

Research Article

Study on the Effect of Apple Size Difference on Soluble Solids Content Model Based on Near-Infrared (NIR) Spectroscopy

Xiaogang Jiang,^{1,2} Mingwang Zhu,^{1,2} Jinliang Yao,^{1,2} Yuxiang Zhang,³ and Yande Liu ^{1,2}

¹School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China

²Institute of Intelligent Mechanical and Electrical Equipment Innovation, Nanchang 330013, China

³Huaiyin Institute of Technology, Huaiyin 223003, China

Correspondence should be addressed to Yande Liu; jxliuyd@163.com

Received 23 April 2022; Accepted 13 May 2022; Published 27 May 2022

Academic Editor: Jose S. Camara

Copyright © 2022 Xiaogang Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soluble solids content (SSC) is a vital evaluation index for the internal quality of apples, and NIR spectroscopy is the preferred technique for predicting the SSC of apples. Due to the differences in fruits' sizes, their SSC prediction models have poor robustness and low prediction accuracy, so it is important to eliminate the effects brought by the differences in fruit sizes to improve the accuracy of fruit sorting models. The NIR spectra of apples with different fruit sizes were collected by applying NIR spectroscopy online detection device, and after various preprocessing of the spectra, the partial least squares (PLS) models of apple SSC were established, respectively, and then the modeling set in the apple fruit size group of 75 mm–85 mm was used to predict the prediction set samples in the apple fruit size group of 65 mm–75 mm and 85 mm–95 mm, respectively. To better address the effects of apple size differences, data fusion techniques were used to perform an intermediate fusion of apple fruit diameter and spectra, firstly, the competitive adaptive reweighting algorithm (CARS) and the continuous projection algorithm (SPA) were used to select spectral variables and build their prediction models for apple SSC, respectively, and the results showed that the models built with 61 spectral variables selected by CARS had better performance, greatly reduced the amount of data involved in modeling, effectively simplified the model, and improved the stability of the model. The apple size variables were added to the wavelength variables selected by CARS, and the data were normalized to establish a PLS model of apple SSC with the normalized spectral and apple fruit diameter data, and the results showed that the size compensation model based on intermediate fusion had the best prediction performance, with the prediction set R_p of 0.886 for fruit diameter of 65 mm–75 mm, RMSEP of 0.536%, and its prediction set R_p was 0.913 and RMSEP was 0.497% for the fruit diameter of 85 mm–95 mm. Therefore, adding the fruit diameter variable to establish the size-compensated model of apple SSC can improve the prediction performance of the model.

1. Introduction

Apples are rich in many vitamins and acids inside, and eating more apples can relieve fatigue and improve brain vitality. Due to the increasing standard of living, the demand for high-quality apples is also increasing. Near-infrared spectroscopy online inspection technology has been applied to detect the internal quality of fruits such as apples, strawberries, citrus, pears, and watermelons as a fast, nondestructive, and green inspection technology [1–4]. The difference in fruit diameter of apples affects the performance of the established SSC model. Therefore, selecting a size-compensated model of apple SSC is necessary.

Scholars at home and abroad have done a lot of research on the internal quality of fruits by applying NIR spectroscopy. Guo et al. [5] built an online detection system for apple heart rot using NIR transmission, and the correlation coefficient of the prediction model they built was 0.92. Liu et al. [6] created a NIR diffuse reflection online detection model for SSC of navel orange, and its prediction correlation coefficient was 0.90. Li et al. [7] built an apple online non-destructive testing equipment using NIR spectroscopy and established a prediction model for the SSC content of apple, whose correlation coefficient reached 0.949 and the root mean square error of the prediction set was 0.449. Han et al. [8, 9] used NIR transmission spectroscopy combined with a

