



Article Nondestructive Testing Model of Mango Dry Matter Based on Fluorescence Hyperspectral Imaging Technology

Zhiliang Kang, Jinping Geng, Rongsheng Fan, Yan Hu 몓, Jie Sun, Youli Wu, Lijia Xu *២ and Cheng Liu

College of Mechanical and Electrical Engineering, Sichuan Agriculture University, Ya'an 625000, China

* Correspondence: xulijia@sicau.edu.cn; Tel.: +86-138-8160-5841

Abstract: The dry matter test of mango has important practical significance for the quality classification of mango. Most of the common fruit and vegetable quality nondestructive testing methods based on fluorescence hyperspectral imaging technology use a single algorithm in algorithms such as Uninformative Variable Elimination (UVE), Random Frog (RF), Competitive Adaptive Reweighted Sampling (CARS) and Continuous Projection Algorithm (SPA) to extract feature spectral variables, and the use of these algorithms alone can easily lead to the insufficient stability of prediction results. In this regard, a nondestructive detection method for the dry matter of mango based on hyperspectral fluorescence imaging technology was carried out. Taking the 'Keitt' mango as the research object, the mango samples were numbered in sequence, and their fluorescence hyperspectral images in the wavelength range of 350–1100 nm were collected, and the average spectrum of the region of interest was used as the effective spectral information of the sample. Select SPXY algorithm to divide samples into a calibration set and prediction set, and select Orthogonal Signal Correction (OSC) as preprocessing method. For the preprocessed spectra, the primary dimensionality reduction (UVE, SPA, RF, CARS), the primary combined dimensionality reduction (UVE + RF, CARS + RF, CARS + SPA), and the secondary combined dimensionality reduction algorithm ((CARS + SPA)-SPA, (UVE + RF)-SPA) and other 12 algorithms were used to extract feature variables. Separately constructed predictive models for predicting the dry matter of mangoes, namely, Support Vector Regression (SVR), Extreme Learning Machine (ELM), and Back Propagation Neural Network (BPNN) model, were used; The results show that (CARS + RF)-SPA-BPNN has the best prediction performance for mango dry matter, its correlation coefficients were $R_C^2 = 0.9710$, $R_P^2 = 0.9658$, RMSEC = 0.1418, RMSEP = 0.1526, this method provides a reliable theoretical basis and technical support for the non-destructive detection, and precise and intelligent development of mango dry matter detection.

Keywords: mango; dry matter; fluorescence hyperspectral; OSC algorithm; BPNN; nondestructive detection

1. Introduction

Mango, the 'king of fruits', is a climacteric tropical fruit with a short shelf life [1]; therefore, mango fruit are usually harvested at the hard green stage (unripe), when they are physiologically mature, but before the onset of the climacteric rise [2]. Fruit picked before physiological maturity will not ripen properly, leading to a poor eating experience. Fruit picked immediately after physiological maturity will ripen, but a further delay in harvest date allows fruit carbohydrate reserves to increase, and such fruit attain a superior eating quality. Fruit dry matter can be used in harvesting time estimation and has been linked to eating quality when the fruitsy are ripe [3]. Dry matter (DM) is the weight of all tissue components except water, but as constitutional components such as cell walls and membranes are relatively constant with fruit maturation. Dry matter is a useful index of soluble and insoluble carbohydrate contents, i.e., sugar and starch in mango fruit. DM with fruit harvest is well correlated with ripened fruit Brix and eating quality [4]. Attempts to maximize profits by supplying fruit at higher market prices may result in the placement of



Citation: Kang, Z.; Geng, J.; Fan, R.; Hu, Y.; Sun, J.; Wu, Y.; Xu, L.; Liu, C. Nondestructive Testing Model of Mango Dry Matter Based on Fluorescence Hyperspectral Imaging Technology. *Agriculture* **2022**, *12*, 1337. https://doi.org/10.3390/ agriculture12091337

Academic Editor: Hongbin Pu

Received: 14 July 2022 Accepted: 22 August 2022 Published: 30 August 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/). under ripe or overripe fruit on the market. Early harvested fruit has a longer shelf life but a lower dry matter. A low DM in fruit can lead to customer dissatisfaction and a reduced buyback rate, so a quick and non-destructive test of the dry matter in mangoes is essential.

More recently, hyperspectral imaging (HSI) [5] is being regarded as an effective nondestructive method to determine the quality of agricultural products. The advantages of HSI are the simultaneous procurement of spatial and spectral information from the sample and the flexibility of area selection for spectral extraction after data acquisitions [6]. Overall, HSI can please provide information on the spatial distribution of physicochemical parameters, enhancing the perception of quality changes within samples [7]. Use of hyperspectral technology to detect agricultural product quality is still in its infancy. However, as fast, and nondestructive detection technology, hyperspectral imaging has great application prospects. There has been only a small amount of research on agricultural product quality detection based on hyperspectral technology [8]. For mango, moisture distribution of dried mango slices [9] and post-harvest quality including color, firmness and TSS [10]. Sharma, et al. [11] investigated the potential of push-sweep near-infrared hyperspectral imaging (NIR-HSI) systems (900–1600 nm) in the classification of maturing stages (under ripe, maturation and overcooking). The performance of five supervised machine learning classifiers, including Support Vector Machines (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), and K-Nearest Neighbor (KNN), for mature stage classification, was developed with Partial Least Squares Regression (PLSR) models for DM prediction. For classification, the LDA showed the best results, with a test accuracy of 100% for both the resonant wavelength and the 135 wavelengths selected by the genetic algorithm. The prediction coefficient (R_p^2) of the PLSR model predicts DM is greater than 0.80, and the prediction root mean square error (RMSEP) is less than 1.6%.

However, hyperspectral still has problems such as the detection peak is not obvious, the noise interference is too large, and the experimental detection speed is slow [12]. In order to solve these problems, some scholars have proposed a method of combining fluorescence spectroscopy with hyperspectral imaging technology. This is because the fluorescence lifetime reflects the decay time of the fluorescence photon, which is only related to the intensity of the excitation light and is not subject to environmental light, fluorescence scattering and other factors. It has the advantages of high sensitivity, high selectivity, ease of use, good reproducibility and good stability [13], and a variety of fluorescence spectroscopic analysis techniques have been applied in the rapidly developing including field of atomic fluorescence spectrum [14], two-dimensional correlation fluorescence spectra [15], three-dimensional fluorescence spectra [16,17], front surface fluorescence spectra [17], (total) synchronous fluorescence spectra [18] and fluorescence hyperspectral imaging technology [19]. This study is based on fluorescence hyperspectral imaging techniques. This technology can use computers to simulate human visual functions without damaging test samples. This technology can extract information such as color, texture, fluorescence intensity and so on from fluorescence imaging for preprocessing and analysis for practical detection. Wang, et al. [20] proposed the use of fluorescence hyperspectral imaging techniques to quantitatively predict the pH value of kiwi fruit nondestructively. The results show that IVMR-VISSA-IRIV-MK-SVR had the best prediction results, with a R_P^2 , R_C^2 and RPD of 0.8512, 0.8580 and 2.66, respectively. Zhuang, et al. [21] investigates the inspection of frozen pork quality attributes without thawing using fluorescence hyperspectral imaging (HSI). Fluorescence hyperspectral imaging technology will continue to play an important role in food nondestructive testing.

