1.46	0.78	1.34	1.73	9.41
	SGCD	UVE	0.84	1.15	0.80	1.19	1.95	10.60
a^*	VN	SPA	0.85	1.39	0.84	1.28	2.10	10.58
	SNV	UVE	0.85	1.49	0.83	1.51	1.78	8.97
b^*	SGCD	SPA	0.85	1.26	0.83	1.28	1.86	11.51
	MSC	UVE	0.84	1.21	0.84	1.25	1.90	11.78

Based on the comparison of the detection model results, the optimal detection model for each color index was determined. The optimal detection model for the color index L^* was SGCD-UVE-PLSR, the optimal detection model for the color index a^* was VN-SPA-PLSR, and the optimal detection model for the color index b^* was MSC-UVE-PLSR.

This study determined the optimal number of LVs based on the J-Score index method [31]. The J-Score comprehensively evaluates the cross-validation error (RMSEcv), the calibration-to-validation error ratio (reflecting overfitting risk), and the regression vector noise index. The model for detecting the color parameter L^* reached its minimum J-Score value (0.23) at LVs = 9, color parameter a^* achieved its minimum J-Score (0.24) at LVs = 7, and color parameter b^* attained its minimum J-Score (0.23) at LVs = 8, indicating that the model achieved the optimal balance between prediction accuracy and robustness at these points.

Figures 3–5 display scatterplots of predicted versus measured values for the optimal prediction models of Korla fragrant pear color parameters L^* , a^* , and b^* , respectively. These plots visually demonstrate the models' generalization capability on the test set through scatter distribution matrices.

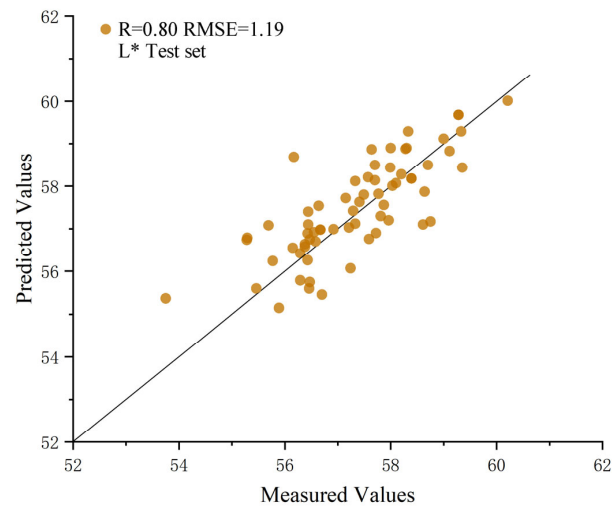


Figure 3. Comparison of predicted and measured values of the optimal prediction model for the color parameter L^* .

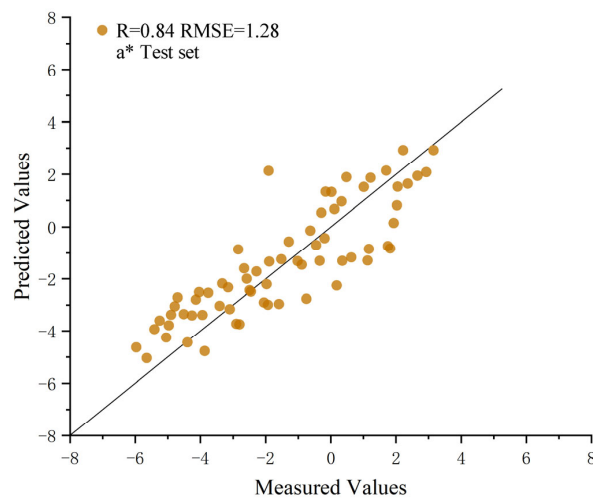


Figure 4. Comparison of predicted and measured values of the optimal prediction model for the color parameter a^* .