band screening method to discriminate two diseases of apple, and their discriminant model accuracy reached 95.7%. Xu et al. [10] studied compared the effect of single- and double-point detection on the accuracy of online detection of apple SSC. The double-branched fiber system proved excellent robustness, while the single-branched fiber proved perfect accuracy with a prediction set coefficient of 0.63. The studies conducted by the above et al. did not consider the effect of apple fruit diameter on the model, and the performance of the established model was low [7, 8, 10]. Liu et al. [11] established the NIR spectrum detection model for navel oranges of different sizes and found through the study that the use of MSC and SNV pretreatment can solve the influence of apple size differences and improve the accuracy of the prediction model. Tian et al. [12] established a discriminative model for apple core mold of different sizes in NIR spectra. They found that the NIR spectrum intensity and optical range were exponentially related and modified the NIR spectra and the modified model. The prediction set discrimination accuracy reached 90.2%, and the method can correct the effect of fruit size on the transmission spectrum to improve the identification of diseased apples. Two prototypes of online NIR systems were developed by McGlone et al. [13], one is based on a time-delay integral spectrometer and the other on a large-aperture spectrometer. The latter system has high accuracy, with a predicted root mean square error of 4.1% after PLS correction. However, only apples with a mean equatorial diameter of 76 mm (SD = 2.8 mm) were selected for the experiment, and the effect of fruit size on the detection of browning tissue was not investigated. Qin and Lu [14, 15] quantified the light transmission in apples using Monte Carlo simulations. They corrected the diffuse reflectance spectrum according to the fruit size to eliminate the light intensity distortion caused by the curved fruit surface. In this paper, we applied NIR spectra. We collected apple NIR spectra at different sizes to establish various preprocessing models, mixed size models, and data fusion-based apple size compensation models to compare the advantages and disadvantages of the prediction performance of the three models and find the best solution for the effect of apple size differences on the model performance.

2. Materials and Methods

2.1. Experimental Materials. A total of 480 apple samples, including 160 of 65 mm–75 mm fruit diameter, 160 of 75 mm–85 mm fruit diameter, and 160 of 85 mm–95 mm fruit diameter, were ordered from an orchard in Yantai, Shandong Province. Upon arrival, the experimental spectra of the apples were collected after wiping off the dust from the surface of the apples with a wet paper towel in order to prevent the scattering effect of the dust on the transmission spectra and leaving the apples in a room with an ambient temperature of 25°C for 24 hours.

2.2. Experimental Device and Spectrum Acquisition. The near-infrared spectrum acquisition device used in this paper is a dynamic online diffuse transmission detection device

developed by our group [16], as shown in Figure 1. The light sources are two rows of halogen lamps, 5 in a row, 10 in total. The parameters of the halogen lamps are 12 V and 100 W, which provide light sources for collecting spectrum information in the diffuse transmission mode. The apples are placed on the fruit cups and transferred to the dark box by the chain. The halogen lamps illuminate the apples passing by, and the light through the interior of the apples is received by the optical fiber and transferred to the computer through the spectrometer, which has a wavelength range of 350~1150 nm, and the exposure time of the samples is adjusted by the supporting spectrum acquisition software. The device was preheated for 30 min before the spectrum acquisition, the detection speed of the device was set to 0.5 m/s, and the exposure time was 100 ms. Each sample was collected four times at the equatorial part, and the average spectrum was taken as the experimental spectrum of that sample.

2.3. Parameter Measurement. The SSC content of apple samples was measured using a refractive digital saccharimeter (PR-101a, Japan). The measurement process was as follows: a fruit knife was used to cut off part of the flesh of the spectrum collection site on the four sides, and the juice was squeezed out of the flesh and dropped on the measurement position of the saccharimeter to measure the saccharimetric value of this side of the apple. The average SSC value of the four sides was taken as the SSC value of the apple sample. The fruit diameter at the equatorial position of the apple was measured using a digital vernier caliper (Mitutoyo-500, Japan). Each apple was measured four times at the equatorial position and the average fruit diameter was taken as the fruit diameter of that fruit.

2.4. Data Processing. The Kennard-Stone (K-S) algorithm was first applied to classify the collected apple spectra. The collected spectrum data were imported using unscrambler software to establish the SSC model of apples. The prediction set correlation coefficient R_p judged the performance of the model and prediction set root mean square error value (RMSEP). The formulae for R_p and RMSEP are shown in equations (1) and (2), respectively.

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

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

where n is the number of samples in the prediction set, \hat{y}_i is the predicted value of the i -th sample in the prediction set, y_i is the true value of the i -th sample in the prediction set, and \bar{y} is the average of the true values of all samples in the prediction set.