Most of the common nondestructive detection methods for fruits and vegetables based on fluorescence hyperspectral imaging technology are Uninformative Variable Elimination (UVE) [22], Random Forest (RF) [23], Competitive Adaptive Reweighted Sam-pling (CARS) [24], Successive Projections Algorithm (SPA) [22], and other single algorithms for feature extraction, combined with Support Vector Regression (SVR) [25] to build a detection model, these algorithms are prone to instability deficiencies when used alone. For example, The Monte Carlo sampling process in the CARS algorithm has randomness, so the characteristic spectral variables proposed by it have a certain degree of randomness; the spectral variables selected by the SPA are easily doped with no information variables, which can reduce the reliability of the predictive model. Therefore, it is very necessary to find a better hyperspectral data feature extraction method and a more suitable detection model to solve the problem of nondestructive testing of mango dry matter. For this reason, this paper adopts the idea of complementary advantages and disadvantages between the algorithms, uses UVE and RF algorithms to eliminate the information-free variables and interference variables in the spectral variables, and directly combines them with the characteristic spectral variables selected by the CARS algorithm to increase the spectral variables rich in the dry matter of mango information. The SPA algorithm is used to degrade the quadratic dimensionality reduction of the combined characteristic spectral variables formed by CARS + RF, UVE + RF and CARS + SPA, which eliminates collinearity between variables and improves the stability of the algorithm. By dividing the spectral variables of the selected characteristics of different dimensionality reduction methods, the Back Propagation Neural Network (BPNN) [26] for predicting the dry matter of mango is constructed. Compared with the prediction results of the models built by the support vector regression (SVR), and the Extreme Learning Machine (ELM), the optimal prediction method is determined. In the field of mango quality testing, this (UVE + RF)-SPA-BPNN method has not been reported, in order to provide a new technical means for nondestructive testing of dry matter of mango.

2. Materials and Methods

2.1. Samples

The experimental material is 'Keitt' mango freshly collected at the mango planting base in Pan-zhihua, Sichuan Province. A total of, 120 mango samples of similar size are selected and numbered sequentially, the epidermis of the mango is wiped with a semidry towel, and it is placed in the laboratory for 24 h, waiting for the acquisition of its fluorescent hyperspectral imaging and then measuring its dry matter, the ambient temperature during the experiment (20 ± 1) °C, the humidity is maintained at 56% to 58%, and the cross painted at the equator of the mango is selected as the collection site of the mango physical and chemical data. Figure 1 is sample 70.



Figure 1. (a) Fluorescence hyperspectral system. (b) Raw hyperspectral fluorescent images. (c) Extraction of fluorescence spectra information of mango.

2.2. Fluorescence Hyperspectral Imaging Acquisition

Fluorescence hyperspectral data of mango samples were acquired using the GaiaFluo (/Pro)-VH-HR series fluorescence hyperspectral test system produced by Jiangsu Dualix Technology Co., Ltd. (Wuxi, China) [19], as shown in Figure 1a. Spectral imaging system sampling built-in push-sweep imaging mode, spectrometer and Scmos detector composed of imaging part under the control of a two-dimensional precision electronic control displacement table can complete the focal length (focusing motor) and push sweep (scanning motor) imaging tasks; The computer completes the collection and storage of the data; Various data preprocessing and calibration of the generated data cubes can be performed: reflectance calibration, radiometric calibration, lens, uniformity calibration, etc. In the system, the xenon lamp light source was used as the excitation light source for the fluorescence imaging system, with a detectable spectral range from 250 nm to 1100 nm [27]. The hyperspectral camera has the advantages of high sensitivity and a high signal-to-noise ratio in the 350 nm to 1100 nm band [28]. By combining multiple excitation filters and fluorescence filters, it was found that under the illumination of four different wavelength bands of excitation light sources in the laboratory, 390 nm excitation filters can better cut off the light input of other bands. Under the influence of excitation light sources, it is necessary in order to pay attention to the fluorescence signal of the sample. The 550 nm fluorescence filter can complete the separation of the fluorescence signal and parasitic light, allowing the final sample captured by the hyperspectral camera to produce the best fluorescence signal.

Experiments are made at an ambient temperature of 20 °C and 50% ambient humidity. The obtained fluorescence images have RGB channels of 638, 551 and 442, respectively, the system moves at a speed of 0.26 nm/s and the camera exposure time is 800 ms. The spectral resolution is 2.8 nm, the pixels are 2048×946 , the spectral range is 350 nm to 1100 nm and, the spectral sampling interval is 0.55 nm (mean).

2.3. Fluorescence Hyperspectral Data Extraction

Before analyzing the fluorescence hyperspectral, it is important to properly select the region of interest, which is directly related to the quality of the extracted data. After data acquisition, it is necessary in order to extract spectral data from the image. Reasonable selection of the region of interest (ROI) [29], which directly related to the quality of the extracted data. ROI is extracted by ENVI 5.3 [30], as shown in Figure 1c, and the spectral average of all pixels in this area calculate as the final spectral value of the sample, and the extracted spectral data corresponds to the physicochemical test data of mango, reducing the coarse error of the sample and effectively reflecting the sample information. Impacted by the fluorescent filter, the final collected spectral range is 475 nm to 1100 nm, with a total of 104 spectra channels, with a total of 125 bands (variables). For convenience, fluorescence hyperspectral images involved below are described in terms of band variable images.

2.4. Determination of the Dry Matter of Mango

Physicochemical test of mango dry matter refers to Xu, Zheng, Huang, Chen and Kang [22]. The determination method of the dry matter of kiwifruit is detected.

2.4.1. Theory

Using the physical properties of the moisture in the food, at 101.3 kPa (one atmosphere pressure), the volatile method is used to determine the weight of drying loss in the sample at 101.3 kPa (one atmosphere pressure) and 105 °C, including hygroscopic water, part of the crystalline water, and the substances that can be volatilized under this condition, and then the content of dry matter is calculated by weighing values before and after drying, As shown in Figure 2.



Figure 2. (a) The mango samples. (b) Flat weighing bottle made of glass (c) DHG-9240A electric constant temperature blast drying oven. (d) Dryer. (e) FA2304N electronic balance.

2.4.2. Instrument

Flat weighing bottle made of glass; DHG-9240A electric constant temperature blast drying oven; Dryer (an effective desiccant is included); FA2304N electronic balance (the inductive volume is 0.1 mg).

2.4.3. Experimental Steps

- 1. Remove the non-edible part of the mango, divide the equatorial cross part of the mango by quartile (as shown in Figure 2a), and then quickly chop and mix well in a porcelain dish, mash for 1–2 min, and load it into a grinding bottle as a specimen for assay.
- 2. Put a filter paper strip and a glass rod in the flat weighing bottle made of glass, place it in the 101~105 °C electric constant temperature drying box, the lid is diagonally supported on the edge of the flat weighing bottle, heated for 1.0 h, take out the lid, put it in the dryer to cool 0.5 h, weigh, re-bake 0.5 h, the same cooling, weighing. The difference in weight between the two times before and after is not more than 2 mg is a constant weight.
- 3. Determination: Take the grinding pattern 2 g, repeat 3 times (accurate to 0.0001 g), move into that flat weighing bottle, the thickness of the specimen does not exceed 5 mm, cover, after precision weighing, placed in a drying box of 101~105 °C, the cap is obliquely supported on the side of the bottle, after drying for 2~4 h, the lid is taken out, put it into the dryer to cool for 0.5 h after weighing. Then, put it into the 101~105 °C drying box for about 1 h, take it out, put it in the dryer, and cool it for 0.5 h

before weighing. And repeat the above operation until the quality difference between the two times before and after does not exceed 2 mg, that is, a constant weight. The minimum weighing time is based on the minimum. Formula (1) for calculating the dry matter of mango:

$$w_{dm} = \frac{m_2 - m_0}{m_1 - m_0} \times 100\tag{1}$$

where w_{dm} represents dry matter of mango; m_0 represents the weight of the Petri dish, g; m_1 represents the weight of the Petri dish and the sample to be measured, g; m_2 represents the weight of the Petri dish and the sample to be measured after drying, g. The measured data results are shown in Table 1.

Table 1. Statistics of mango dry matter and sample division results based on SPXY algorithm.