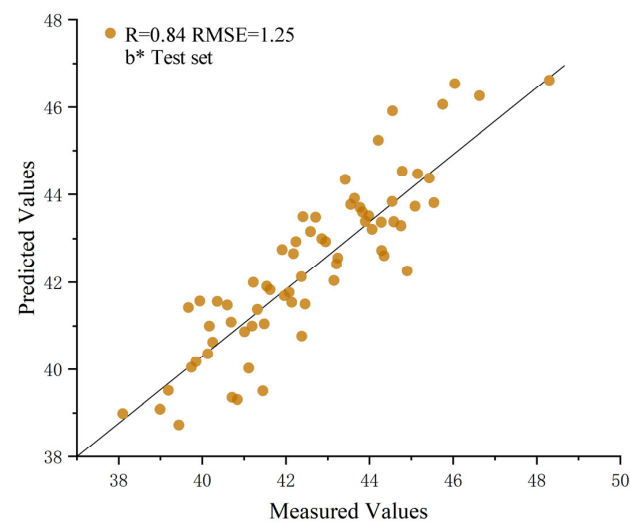


Figure 5. Comparison of predicted and measured values of the optimal prediction model for the color parameter b^* .

4. Discussion

Machine vision technology is widely applied in fruit color detection; yet, its accuracy and consistency are often compromised by lighting variations, surface heterogeneity, and differences in fruit shape and size [32]. Additionally, while fruit quality assessment requires parameters such as hardness, sugar content, and vitamins, machine vision cannot comprehensively detect these indicators online [7]. In contrast, near-infrared spectroscopy (NIRS) enables simultaneous multi-parameter detection. For instance, Liu et al. [33] demonstrated NIRS's capability to non-destructively predict soluble solid content, hardness, and moisture content during pear ripening, while Hao et al. [34] achieved accurate quantitative predictions of pear hardness and color. These studies highlight NIRS's superiority over machine vision for multi-quality fruit detection.

The effectiveness of NIRS combined with chemometric methods is further supported by recent advances. Liu et al. [12] identified PLSR as the optimal model for non-destructive color detection in Fengshui pears, attributing its success to handling high-dimensional multicollinear spectral data—a finding consistent with our study. Pre-processing methods (e.g., SGCS, SNV, and SGCD) and feature selection techniques (SPA and UVE) enhanced model accuracy by reducing noise and extracting informative spectral bands, as evidenced in studies on kiwi fruit [17], coco peat [35], and Korla pears [21]. Our integration of these strategies—spectral pre-processing to improve data quality, feature extraction to reduce complexity, and PLSR to address multicollinearity—significantly improved the pear color detection model's robustness, underscoring the synergy of spectral optimization and algorithmic design.

This study provides a theoretical reference for developing Korla fragrant pear color detection instruments. However, due to potential differences in pears grown in different environments and conditions, the current model's applicability needs validation with samples from various regions. Future research should include samples from different locations to further improve model accuracy and applicability and to optimize algorithms by incorporating advanced machine learning frameworks to enhance detection accuracy and processing efficiency.

5. Conclusions

Compared to the pear color detection model established by the original spectral data, the spectral data processed using SGCS, SGCD, MSC, SNV, VN, and MMN methods combined with the SPA and UVE methods resulted in higher-precision detection models. The SGCS-UVE-PLSR model exhibits high detection accuracy and strong practical applicability for the color index L^* ($R = 0.80$, $RMSE = 1.19$, $RPD = 1.95$, and $RER = 10.60$). The VN-SPA-PLSR model achieved the highest accuracy for detecting the color index a^* ($R = 0.84$, $RMSE = 1.28$, $RPD = 2.10$, and $RER = 10.58$), and the MSC-UVE-PLSR model was most effective for detecting the color index b^* ($R = 0.84$, $RMSE = 1.25$, $RPD = 1.90$, and $RER = 11.78$). In this study, advanced spectral pre-treatment methods were integrated (SGCD, MSC, VN, etc.) with UVE and SPA feature selection. The color detection model of Korla fragrant pears based on near-infrared spectra was established. Building upon the achievement of near-infrared spectroscopy technology in the non-destructive detection of hardness and soluble solid content of Korla fragrant pears, detection models for the color parameters (L^* , a^* , and b^*) were further developed, which can be utilized for quality grading of Korla fragrant pears. Future efforts will prioritize optimizing algorithm efficiency to meet industrial speed requirements, expanding sample diversity to ensure cross-regional applicability, and validating the framework's scalability for other fruit varieties. These advancements aim to translate laboratory precision into field-ready solutions, ultimately