PLS is the most commonly used multivariate linear correction technique, which is widely used in NIR spectroscopy to predict the internal quality of fruits quantitatively, and the principle of PLS prediction is shown in

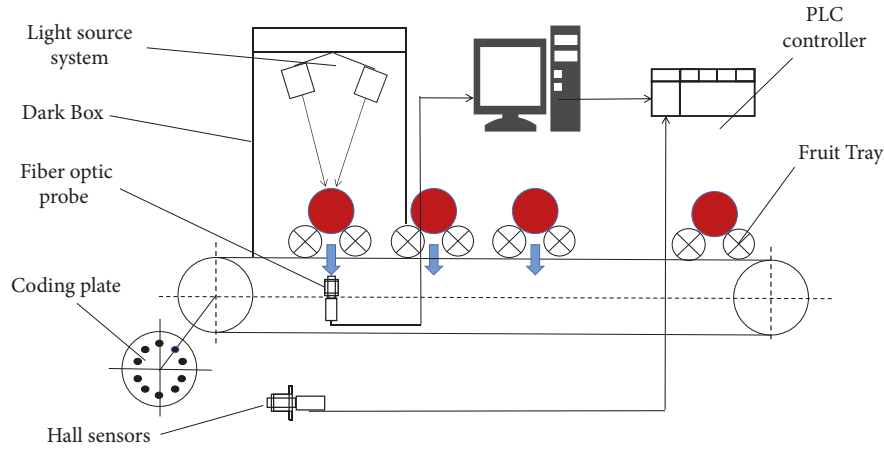


FIGURE 1: NIR diffuse transmission online detection device.

$$Y = \sum_{i=1}^n \beta_i \lambda_i + B, \quad (3)$$

where Y is the model prediction, i denotes the i -th wavelength point, β_i denotes the regression coefficient value corresponding to the i -th wavelength point, λ_i is the spectrum energy value corresponding to the i -th wavelength point, n is the number of wavelength points, and B is the intercept.

3. Results

3.1. Analysis of Apple SSC and Measurement Results. The experimental samples of 480 apples were classified separately using the K-S algorithm for the modeling set and the prediction set, 160 samples under each fruit diameter group, of which 120 were in the modeling set and 40 in the prediction set, and the SSC measurements of apples are shown in Table 1. The SSC content range of the modeling set was more comprehensive (9.05–16.4 Brix) than that of the prediction set (9.65–14.85 Brix), which could achieve a better prediction for the apple SSC model.

3.2. Analysis of NIR Spectrum Characteristics of Apples with Different Fruit Diameters. The average spectra of apples in the three fruit diameters were taken and compared with the spectra of three different fruit diameters, as shown in Figure 2. The peaks at 640 nm, 710 nm, and 800 nm and the troughs at 675 nm and 755 nm are mainly related to the multiplicative stretching vibrations of C-H and O-H bonds at 710 nm [17, 18]. The peaks at 805 nm are mainly associated with the secondary multiplicative absorption of C-H and N-H bonds [19, 20].

The energy spectrum of apples with a fruit diameter of 65 mm–75 mm is higher than apples with a fruit diameter of 75 mm–85 mm and 85 mm–95 mm. This phenomenon is because the energy carried by near-infrared light inside the apple decays as the light range increases. At a given wavelength, the extinction rate of light entering the apple interior is approximated as an exponential decay function [21], which can be fitted as

$$I = I_0 \exp(-u_e d), \quad (4)$$

where I_0 is the light intensity entering the interior of the apple, I is the light intensity received by the fiber optic probe below the apple, u_e is the extinction coefficient, and d is the distance from the point where the light enters the apple to the point where the light exits the apple.

In the NIR online detection device, as in Figure 1, d in formulae (3) is the fruit diameter of the apple. As the fruit diameter increases, the energy of the light is absorbed more and more inside the apple, resulting in a lower energy value of its collected NIR spectrum. It is thus speculated that different fruit diameters of the apple will cause differences in its NIR spectrum, which will impact the performance of the apple SSC prediction model built from the NIR spectrum.

3.3. Apple Soluble Solids Content Prediction Model for Each of the Three Fruit Diameters. The PLS was used to build the apple SSC prediction model, and the number of LVs was set to 1–20 to prevent the model from overfitting or underfitting. The spectra were pretreated by MSC, SNV, and S-G smoothing alternatively, and the results of the PLS model of sugar content built for the three groups of fruit diameter apples are shown in Table 2.