Sample Set	Number of Samples	Maximum	Minimum	Mean	Standard Deviation
Total	120	7.300	5.066	6.329	0.7885
Calibration set	90	7.300	5.066	6.295	0.8187
Prediction set	30	7.196	5.067	6.430	0.7098

2.5. Sample Partitioning

2.5.1. Kennard-Stone (KS)

KS algorithm [31] regards all the samples as candidates for a calibration set, and selects the two samples furthest from Euclidean to enter the calibration set. Then, by calculating the Euclidean distance between the remaining samples and the known samples in the calibration set, the two samples closest to the selected samples are selected and placed in the calibration set, repeating the above steps until the number of samples reaches the setpoint. Formula (2) for calculating the Euclidean distance is as following:

$$d_x(p,q) = \sqrt{\sum_{i=1}^n \left[x_p(i) - x_q(i) \right]^2}; p,q \in [1,n]$$
(2)

where x_p and x_q represents two different samples and represents the number of bands of the spectrum.

2.5.2. Sample Set Partitioning Based on Joint X-Y Distances (SPXY)

It was first proposed by [32], and was developed based on KS algorithm. Experiments show that SPXY algorithm can be effectively used for the establishment of hyperspectral quantitative models. SPXY takes both and variables into account when calculating sample distances. Distance equations are as Formulas (3)–(5):

$$d_{x} = \sqrt{(x_{p} - x_{q})^{2}} = |x_{p} - x_{q}|; p, q \in [1, n],$$
(3)

$$d_{y}(p,q) = \sqrt{(y_{p} - y_{q})^{2}} = |y_{p} - y_{q}|; p,q \in [1,n]$$
(4)

$$d_{x,y}(p,q) = \frac{d_x(p,q)}{\max_{p,q \in (1,n)} d_x(p,q)} + \frac{d_y(p,q)}{\max_{p,q \in (1,n)} d_y(p,q)}; p,q \in [1,n]$$
(5)

where $d_x(p,q)$ represents the spectral distance; $d_y(p,q)$ represents a chemically measured distance.

2.6. Preprocessing Methods of Spectral Data

In the process of spectral acquisition, in order to reduce random noise and disorder fluctuations in spectral data due to factors unrelated to the nature of the sample to be measured, De-trending (DT) [33], Savitzky-Golay polynomial smoothing (S-G) [34], Standard

Normal Variate (SNV) [33], Multiplicative Scatter Correction (MSC) [33], and Orthogonal Signal Correction (OSC) [35], 5 preprocessing algorithms to eliminate noise in the original spectral data, and establish the corresponding SVR prediction model respectively. All data preprocessing and modeling use matlab2020b.

2.7. Methods of Extracting Effective Variables and Modeling

Preprocessed spectral data not only have a high dimensionality, but also contain many collinear variables, which are not conducive to the establishment of the model. Therefore, it is important to reduce the dimensionality of the preprocessed spectral data to extract valid variables. Back Propagation Neural Network (BPNN) is constructed using primary dimensionality reduction (UVE, SPA, RF, CARS), quadratic combination dimensionality reduction algorithm ((CARS + SPA)-SPA, (UVE + RF)-SPA), (CARS + RF)-SPA), respectively. Support Vector Regression (SVR), and Extreme Learning Machine (ELM) [33] prediction models. The algorithm flow chart is shown in Figure 3.



Figure 3. Algorithm flowchart of this study.

2.8. Evaluate the Parameters of the Corrected Model

After selecting the calibration set samples by KS method and SPXY algorithm, the quantitative correction model between the actual dry matter of mango and the fluorescence hyperspectral spectrum was established by spectral pretreatment and feature extraction, and then the quantitative correction model between the actual dry matter of mango and the fluorescence hyperspectral was established by three methods, respectively. The model evaluation is based on the coefficient of determination (R^2) [36], and root mean square error (RMSE) [36]; the calculation methods are shown in Equations (6) and (7).

$$R^{2} = \frac{\sum_{i=1}^{N} (\hat{y}_{i} - \overline{y})^{2}}{\sqrt{\sum_{i=1}^{N} (y_{i} - \overline{y})^{2}}} = 1 - \frac{\sum_{i=1}^{N} (y_{i} - \hat{y}_{i})^{2}}{\sqrt{\sum_{i=1}^{n} (y_{i} - \overline{y}_{i})^{2}}}$$
(6)

where *N* is the number of samples; y_i is the actual value of the *i* sample; \hat{y}_i is the predicted value for the *i* sample; \bar{y} is the average of all samples. The coefficient of determination is a measure of the degree of correlation between variables, which determines the degree of closeness of the correlation, generally denoted as R^2 . In multiple regression analysis, the determining coefficient is the square of the correlation coefficient. The larger the R^2 , the higher the correlation.

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

where *N* is the number of samples; y_i is the actual value; \hat{y}_i is the predicted value. Root mean square error is a measure of the deviation between the predicted value and the actual value. The root mean square error includes root mean square error of calibration (RMSEC), the root mean square error of prediction (RMSEP), and the root mean square error of cross validation (RMSECV). The smaller the RMSEP of a model, the better the model.

3. Results

3.1. Sample Division

The outliers were discriminated and excluded using the Monte Carlo Partial Least Squares (MCPLS) [37] detection method, and the results showed no outliers, and then they were divided into 90 correction set samples and 30 prediction set samples in a 3:1 ratio with KS and SPXY, model and predict the preprocessed spectral data using SVR. Modeling results based on different preprocessing algorithms and different sample division algorithms by the coefficients of determination of the calibration set, the calibration set root mean square error (R_C^2 , RMSEC), and the prediction set determination coefficient, the prediction set root mean square error (R_P^2 , RMSEP) are shown in Table 2.

Table 2. Results of two diversity methods and five preprocessing methods.

	D	Calibration		Prediction	
Models	Preprocessing —	R _C ²	RMSEC	R _P ²	RMSEP
	Raw	0.8412	0.3247	0.5706	0.4758
	SG	0.8839	0.2798	0.6598	0.4243
	OSC	0.9138	0.2396	0.9145	0.2280
SPXY-SVR	MSC	0.7229	0.4343	0.7017	0.4688
	SNV	0.6311	0.5071	0.6186	0.4860
	DT	0.6941	0.4489	0.5962	0.4835
	Raw	0.8546	0.3105	0.6746	0.4332
	SG	0.6963	0.4445	0.6384	0.4493
KC CUD	OSC	0.9085	0.2468	0.9233	0.3194
K5-SVK	MSC	0.7990	0.3636	0.7860	0.6948
	SNV	0.6302	0.5032	0.7185	0.4010
	DT	0.6831	0.4537	0.7552	0.3687

3.2. Spectral Preprocessing

Figures 1b and 4a show the same images, both of which are raw fluorescence hyperspectral images. The difference between the two is that the abscissa of Figure 4a is the wavelength (unit: nm) and the abscissa of Figure 4a is the number of band variables. The fluorescence spectra recorded from the peel as well as pulp of all samples shows three major peaks at 540, 680, and 740 nm, as shown in Figure 4a.



Figure 4. (a) Raw hyperspectral fluorescent images; (b) DT preprocessing of spectral images; (c) S-G preprocessing of spectral images; (d) SNV preprocessing of spectral images; (e) MSC preprocessing of spectral images; (f) OSC preprocessing of spectral images.

The spectral shapes of different mangoes are similar, but the difference in their absorption peaks is very large, which indicates that the experimental samples react sensitively under fluorescence irradiation. There are two obvious peaks, about 680 nm and 740 nm, this is because chlorophyll has a bimodal distribution in the near-infrared region, and different chlorophyll levels lead to different fluorescence intensities; Chlorophyll pigments absorb light in two bands 400–500 nm and 600–700 nm and emit at 680 nm and 740 nm. In this case, the absorption has been made at 460 nm. Fluorescence should have occurred at 680 nm due to high chlorophyll content. Re-emission of the absorbed 680 nm occurs at 705 nm [39,40].