supporting standardized quality control, minimizing postharvest losses, and aligning with the food industry's demand for rapid non-destructive inspection technologies.

Author Contributions: Conceptualization, Y.X. and Y.L.; software, J.C.; validation, Y.X., J.C. and Q.L.; formal analysis, H.Z.; resources, Y.L. and Y.X.; data curation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, Y.L.; visualization, H.Z.; supervision, H.Z. and Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Chinese Natural Science Foundation: 32260618 and 32202139; the Team Project of the President Fund of Tarim University: TDZKCX202203; the Tarim University President Fund Project: TDZKSS202427; and the Bingtuan Guiding Science and Technology Plan Program 2022ZD094.

Data Availability Statement: No new data were created or analyzed in this study. Data will be made available upon request.

Acknowledgments: The authors acknowledge the support provided by the Horizontal Project. The authors are grateful to the anonymous reviewers for their comments.

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

References

1. Wang, B.H.; Sun, X.X.; Dong, F.Y.; Zhang, F.; Niu, J. Cloning and expression analysis of an MYB gene associated with calyx persistence in Korla fragrant pear. *Plant Cell Rep.* **2014**, *33*, 1333–1341. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Sardar, H. Fruit Quality Estimation by Color for Grading. *Int. J. Model. Optim.* **2014**, *4*, 38–42. [\[CrossRef\]](#)
3. Anjali; Ankita, J.; Ayushi, B.; Sadhna, M.; Ishika, J.; Nandini, P.; Nishita, S.; Nitiksha, J.; Renu, P.; Shakshi, K.; et al. State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: Principles, applications, and future directions. *Food Prod. Process. Nutr.* **2024**, *6*, 56. [\[CrossRef\]](#)
4. Huang, J.; Li, X.J. The Korla Fragrant Pear Color Grading Based on Colorimeter. *North. Hortic.* **2018**, *17*, 38–44.
5. Qu, N.W.; Ma, B.X.; Wang, W.X. Color Grading of Korla Fragrant Pears Based on Neural Network and Machine Vision. *J. Shihezi Univ. (Nat. Sci. Ed.)* **2010**, *28*, 514–518. [\[CrossRef\]](#)
6. Guo, S.P.; Huang, C.J.; Lu, Z.P. Application of Color Difference Meter in the Evaluation of Eggplant Color. *Chin. Fruit Veg.* **2022**, *42*, 61–65. [\[CrossRef\]](#)
7. Yang, Z.Q.; Li, Z.M.; Hu, N.; Zhang, M.X.; Zhang, W.B.; Gao, L.X.; Ding, X.Y.; Qi, Z.P.; Duan, S.Y. Multi-Index Grading Method for Pear Appearance Quality Based on Machine Vision. *Agriculture* **2023**, *13*, 290. [\[CrossRef\]](#)
8. Shao, Y.Y.; Ji, S.H.; Shi, Y.K.; Xuan, G.T.; Jia, H.J.; Guan, X.L.; Chen, L. Growth period determination and color coordinates visual analysis of tomato using hyperspectral imaging technology. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2024**, *319*, 124538. [\[CrossRef\]](#)
9. Rodríguez, P.; Villamizar, J.; Londoño, L.; Tran, T.; Davrieux, F. Quantification of Dry Matter Content in Hass Avocado by Near-Infrared Spectroscopy (NIRS) Scanning Different Fruit Zones. *Plants* **2023**, *12*, 3135. [\[CrossRef\]](#)
10. Patel, D.; Bhise, S.; Kapdi, S.S.; Bhatt, T. Non-destructive hyperspectral imaging technology to assess the quality and safety of food: A review. *Food Prod. Process. Nutr.* **2024**, *6*, 69. [\[CrossRef\]](#)
11. Nicácia, P.M.; José, C.F.; Simone, P.G.; Débora, L.B.; Mateus, S.P.; Juliano, D.S. Pear quality characteristics by Vis/NIR spectroscopy. *An. Acad. Bras. Ciênc.* **2012**, *84*, 853–863. [\[CrossRef\]](#)
12. Liu, Y.D.; Chen, X.M.; Ouyang, A.G. Non-Destructive Measurement of Surface Color of Pear by Visible/Near-Infrared Diffuse Reflectance Spectra. *J. Infrared Millim. Waves* **2008**, *4*, 266–268.
13. Che, J.; Liang, Q.; Xia, Y.; Liu, Y.; Li, H.; Hu, N.; Cheng, W.; Zhang, H.; Zhang, H.; Lan, H. The Study on Nondestructive Detection Methods for Internal Quality of Korla Fragrant Pears Based on Near-Infrared Spectroscopy and Machine Learning. *Foods* **2024**, *13*, 3522. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kusumiyati, K.; Indah, K.; Oktavia, A.R. The color detection of two cucumber cultivars by NIR Spectroscopy. *Asian J. Agric.* **2017**, *1*, 59–65. [\[CrossRef\]](#)
15. Chen, C.; Lu, X.X.; Zhang, P.; Chen, S.H.; Li, J.K. Measurement of surface color of Manai grape by VIS/NIR diffuse reflectance spectra. *Food Ind. Div.* **2019**, *36*, 308–311. [\[CrossRef\]](#)
16. Mancini, M.; Mazzoni, L.; Gagliardi, F.; Balducci, F.; Duca, D.; Toscano, G.; Mezzetti, B.; Capocasa, F. Application of the Non-Destructive NIR Technique for the Evaluation of Strawberry Fruits Quality Parameters. *Foods* **2020**, *9*, 441. [\[CrossRef\]](#)