The results showed that the NIR spectra of apples with fruit diameters of 65 mm–75 mm were treated with SNV. Their model predictions were the best, with R_p of 0.885 and RMSEP of 0.771%. Their scatter diagrams are shown in Figure 3(a). The NIR spectra of apples with fruit diameters of 75 mm–85 mm were treated with SNV, and their model predictions were the best, with R_p of 0.959 and RMSEP of the scatter diagram shown in Figure 3(b). The NIR spectra of apples with the fruit diameter of 85 mm–95 mm were best predicted by SNV, with R_p of 0.937 and RMSEP of 0.421%, and the scatter diagram is shown in Figure 3(c). The modeling results of the original spectra at the three fruit sizes and the modeling results after SNV pretreatment show that SNV could solve the problem of poor performance of its sugar prediction model due to differences in apple size to some extent, because SNV, as a pretreatment method that

TABLE 1: The statistical results of SSC of apple.

	Data set	Min/Brix	Max/Brix	Mean/Brix	Standard deviation
65 mm–75 mm	All	160	9.8	17.2	13.8
	Calibration set	120	9.8	17.2	13.8
	Prediction set	40	10.9	16.7	13.8
75 mm–85 mm	All	160	8.3	17.2	13.4
	Calibration set	120	8.3	17.2	13.2
	Prediction set	40	8.7	16.5	12.73
85 mm–95 mm	All	160	10.9	15.3	13.3
	Calibration set	120	10.9	15.3	13.3
	Prediction set	40	11.3	15.3	13.3

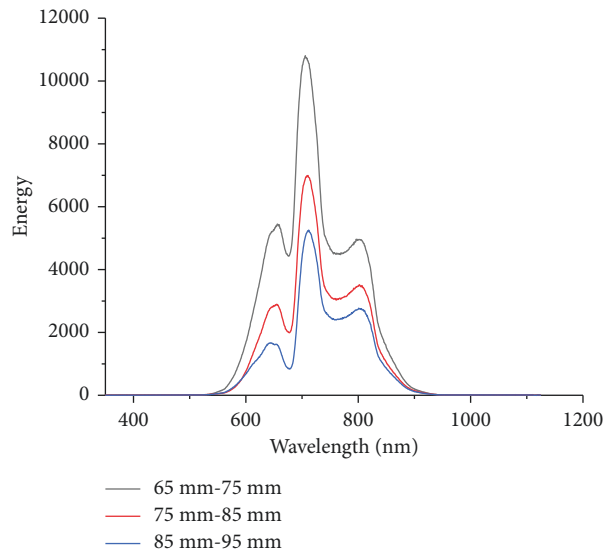


FIGURE 2: Near-infrared spectra of three fruit sizes of apples.

TABLE 2: Modeling results of apple SSC modeled for each of the three fruit diameters.

Fruit diameter range	Preprocessing methods	LVs	Rc	Rmsec/%	Rp	RMSEP/%
65 mm–75 mm	Raw	12	0.938	0.562	0.879	0.782
	MSC	13	0.921	0.664	0.859	0.789
	SNV	13	0.958	0.412	0.885	0.771
	S-G smoothing	12	0.912	0.667	0.878	0.789
75 mm–85 mm	Raw	12	0.976	0.375	0.947	0.643
	MSC	13	0.988	0.262	0.957	0.593
	SNV	14	0.989	0.257	0.959	0.578
	S-G smoothing	13	0.968	0.434	0.948	0.635
85 mm–95 mm	Raw	12	0.962	0.242	0.918	0.493
	MSC	10	0.933	0.318	0.932	0.482
	SNV	10	0.956	0.295	0.937	0.421
	S-G smoothing	12	0.931	0.322	0.918	0.492

can eliminate sample particle size, surface scattering, and light range variation [22], can solve the effect of spectrum scattering caused by uneven sample size.