The study also focused on the behavior of the carotene fluorescence peak at 540 nm with the dry matter content of mango. Carotene is a general name for hydrocarbons, mostly founds in fruits and vegetables [41]. Mango is an excellent source of carotene. The most dominant types of carotenes reported in mangoes are violaxanthin, antheraxanthin, and β -carotene. Composition as well as concentration of carotenes changes during ripening [42–45]. Therefore, the peak intensity of carotene steadily increased during ripening, as shown in Figure 4a. Increased carotene content may be an alternative method for predicting dry matter in mangoes. In addition, the carotene fluorescence peak undergoes a

slight red shift from 540 nm to 580 nm, as shown in Figure 4d,e. Probably this could be due to higher concentration of β -carotene compared to other carotenes in ripe mango.

As can be observed in Figure 4a,b, the DT algorithm eliminates the baseline drift of the diffuse reflection spectrum. After the spectral data are processed by DT, the characteristics of peaks, and troughs are more obvious [46]. However, R_C^2 and R_P^2 are only 0.6941 and 0.7054, respectively, and the modeling effect is not ideal. The basic idea of the S-G algorithm is to obtain the purpose of smoothing by polynomials to perform polynomial least squares fitting to the data in the moving window. In this paper, a three-way polynomial movement window is used, but as can be seen from Figure 4c, there is still a lot of data noise in the 60 to 80 bands, which need to be further processed. Both SNV and MSC are designed to eliminate scattering effects caused by uneven particle distributions and different particle sizes, so they look similar [47] as shown in Figure 4d,e. After correction, there is still a large scattering in the 30th to 60th bands, and these two methods are not the best pretreatment methods. Combining Table 2, and Figure 4f, the pretreatment effect of OSC algorithm is the best, using the concentration matrix (Y) and the spectral matrix (X) orthogonal to filter out the signals in the spectrum that is not related to the concentration matrix [48].

Figure 5 shows the SVR modeling results of the dataset divided by the KS algorithm and the SPXY algorithm. Calibration set determination coefficient R_C^2 range is 0.63 to 0.91. Among them, the OSC-SPXY-SVR and OSC-KS-SVR models have the best effect, and the determination coefficients of the calibration set R_C^2 are 0.9138 and 0.9085, the RMSEC of the calibration set is 0.2396 and, 0.2468, the prediction set determination coefficients are 0.9145 and, 0.9233, and the RMSEP of the prediction set is 0.2280 and 0.3194, respectively. Models built by general KS algorithms are more likely to overfit. So, SPXY-SVR models are generally superior to KS-SVR models [34]. Secondly, the OSC-SPXY-SVR compares with the RMSEC and RMSEP of OSC-KS-SVR, respectively Obviously, 0.2396 < 0.2468; 0.2280 < 0.3194. This shows that the model established with OSC as the preprocessing algorithm and SPXY as the sample segmentation method not only has superior accuracy, but also has good robustness. In summary, OSC algorithm was finally selected to preprocess the original fluorescence hyperspectral data of mango.





Figure 5. SVR modeling results based on KS and SPXY algorithm datasets.

3.3. Extracting Effective Variables

After using the OSC algorithm preprocessing, the data noise has been improved to a certain extent, but there still a lot of information in the data that not related to the dry matter prediction. If the spectral number is not extracted further, the fluorescence hyper-spectral data will undoubtedly affect the accuracy and robustness of the model. Uninformative Variable Elimination (UVE) eliminates variables without information, and in the end, all that remains are useful variables for predicting chemical components. A total of 18 bands were selected, indicating that 14.4% of the functional information-free variables exist in the preprocessing spectrum. Removing the spectral bands that contribute less to the mango dry matter prediction model, the spectral dimension is changed from the original 120 dimensions to 21 dimensions, and it is not difficult to find that plenty of adjacent bands are retained through Figure 6a, but due to the large correlation of adjacent bands in hyperspectral data.



Figure 6. (a) The extraction process of UVE; (b) The extraction process of RF; (c) The extraction process of SPA; (d) preserve the wavelength number transformation trend; (e) RMSECV tendency shifts; (f) number of runs.

valid variables.

By setting the initial variable number to 5, the number of iterations to 1000, and the threshold of 0.15, the valid variables are extracted from the preprocessed spectral data using the Random Forest (RF). The extraction process of RF is given in Figure 6b. Figure 6b shows a plot of the probability distribution of usable variables. Distribution of valid variables within the total spectral interval is balanced. The threshold is configured on 0.15, and 10 variables with a probability greater than or equal to 0.15 are finally selected as

When using a Successive Projections Algorithm (SPA) to extract spectral variables, the RMSE of the SPA under different numbers of eigenspectral variables is calculated separately, and the smallest RMSE value is usually selected to determine the optimal number of eigenspectral variables. Set the number of eigenspectral variables to vary from 1 to 125, when RMSE = 0.2964 is the minimum value, at this time, 2 eigenspectral variables are extracted, accounting for 1.6% of the full spectral band, the distribution is shown in Figure 6c.

When extracting feature spectral variables, the Competitive Adaptive Reweighted Sampling (CARS) algorithm sets Monte Carlo to sample 50 times, using a 5-fold cross-validation method. Figure 6d shows the selection process of the exponential decay function, and the number of characteristic spectral variables decreases rapidly with the increase of the number of samples, and then decreases gently, with two characteristics: "coarse selection" and "selection" [49]. It can be seen from Figure 6e that as Monte Carlo samples increased, RMSECV showed a tendency to decrease slowly and then increase sharply, due to over selection and removal of information-rich key variables, resulting in a decrease in the predictive performance of the model. Figure 6f is the regression coefficient path diagram of the characteristic spectral variables with the sampling times, when the value of RMSECV in Figure 6e reaches the minimum value, the regression coefficient of each eigenspectral variable is located at the vertical line position where the '*' in Figure 6f is located, and the sampling runs 5 times, and finally 35 eigenspectral variables are extracted. The results of the bands selected by the 4 feature extraction methods are shown in Table 3.

Extraction Method	Number of Effective Variables
UVE	3, 20, 6, 27, 28, 29, 30, 87, 88, 89, 98, 99, 101, 103, 113, 114, 115, 116
SPA	108, 99
RF	28, 114, 01, 3, 102, 61, 91, 90, 17, 18
CARS	36, 7, 8, 9, 12, 13, 17, 18, 20, 21, 27, 28, 50, 69, 75, 77, 81, 82, 84, 85, 89, 90, 97, 98, 99, 102, 103, 114, 115, 117, 119, 121, 122, 124

Table 3. Results of two diversity methods and five preprocessing methods.

3.4. Combination of Characteristic Spectral Variables

According to the characteristic spectral variables extracted by the above four methods, it is not difficult to find that the number of feature band extractions of the four methods varies greatly, and there may be omissions or collinear problems, so this paper designs the combined extraction of 8 methods, as shown in Table 4, and establishes corresponding SVR, ELM and BPNN models to predict the dry matter of mango.

Extraction Method	Number of Effective Variables	Valid Variables Percentage
UVE	18	14.4%
SPA	2	1.6%
RF	10	8%
CARS	35	28%
UVE-SPA	2	1.6%
CARS-SPA	4	3.2%
CARS + SPA	37	29.6%
(CARS + SPA)-SPA	2	1.6%
UVE + RF	28	22.4%
(UVE + RF)-SPA	2	1.6%
CARS + RF	45	36%
(CARS + RF)-SPA	4	3.2%

Table 4. The number of feature variables extracted by different feature extraction methods.

Note: For example, CARS + RF represents a direct combination of the 35 feature spectral variables extracted by the CARS algorithm and the 10 feature spectral variables extracted by the RF algorithm; (CARS + RF)-SPA represents the use of the SPA algorithm to perform the secondary dimensional reduction on the combined feature spectral variable formed by CARS + RF.

3.5. Building the Models and Analyzing the Results

3.5.1. SVR Model

Support Vector Machine (SVM) can be split into Support Vector Classification (SVC) and Support Vector Regression (SVR) according to the implementation goal. According to whether the relaxation variable is introduced, it can be divided into hard interval support vector machine; According to whether the data is linearly separable, it can be divided into linear support vector machine and kernel method support vector machine. Due to the introduction of subtle spacing, kernel methods and other ideas, SVM is particularly suitable for solving the problems of small sample, nonlinearity, and high-dimensional pattern recognition.