17. Zeng, S.C.; Zhang, Z.Y.; Cheng, X.D.; Cai, X.; Cao, M.K.; Guo, W.C. Prediction of soluble solids content using near-infrared spectra and optical properties of intact apple and pulp applying PLSR and CNN. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *304*, 123402. [\[CrossRef\]](#)
18. Liu, F.F. Key Technology Research and Application of Real-Time Non-Destructive Detection for Kiwifruit Sugar. Master's Thesis, Xijing University, Xi'an, China, 2023. [\[CrossRef\]](#)
19. Ping, F.J.; Yang, J.H.; Zhou, X.J. Quality Assessment and Ripeness Prediction of Table Grapes Using Visible-Near-Infrared Spectroscopy. *Foods* **2023**, *12*, 2364. [\[CrossRef\]](#)
20. Toktam, M.M.; Seyed, M.A.R.; Ameneh, S.; Masoud, T. Predicting the moisture content and textural characteristics of roasted pistachio kernels using Vis/NIR reflectance spectroscopy and PLSR analysis. *J. Food Meas. Charact.* **2018**, *12*, 346–355. [\[CrossRef\]](#)
21. Xie, C.J.; Qiao, M.M.; Yang, L.; Zhang, D.X.; Cui, T.; He, X.T.; Du, Z.H.; Xiao, T.P.; Li, H.S. Establishment of a general prediction model for protein content in various varieties and colors of peas using visible-near-infrared spectroscopy. *J. Food Compos. Anal.* **2024**, *127*, 105965. [\[CrossRef\]](#)
22. Chen, D.J.; Jiang, P.H.; Guo, F.J.; Zhang, Y.H.; Zhang, C.F. Effects of Prediction Model of Kolar Pear Based on NIR Diffuse Transmission under Different Moving Speed on Online. *Spectrosc. Spectr. Anal.* **2020**, *40*, 1839–1845. Available online: <https://www.gpxygpfx.com/CN/abstract/abstract11387.shtml> (accessed on 24 March 2025).
23. Chen, Q.M. Leaf Color Analysis and Color Application of Four Japanese Colorful Leaf Maple Based on Lab Model. *Shandong For. Sci. Technol.* **2019**, *49*, 37–40.
24. Wang, F.; Liu, F.; Zhang, Y.Y.; Zhang, A.; Cao, Y.Z.; Li, J.Q.; Zhang, S.B. The study of applying CIE 1976 (L*a*b*) colour space in the measurement of colour of wine. *Sino-Overseas Grapevine Wine* **2015**, *4*, 6–11. [\[CrossRef\]](#)
25. Sun, X.; Subedi, P.; Walker, R.; Walsh, K.B. NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment. *Postharvest Biol. Technol.* **2020**, *163*, 111140. [\[CrossRef\]](#)