3.4. Fruit Diameter Group 75 mm–85 mm Predicted Other Fruit Diameter Group SSC Prediction Model. As shown in Table 2, the modeling set of apple SSC model with fruit diameter 75 mm–85 mm had better performance than the other fruit diameter groups, so the modeling set with fruit

diameter 75 mm–85 mm was selected as the modeling set of the hybrid prediction model to investigate the performance of the apple SSC prediction model when the apple fruit diameter in the modeling set was different from the apple fruit diameter in the prediction set. The modeling set in the fruit diameter 75 mm–85 mm group was used to predict the prediction set in the fruit diameter 65 mm–75 mm and fruit diameter 85 mm–95 mm groups. The prediction model effects are shown in Table 3. The scatter diagrams are shown in Figures 4(a) and 4(b). The

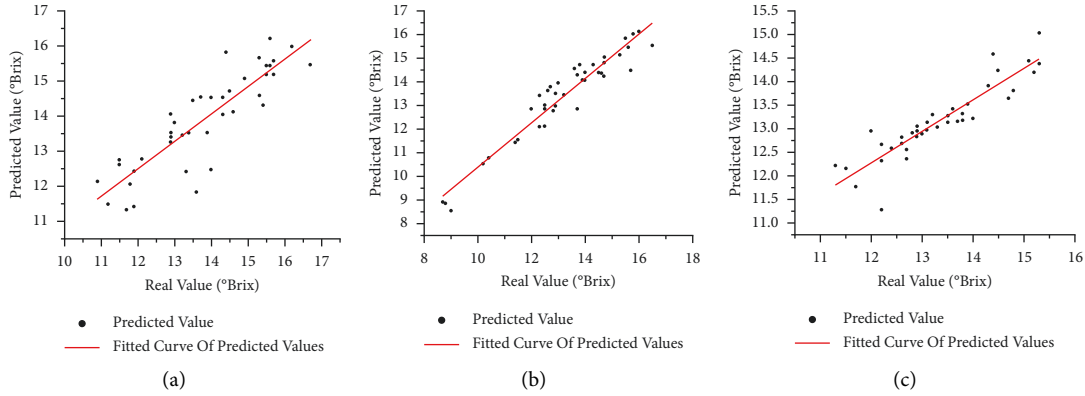


FIGURE 3: Predicted scatter diagram for each fruit size group. (a) Fruit diameter 65 mm–75 mm, (b) fruit diameter 75 mm–85 mm, and (c) fruit diameter 85 mm–95 mm.

TABLE 3: Results of the prediction model for SSC of apples in the remaining fruit size groups predicted from 75 mm–85 mm fruit diameter.

Modeling set fruit diameter	Predicted fruit set diameter	LVs	R_c	RmseC/%	R_p	RMSEP/%
75 mm–85 mm	65 mm–75 mm	9	0.918	0.686	0.779	0.877
	85 mm–95 mm	13	0.958	0.412	0.745	0.914

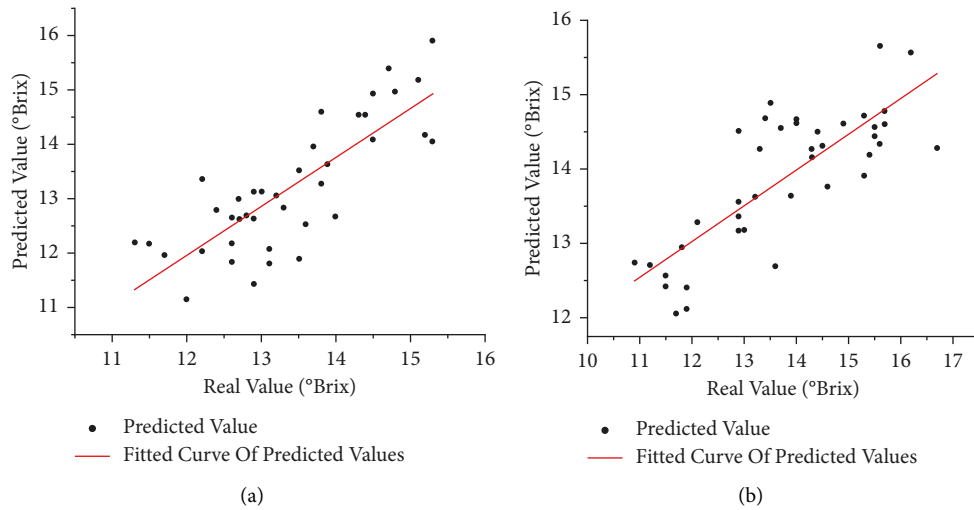


FIGURE 4: Scatter diagram of SSC prediction model. (a) 75 mm–85 mm fruit size group predicted 65 mm–75 mm fruit size group. (b) 75 mm–85 mm fruit size group predicted 85 mm–95 mm fruit size group.