Support vector regression (SVR) is essentially no different from support vector classification. Support vector classification is to solve the maximum separation geometric interval, while support vector regression is to solve the maximum tolerable deviation ϵ between the predicted value and the true value. The derivation process of SVR is as follows [50]:

$$\min_{w,b} \frac{\ell}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \ell_i (f(\mathbf{x}_i) - y_i)$$
(8)

where, *C* is the regularization parameter, ℓ_i is the insensitive loss function of ϵ , the specific expression is shown in (9).

$$\ell_{\epsilon}(z) = \begin{cases} 0, & \text{if } |z| \leq \epsilon \\ |z| - \epsilon, & \text{otherwise} \end{cases}$$
(9)

Introducing slack variables *b* and *v*, Equation (8) can be rewritten as:

$$\min_{\substack{w,b,\xi_i,\hat{\xi}_i}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i)$$

s.t. $f(x_i) - y_i \leq \epsilon + \xi_i$
 $y_i - f(\mathbf{x}_i) \leq \epsilon + \hat{\xi}_i$
 $\xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, \dots, m$ (10)

The Lagrange multiplier method is used to solve the problem, and the Lagrange multiplier $\mu_i \ge 0$, $\hat{\mu}_i \ge 0$, $\hat{\alpha}_i \ge 0$, $\hat{\alpha}_i \ge 0$ is introduced to obtain the Lagrange function of (10).

$$L(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\alpha}, \boldsymbol{\zeta}, \boldsymbol{\zeta}, \boldsymbol{\mu}, \boldsymbol{\mu}) = \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{m} (\tilde{\zeta}_i + \hat{\zeta}_i) - \sum_{i=1}^{m} \mu_i \tilde{\zeta}_i - \sum_{i=1}^{m} \hat{\mu}_i \hat{\zeta}_i + \sum_{i=1}^{m} \alpha_i (f(\boldsymbol{x}_i) - \boldsymbol{y}_i - \boldsymbol{\epsilon} - \tilde{\zeta}_i) + \sum_{i=1}^{m} \hat{\alpha}_i (\boldsymbol{y}_i - f(\boldsymbol{x}_i) - \boldsymbol{\epsilon} - \hat{\zeta}_i)$$
(11)

Taking the partial derivatives of ω , *b*, ξ_i and $\hat{\xi}_i$ from Equation (11), we get:

$$w = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i$$

$$0 = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i)$$

$$C = \alpha_i + \mu_i$$

$$C = \hat{\alpha}_i + \hat{\mu}_i$$

(12)

Substitute (12) into (10) to get the Dual problem of SVR

$$\max_{\substack{\alpha, \hat{\alpha} \\ \alpha, \hat{\alpha}}} \sum_{i=1}^{m} y_i(\hat{\alpha}_i - \alpha_i) - \epsilon(\hat{\alpha}_i + \alpha_i) \\ -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$$
(13)
s.t.
$$\sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) = 0 \\ 0 \leqslant \alpha_i, \hat{\alpha}_i \leqslant C$$

According to the KKT condition, it can be solved:

$$b = y_i + \epsilon - \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{x}$$
(14)

$$\boldsymbol{w} = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) \boldsymbol{\phi}(\boldsymbol{x}_i)$$
(15)

Substitute (14) and (15) into the hyperplane model $f(\mathbf{x}) = \mathbf{w}^{T} \phi(\mathbf{x}) + b$ of the SVR kernel method, and the SVR can be expressed as:

$$f(\mathbf{x}) = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) \kappa(\mathbf{x}, \mathbf{x}_i) + b$$
(16)

where, $\kappa(\mathbf{x}, \mathbf{x}_i)$ is a kernel function, \mathbf{x}_i is the input variable, $\hat{\alpha}_i$, α for the Lagrange multiplier, b is the deviation of the SVR model.

The performance of the SVR model is mainly affected by the regularization parameter *C* and the kernel function parameter γ . The smaller the value of *C*, the lower the complexity of the model and the stronger the generalization ability of the model; otherwise, the higher the fitting accuracy of the model, the lower the generalization ability; while γ reflects the influence of a single sample on the model, the larger γ is, the smaller the influence of a single sample on the model, and vice versa [51]. Therefore, this paper uses the grid search algorithm to select the optimal parameters for the two parameters to obtain the optimal SVR model.

It can be seen from Table 5 that compared with the original full-spectrum data is shown in Table 4, the SVR model after feature extraction by 4 methods is not effective, but the SVR model established by the primary dimensionality reduction UVE-SPA and CARS-SPA, R_C^2 , R_P^2 has been significantly improved, indicating that SPA can effectively eliminate the correlation between the bands, and then use the primary combined dimen-

sionality reduction and the secondary combinatorial dimensionality reduction algorithm, (CARS + RF)-SPA-SVR prediction effect is best as shown in Figure 7a,b. R_C^2 and R_P^2 were 0.9038, and 0.9420, respectively. In addition, it is not difficult to find from Table 5 that the difference between R_P^2 and R_C^2 of the UVE-SPA-SVR combination model is too large, and overfitting occurs. The RMSEC and RMSEP values of the (UVE + RF)-SPA-SVR model and the UVE-SPA model are respectively equal, indicating the existence of collinearity.



Figure 7. (a) Calibration set of (CARS + RF)-SPA-SVR model; (b) Prediction set of (CARS + RF)-SPA-SVR model; (c) (CARS + RF)-SPA-ELM model; (d) (CARS + RF)-SPA-BPNN model.

Extraction Method	Variable Number	R _C ²	RMSEC	R _P ²	RESEP
UVE	18	0.8839	0.2746	0.9075	0.2377
SPA	2	0.9224	0.2269	0.8214	0.3413
RF	10	0.9007	0.2563	0.8863	0.2648
CARS	35	0.8817	0.2816	0.9105	0.2357
UVE-SPA	2	0.8932	0.2668	0.9322	0.2550
CARS-SPA	4	0.9030	0.2514	0.9220	0.1764
CARS + SPA	37	0.8805	0.2832	0.9187	0.2216
(CARS + SPA)-SPA	2	0.9159	0.2587	0.8364	0.3339
UVE + RF	28	0.9092	0.2462	0.9100	0.2330
(UVE + RF)-SPA	2	0.8932	0.2668	0.9322	0.2550
CARS + RF	45	0.8839	0.2784	0.9075	0.2377
(CARS + RF)-SPA	4	0.9038	0.2524	0.9420	0.1766

Table 5. Prediction results of SVR model established by different feature extraction methods.

3.5.2. ELM Model

Extreme learning machine (ELM) is a single hidden layer feedforward neural network. Unlike BPNN, this algorithm requires continuous iterative connection weights and thresholds, and only adjusts the number of hidden layer neurons to parse the output weights [52]. So, it converges very fast and generalizes well. The specific derivation is as follows [53]:

Let the training set be {*X*, *T*}, where *X* is the $N \times m$ samples, and *T* is the $N \times 1$ expected output matrix; the connection weight between neurons is written as *w*, and the neuron activation threshold is written as *b*, the output of the kth neuron in the hidden layer as $h_k(x)$, and the calculation formula of $h_k(x)$ is shown in (17).

$$h_k(x) = g(x; w_k, b_k) \tag{17}$$

where, $g(\bullet)$ is the transfer (activation) function, generally, select the sigmoid function, tanh function, Relu function as the activation function.

The output of ELM is denoted as $f(x) = h(x)\beta$, where h(x) is the output vector of the hidden layer and, β is the output weight, and β is solved by the least square method, as shown in formula (18).