26. Peng, X.T.; Shi, T.Z.; Song, A.H.; Chen, Y.Y.; Gao, W.X. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **2014**, *6*, 2699–2717. [\[CrossRef\]](#)
27. Li, Q.Q.; Huang, Y.; Tian, K.D. Optimal modeling pattern of variables selection on analog complex using UVE-PLS regression. *IOP SciNotes* **2020**, *1*, 014201. [\[CrossRef\]](#)
28. Wang, H.L.; Peng, J.P.; Xie, C.Q.; Bao, Y.D.; He, Y. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. *Sensors* **2015**, *15*, 11889–11927. [\[CrossRef\]](#)
29. Douglas, R.K.; Nawar, S.; Alamar, M.C.; Coulon, F.; Mouazen, A.M. Rapid detection of alkanes and polycyclic aromatic hydrocarbons in oil- contaminated soil with visible near-infrared spectroscopy. *Eur. J. Soil Sci.* **2019**, *70*, 140–150. [\[CrossRef\]](#)
30. Wang, Z.Z.; Wu, Q.Y.; Kamruzzaman, M.H. Portable NIR spectroscopy and PLS based variable selection for adulterati on detection in quinoa flour. *Food Control* **2022**, *138*, 108970. [\[CrossRef\]](#)
31. Ezenarro, J.; Schorn-García, D.; Aceña, L.; Mestres, M.; Busto, O.; Boqué, R. J-Score: A new joint parameter for PLSR model performance evaluation of spectroscopic data. *Chemom. Intell. Lab. Syst.* **2023**, *240*, 104883. [\[CrossRef\]](#)
32. Lukinac, J.; Mastanjević, K.; Mastanjević, K.; Nakov, G.; Jukić, M. Computer Vision Method in Beer Quality Evaluation—A Review. *Beverages* **2019**, *5*, 38. [\[CrossRef\]](#)
33. Liu, D.Y.; Wang, E.F.; Wang, G.L.; Ma, G.K. Nondestructive determination of soluble solids content, firmness, and moisture content of “Longxiang” pears during maturation using near-infrared spectroscopy. *J. Food Process. Preserv.* **2022**, *46*, e16332. [\[CrossRef\]](#)
34. Hao, Y.; Sun, X.D.; Pan, Y.Y.; Gao, R.J.; Liu, Y.D. Detection of Firmness and Surface Color of Pear by Near Infrared Spectroscopy Based on Monte Carlo Uninformative Variables Elimination Method. *Spectrosc. Spectr. Anal.* **2011**, *31*, 1225–1229.
35. Lu, B.; Liu, N.H.; Li, H.L.; Yang, K.F.; Hu, C.; Wang, X.F.; Li, Z.X.; Shen, Z.X.; Tang, X.Y. Quantitative determination and characteristic wavelength selection of available nitrogen in coco-peat by NIR spectroscopy. *Soil Tillage Res.* **2019**, *191*, 266–274. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.