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2$$

$$s.t.h(x_i)\beta = t_i^T - \xi_i^T, i = 1, 2, \cdots, N$$

$$(18)$$

where, $\frac{1}{2} \|\beta\|^2$ is the regular term, $\frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2$ is the sum of the prediction errors, *C* is the coefficient, converting this problem into an unconstrained problem is:

$$\beta^{\hat{A}} - CH^T \left(T - H\beta^{\hat{A}} \right) = 0 \tag{19}$$

When the number of training samples is less than the number of neurons in the hidden layer:

$$\beta^{\hat{A}} = H^T \left(H H^T + \frac{I_N}{C} \right)^{-1} T \tag{20}$$

When the number of training samples is greater than the number of neurons in the hidden layer:

$$\beta^{\hat{t}} = \left(H^T H + \frac{1_L}{c}\right)^{-1} H^T T \tag{21}$$

As can be seen from Table 6, Longitudinal comparison of 12 feature extraction methods, the best predictive method is (CARS + RF)-SPA-ELM, as shown in Figure 7c, R_C^2 and R_P^2 are 0.8740 and, 0.9336, respectively. The worst model is RF-ELM, R_P^2 is only 0.8328, RMSEP

is 0.3796. Compare the results in the table, UVE-SPA, UVE, CARS-SPA, (CARS + SPA)-SPA, and (UVE + RF)-SPA Both are severely overfitted, and the feature combination extraction has no significant effect in the ELM model. Therefore, a horizontal comparison of the 3 models predicts the effect, the RMSEC of 12 features extracted exceeded 0.3, and the range of RMESEP was 0.2591 to 0.3796, which was no less than 0.2, at the same time, the difference between R_P^2 and R_C^2 is too large, and there is serious overfitting, so the prediction performance of ELM was poor.

Extraction Method	Variable Number	R _C ²	RMSEC	R _P ²	RESEP
UVE	18	0.8363	0.3863	0.9239	0.2888
SPA	2	0.7981	0.4068	0.9064	0.2932
RF	10	0.8350	0.3678	0.8328	0.3796
CARS	35	0.8341	0.3694	0.8459	0.3601
UVE-SPA	2	0.7467	0.4595	0.8424	0.3470
CARS-SPA	4	0.9118	0.2946	0.8634	0.3569
CARS + SPA	37	0.9052	0.2998	0.8932	0.3013
(CARS + SPA)-SPA	2	0.8010	0.4038	0.9087	0.2772
UVE + RF	28	0.8824	0.3105	0.9273	0.2889
(UVE + RF)-SPA	2	0.8372	0.3853	0.8674	0.3546
CARS + RF	45	0.8393	0.3856	0.8885	0.3129
(CARS + RF)-SPA	4	0.8740	0.3214	0.9336	0.2591

Table 6. Prediction results of ELM model established by different feature extraction methods.

3.5.3. BPNN Model

A neural network is an extensive parallel interconnected network of adaptive simple units whose organization can simulate the interaction of biological nervous systems with real-world objects [54]. Error Back Propagation Neural Network (BPNN) is one of the most classic types of neural network models. BPNN starts the forward propagation of the signal at random initial weights and thresholds, gives a predicted output in the output layer and compares it with the expected output, and then propagates the deviation back to the neurons in each layer to guide the connection weights and the bias update so iteratively, is that the predicted value is constantly approaching the expected value.

The BPNN training process is now described as follows [54]:

- 1. Randomly set an initial link weight ω , activation threshold θ of each neuron, and learning rate η in the interval (0,1).
- 2. Determine the number of neurons in the input layer according to the input variable, assuming that there is a neuron in the hidden layer and a neuron in the output layer, denoted as. The input layer neuron directly inputs the input signal to the hidden layer without any transformation, and the input and output of the *j*th neuron in the hidden layer are respectively expressed as formulas (20) and (23).

$$\alpha_j = \sum_{i=1}^d w_{ij} x_i \tag{22}$$

$$\hat{\varphi}_j = f\left(\alpha_j - \theta_j\right) \tag{23}$$

where, θ_j is the hidden layer threshold, and $f(\bullet)$ is the neuron transfer (activation) function. The input and output of the output layer can be expressed as:

$$\beta = \sum_{j=1}^{m} w_j b_j \tag{24}$$

$$\hat{y} = f(\beta - y) \tag{25}$$

- 3. The square of the difference between the predicted value and the actual value is used as the prediction error, that is, $E = \frac{1}{2}(\hat{y} y)^2$, where the function of $\frac{1}{2}$ is to facilitate the calculation.
- 4. According to the gradient descent algorithm, the connection weight of the output layer and the hidden layer and the gradient of the threshold are calculated. The calculation formula is as follows:

$$w_j := w_j - \nabla w_j = w_j - \eta \frac{\partial E}{\partial w_j}$$
(26)

$$\theta_j := \theta_j - \nabla \theta_j = \theta_j - \eta \frac{\partial E}{\partial \theta_j}$$
(27)

5. After updating the weights and thresholds, repeat steps 2–4 until the iteration termination condition is reached.

The activation function selected in this paper is the sigmoid function, the learning rate is selected as 0.01, and the number of neurons in the hidden layer is 12. The BPNN implementation is implemented by calling the neural network toolbox in MATLAB2020b.

As obtained from Table 7, (CARS + RF)-BPNN and (CARS + RF)-SPA-BPNN are the most effective models among the 12 extraction methods, R_C^2 are 0.9400, 0.9710; R_P^2 are 0.9527, 0.9658, respectively. It shows that the effect of the secondary combination dimensionality reduction algorithm (CARS + RF)-SPA-BPNN of feature extraction is slightly higher than that of (CARS + RF)-BPNN of R_P^2 , but the former requires 45 characteristic spectral variables, resulting in a large computational load, and the spectral variables extraction method of CARS + RF has certain randomness, resulting in the detection results of the preprocessing method are not stable. The secondary combinatorial dimensionality reduction method of (CARS + RF)-SPA can reduce the number of characteristic spectral variables and eliminate collinearity between variables, and further reduce the randomness caused by the CARS algorithm, so that the corresponding BPNN model has high precision and a stable and reliable prediction performance. From Tables 5 and 6, the prediction accuracy of (CARS + RF)-SPA under both SVR and ELM models is very high, indicating that the applicability of the feature extraction method is strong. There is a certain difference between the model evaluation R_{C}^{2} and R_{P}^{2} of the various methods from Tables 5–7, which is caused by the small number of samples and small data of the correction set and the prediction set, and the model prediction results can be improved by expanding the sample data volume in practical applications. Combining the premature results from Tables 5-7, it can be determined that (CARS + RF)-SPA-BPNN is the best prediction method for the dry matter of mango, as shown in Figure 7d.

Extraction Method	Variable Number	R _C ²	RMSEC	R _P ²	RESEP
UVE	18	0.9114	0.2496	0.8768	0.3343
SPA	2	0.8659	0.3673	0.9481	0.2559
RF	10	0.7743	0.4512	0.8533	0.3708
CARS	35	0.9143	0.2400	0.9383	0.2313
UVE-SPA	2	0.8940	0.3006	0.8019	0.3979
CARS-SPA	4	0.9321	0.2347	0.8930	0.3282
CARS + SPA	37	0.9514	0.2097	0.8940	0.3006
(CARS + SPA)-SPA	2	0.8514	0.3846	0.9377	0.2195
UVE + RF	28	0.9305	0.2236	0.8525	0.3731
(UVE + RF)-SPA	2	0.9306	0.2205	0.8861	0.3267
CARS + RF	45	0.9400	0.2106	0.9527	0.2013
(CARS + RF)-SPA	4	0.9710	0.1418	0.9658	0.1526

Table 7. Prediction results of BPNN model established by different feature extraction methods.

4. Conclusions

The results of this study show that fluorescence hyperspectral knot technology and machine learning can be used for nondestructive testing of mango dry matter.

- 1. Comparing the prediction results of the SVR model established by DT, SG, SNV, MSC, and OSC algorithm. OSC algorithm pretreatment has the best effect, which can effectively reduce the influence of spectral baseline drift and tilt, while retaining spectral information to the greatest extent.
- 2. Using the idea of complementary advantages and disadvantages between algorithms, a total of 12 feature extraction methods are used for the spectrum after OSC algorithm, including the primary dimensionality reduction, the primary combined dimensionality reduction, and the secondary combination dimensionality reduction, which makes up for the shortcomings of insufficient stability when the feature extraction algorithm is used alone.
- 3. Based on the above 12 feature extraction methods. SVR, ELM and BPNN models for predicting the dry matter of mango were established. Experimental results show that the BPNN model has the best prediction performance, while the ELM model has the worst prediction performance, (CARS + IRIV)-SPA algorithm extracts feature spectral variables into SVR and ELM models, which obtain better predictions than other methods. comparing the prediction accuracy and stability of different methods, OSC-SPXY-(CARS + RF)-SPA-BPNN is finally determined to be the optimal prediction method for detecting the dry matter of mango, and its correlation coefficients $R_C^2 = 0.9710$, $R_P^2 = 0.9658$, RMSEC = 0.1418, RMSEP = 0. 1526.

This study proves the feasibility of fluorescence hyperspectral technology combined with machine learning in nondestructive detection of mango dry matter. although fluorescent hyperspectral technology can achieve high-precision nondestructive detection of mango dry matter, but there are still some errors with the actual physical and chemical values, and the prediction accuracy needs to be further improved. In addition, fluorescence hyperspectral instruments are relatively expensive and bulky, and they cannot achieve rapid, online detection, which hinders the promotion and use of this technology. In future research, we will focus on improving the accuracy of mango quality prediction through the combination of graph fusion technology and feature fusion method, and our team is also working on improving the miniaturization of equipment. Therefore, our team will improve on the following areas: 1. Establish a fluorescence hyperspectral nondestructive testing database so that users can download it online in real time; 2. R&D and production of portable fluorescence hyperspectral instruments; 3. Establishment of mango atlas fusion technology segment and content visualization through gray level co-occurrence matrix.

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

Funding: This work was supported by the subject double support program of Sichuan Agricultural University (Grant NO. 035-1921993093).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://github.com/GengJinping/GengJinping/tree/main (accessed on 14 July 2022).

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

References

- Anderson, N.T.; Subedi, P.P.; Walsh, K.B. Manipulation of mango fruit dry matter content to improve eating quality. *Sci. Hortic.* 2017, 226, 316–321. [CrossRef]
- Saranwong, S.; Sornsrivichai, J.; Kawano, S. Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. *Postharvest Biol. Technol.* 2004, 31, 137–145. [CrossRef]
- 3. Subedi, P.P.; Walsh, K.B.; Owens, G. Prediction of mango eating quality at harvest using short-wave near infrared spectrometry. *Postharvest Biol. Technol.* 2007, 43, 326–334. [CrossRef]
- 4. Sun, X.; Subedi, P.; Walsh, K.B. Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content. *Postharvest Biol. Technol.* 2020, *162*, 111177. [CrossRef]
- Cui, M.; Sun, Y.; Huang, C.; Li, M. Water Turbidity Retrieval Based on UAV Hyperspectral Remote Sensing. *Water* 2022, 14, 128. [CrossRef]
- Li, L.; Huang, J.; Wang, Y.; Jin, S.; Li, M.; Sun, Y.; Ning, J.; Chen, Q.; Zhang, Z. Intelligent evaluation of storage period of green tea based on VNIR hyperspectral imaging combined with chemometric analysis. *Infrared Phys. Technol.* 2020, 110, 103450. [CrossRef]
- ElMasry, G.; Sun, D.-W.; Allen, P. Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. J. Food Eng. 2013, 117, 235–246. [CrossRef]
- 8. Luo, X.; Xu, L.J.; Huang, P.; Wang, Y.C.; Liu, J.; Hu, Y.; Wang, P.; Kang, Z.L. Nondestructive Testing Model of Tea Polyphenols Based on Hyperspectral Technology Combined with Chemometric Methods. *Agriculture* **2021**, *11*, 673. [CrossRef]
- 9. Pu, Y.Y.; Sun, D.W. Vis-NIR hyperspectral imaging in visualizing moisture distribution of mango slices during microwave-vacuum drying. *Food Chem.* **2015**, *188*, 271–278. [CrossRef]
- Makino, Y.; Isami, A.; Suhara, T.; Oshita, S.; Tsukada, M.; Ishiyama, R.; Serizawa, M.; Kuroki, S.; Kawagoe, Y.; Purwanto, Y.A.; et al. Non-destructive analysis of internal and external qualities of mango fruits during storage by Hyperspectral imaging. *Acta Hortic.* 2013, 1011, 443–450. [CrossRef]
- 11. Sharma, S.; Sumesh, K.C.; Sirisomboon, P. Rapid ripening stage classification and dry matter prediction of durian pulp using a pushbroom near infrared hyperspectral imaging system. *Measurement* **2022**, *189*, 110464. [CrossRef]
- Hu, Y.; Kang, Z. The Rapid Non-Destructive Detection of Adulteration and Its Degree of Tieguanyin by Fluorescence Hyperspectral Technology. *Molecules* 2022, 27, 1196. [CrossRef]
- 13. Yu, Y.; Qu, Y.; Zhang, M.; Guo, X.; Zhang, H. Fluorescence detection of paclobutrazol pesticide residues in apple juice. *Optik* **2020**, 224, 165542. [CrossRef]
- 14. Liu, Y.; Zou, J.; Luo, B.; Yu, H.; Zhao, Z.; Xia, H. Ivy extract-assisted photochemical vapor generation for sensitive determination of mercury by atomic fluorescence spectrometry. *Microchem. J.* **2021**, *169*, 106547. [CrossRef]
- Guo, X.-J.; He, X.-S.; Li, C.-W.; Li, N.-X. The binding properties of copper and lead onto compost-derived DOM using Fouriertransform infrared, UV-vis and fluorescence spectra combined with two-dimensional correlation analysis. *J. Hazard. Mater.* 2019, 365, 457–466. [CrossRef]
- 16. Hu, Y.; Zhang, H.; Zhao, D. Transform method in three-dimensional fluorescence spectra for direct reflection of internal molecular properties in rapid water contaminant analysis. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, 250, 119376. [CrossRef]
- 17. Li, G.; Liao, Y.; Wang, X.; Sheng, S.; Yin, D. In situ estimation of the entire color and spectra of age pigment-like materials: Application of a front-surface 3D-fluorescence technique. *Exp. Gerontol.* **2006**, *41*, 328–336. [CrossRef]
- 18. Sunuwar, S.; Manzanares, C.E. Excitation, emission, and synchronous fluorescence for astrochemical applications: Experiments and computer simulations of synchronous spectra of polycyclic aromatic hydrocarbons and their mixtures. *Icarus* 2021, 370, 114689. [CrossRef]
- 19. Hu, Y.; Xu, L.; Huang, P.; Luo, X.; Wang, P.; Kang, Z. Reliable Identification of Oolong Tea Species: Nondestructive Testing Classification Based on Fluorescence Hyperspectral Technology and Machine Learning. *Agriculture* **2021**, *11*, 1106. [CrossRef]
- Wang, X.; Xu, L.; Chen, H.; Zou, Z.; Huang, P.; Xin, B. Non-Destructive Detection of pH Value of Kiwifruit Based on Hyperspectral Fluorescence Imaging Technology. *Agriculture* 2022, 12, 208. [CrossRef]
- 21. Zhuang, Q.; Peng, Y.; Yang, D.; Wang, Y.; Zhao, R.; Chao, K.; Guo, Q. Detection of frozen pork freshness by fluorescence hyperspectral image. *J. Food Eng.* **2022**, *316*, 110840. [CrossRef]
- 22. Xu, L.J.; Zheng, L.N.; Huang, P.; Chen, H.; Kang, Z.L. Detection of kiwifruit dry matter content based on hyperspectral technology using uninformed variable elimination coupled with successive projection algorithm. *Dyna* **2020**, *95*, 654–660. [CrossRef]
- Huang, L.; Liu, Y.; Huang, W.; Dong, Y.; Ma, H.; Wu, K.; Guo, A. Combining Random Forest and XGBoost Methods in Detecting Early and Mid-Term Winter Wheat Stripe Rust Using Canopy Level Hyperspectral Measurements. *Agriculture* 2022, 12, 74. [CrossRef]
- Xing, Z.; Du, C.; Shen, Y.; Ma, F.; Zhou, J. A method combining FTIR-ATR and Raman spectroscopy to determine soil organic matter: Improvement of prediction accuracy using competitive adaptive reweighted sampling (CARS). *Comput. Electron. Agric.* 2021, 191, 106549. [CrossRef]
- Yao, K.S.; Sun, J.; Zhang, L.; Zhou, X.; Tian, Y.; Tang, N.Q.; Wu, X.H. Nondestructive detection for egg freshness based on hyperspectral imaging technology combined with harris hawks optimization support vector regression. *J. Food Saf.* 2021, 41, e12888. [CrossRef]
- Ma, Q.; Teng, Y.; Li, C.; Jiang, L. Simultaneous quantitative determination of low-concentration ternary pesticide mixtures in wheat flour based on terahertz spectroscopy and BPNN. *Food Chem.* 2022, 377, 132030. [CrossRef]

- 27. Hong, Z.; Zhang, C.; Kong, D.; Qi, Z.; He, Y. Identification of storage years of black tea using near-infrared hyperspectral imaging with deep learning methods. *Infrared Phys. Technol.* **2021**, *114*, 103666. [CrossRef]
- Wang, Y.-J.; Jin, G.; Li, L.-Q.; Liu, Y.; Kianpoor Kalkhajeh, Y.; Ning, J.-M.; Zhang, Z.-Z. NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves. *Infrared Phys. Technol.* 2020, 108, 103365. [CrossRef]
- Delwiche, S.R.; Baek, I.; Kim, M.S. Does spatial region of interest (ROI) matter in multispectral and hyperspectral imaging of segmented wheat kernels? *Biosyst. Eng.* 2021, 212, 106–114. [CrossRef]
- Ni, Z.; Lü, X.; Huang, G. Impact of Meteorological Factors on Thermokarst Lake Changes in the Beilu River Basin, Qinghai-Tibet Plateau, China (2000–2016). Water 2021, 13, 1605. [CrossRef]
- Nawar, S.; Mouazen, A.M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. Comput. Electron. Agric. 2018, 151, 469–477. [CrossRef]
- Galvão, R.K.H.; Araujo, M.C.U.; José, G.E.; Pontes, M.J.C.; Silva, E.C.; Saldanha, T.C.B. A method for calibration and validation subset partitioning. *Talanta* 2005, 67, 736–740. [CrossRef]
- Pan, R.R.; Luo, Y.F.; Wang, C.; Zhang, C.; He, Y.; Feng, L. Classifications of Oilseed Rape and Weeds Based on Hyperspectral Imaging. Spectrosc. Spectr. Anal. 2017, 37, 3567–3572. [CrossRef]
- Du, S.X.; Du, Y.F.; Wu, X.L. The Surface Smoothing Methods for Three-Dimensional Fluorescence Spectrometry Based on Savitzky-Golay Polynomial Smoothing. Spectrosc. Spectr. Anal. 2011, 31, 440–443. [CrossRef]
- Gessell, A.; Small, G.W. Longitudinal Study Comparing Orthogonal Signal Correction Algorithms Coupled with Partial Least-Squares for Quantitative Near-Infrared Spectroscopy. *Anal. Lett.* 2022, 55, 449–466. [CrossRef]
- 36. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]
- He, X.; Jiang, X.; Fu, X.; Gao, Y.; Rao, X. Least squares support vector machine regression combined with Monte Carlo simulation based on the spatial frequency domain imaging for the detection of optical properties of pear. *Postharvest Biol. Technol.* 2018, 145, 1–9. [CrossRef]
- Ullah, R.; Khan, S.; Bilal, M.; Nurjis, F.; Saleem, M. Non-invasive assessment of mango ripening using fluorescence spectroscopy. Optik 2016, 127, 5186–5189. [CrossRef]
- Gitelson, A.A.; Buschmann, C.; Lichtenthaler, H.K. Leaf chlorophyll fluorescence corrected for re-absorption by means of absorption and reflectance measurements. J. Plant Physiol. 1998, 152, 283–296. [CrossRef]
- 40. Gitelson, A.A.; Buschmann, C.; Lichtenthaler, H.K. The Chlorophyll Fluorescence Ratio F735/F700 as an Accurate Measure of the Chlorophyll Content in Plants. *Remote Sens. Environ.* **1999**, *69*, 296–302. [CrossRef]
- Bilal, M.; Ullah, R.; Khan, S.; Ali, H.; Saleem, M.; Ahmed, M. Lactate based optical screening of dengue virus infection in human sera using Raman spectroscopy. *Biomed. Opt. Express* 2017, *8*, 1250–1256. [CrossRef] [PubMed]
- Abbasi, N.A.; Iqbal, Z.; Maqbool, M.; Hafiz, I.A. Postharvest Quality of mango (*Mangifera indica* L.) fruit as affected by chitosan coating. *Pak. J. Bot.* 2009, *41*, 343–357.
- Ben-Amotz, A.; Fishier, R. Analysis of carotenoids with emphasis on 9-cis β-carotene in vegetables and fruits commonly consumed in Israel. *Food Chem.* 1998, 62, 515–520. [CrossRef]
- 44. Cano, M.P.; de Ancos, B. Carotenoid and Carotenoid Ester Composition in Mango Fruit As Influenced by Processing Method. *J. Agric. Food Chem.* **1994**, 42, 2737–2742. [CrossRef]
- 45. Léchaudel, M.; Joas, J. An overview of preharvest factors influencing mango fruit growth, quality and postharvest behaviour. *Braz. J. Plant Physiol.* **2007**, *19*, 287–298. [CrossRef]
- Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Correction to the Description of Standard Normal Variate (SNV) and De-Trend (DT) Transformations in Practical Spectroscopy with Applications in Food and Beverage Analysis—2nd Edition. *J. Near Infrared* Spectrosc. 1993, 1, 185–186. [CrossRef]
- Fearn, T.; Riccioli, C.; Garrido-Varo, A.; Guerrero-Ginel, J.E. On the geometry of SNV and MSC. *Chemom. Intell. Lab. Syst.* 2009, 96, 22–26. [CrossRef]
- Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* 1998, 44, 175–185. [CrossRef]
- Song, X.Z.; Du, G.R.; Li, Q.Q.; Tang, G.; Huang, Y. Rapid spectral analysis of agro-products using an optimal strategy: Dynamic backward interval PLS-competitive adaptive reweighted sampling. *Anal. Bioanal. Chem.* 2020, 412, 2795–2804. [CrossRef]
- 50. Vapnik, V.N. An overview of statistical learning theory. IEEE Trans. Neural Netw. 1999, 10, 988–999. [CrossRef]
- Jayanthi, S.L.; Keesara, V.R.; Sridhar, V. Prediction of Future Lake Water Availability Using SWAT and Support Vector Regression (SVR). Sustainability 2022, 14, 6974. [CrossRef]
- 52. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
- Zhu, S.; Feng, L.; Zhang, C.; Bao, Y.; He, Y. Identifying Freshness of Spinach Leaves Stored at Different Temperatures Using Hyperspectral Imaging. *Foods* 2019, *8*, 356. [CrossRef]
- Liu, P.; Liu, Z.; Hu, Y.; Shi, Z.; Pan, Y.; Wang, L.; Wang, G. Integrating a Hybrid Back Propagation Neural Network and Particle Swarm Optimization for Estimating Soil Heavy Metal Contents Using Hyperspectral Data. *Sustainability* 2019, 11, 419. [CrossRef]