results showed that the prediction model of apple SSC was poor when the difference between the modeling set and the prediction set of apple fruit diameter was significant, and compared with the modeling results of separate groups of fruit diameter in Table 2, the prediction set of apple SSC model for fruit diameter 65 mm–75 mm R_p decreased from 0.879 to 0.779 and RMSEP increased from 0.782% to 0.877%. The prediction set R_p decreased from 0.918 to 0.745 and RMSEP was risen from 0.493% to 0.914% for the model of the SSC of apples with 85 mm–95 mm fruit diameter. The difference in apple size had a significant influence on the prediction model of the SSC of apples. In the actual fruit sorting line, the size of apples can vary greatly, which can lead to the poor sorting

performance of the fruit sorting model, so size compensation of the apple SSC model is needed to increase the prediction performance of the sorting model.

3.5. Prediction Model of Soluble Solids Content under Apple Size Compensation

3.5.1. Analysis of the Relationship between Apple Size and Spectrum. Apples with fruit diameters of 65 mm, 67 mm, 69 mm, 71 mm, 73 mm, 75 mm, 77 mm, 79 mm, 81 mm, 83 mm, 85 mm, 87 mm, 89 mm, 91 mm, 93 mm, 95 mm were taken separately to establish the relationship between their fruit diameter and the light intensity of their collected NIR

spectra at 750 nm, as shown in Figure 5, from which it can be seen that the size of the apple will have an effect on its NIR spectrum and thus will have an effect on the performance of the apple SSC prediction model established by its NIR spectrum, so it is necessary to compensate the size of the apple SSC model to increase the prediction performance of the sorting model.

3.5.2. Mixed Apple Size Soluble Solids Content Prediction Model. From Table 3 and Figure 5, it is known that apple size differences will affect its NIR spectrum, so it is necessary to build an apple size compensation model to solve the influence of apple size differences on the model. 120 representative apple samples from each fruit size group were selected as the modeling set of the mixed apple size SSC prediction model using the K-S algorithm, and 40 apple samples were selected as the prediction set. The PLS prediction models for SSC of different apple sizes were constructed, the model effects are shown in Table 4, and the scatter diagrams are shown in Figure 6. The results show that the prediction models established for mixed apple sizes fit better, the model performance is better, the correlation coefficient R_p is significantly improved, the root mean square error value RMSEP of the prediction set is significantly reduced, and the model stability is significantly improved compared with the prediction models established in Table 3. The model stability was significantly improved, and the influence of the model on the SSC of apple due to apple size could be reduced.

As shown from Table 4, the constructed PLS model of mixed apple size SSC has improved its model prediction performance relative to Table 3 when the modeling set and prediction set of fruit diameter is different. Still, its prediction set root mean square error value RMSEP is as high as 0.911%. In the actual fruit sorting line, such a high error value will lead to inaccurate apple quality sorting. Add the size variable, and then build its SSC prediction model. Its model prediction results are shown in Table 5.

As shown in Table 5, the model prediction performance of the mixed apple size SSC prediction model improved after adding the size variable. The root means square error value RMSEP of the prediction set decreased from 0.911% to 0.822%, but the improvement of its size variable on the model performance was negligible. In the apple mixed size solids content model, its modeling set containing each group of the effect of size variables was diluted in the mixed size solids content model, resulting in an insignificant impact.

3.5.3. Development of a Size-Compensated Soluble Solids Content Prediction Model for Apples. In this study, the effect of apple fruit size on the SSC model of apple in NIR spectroscopy was investigated by the data fusion technique. The main objective of this technique is to optimize the amount of information on the tested sample metrics through the synergistic effect between the individual assays of the same sample [23], which consists of three levels of fusion: primary, intermediate, and advanced fusion. Primary fusion is the fusion and modeling of the raw data from multiple

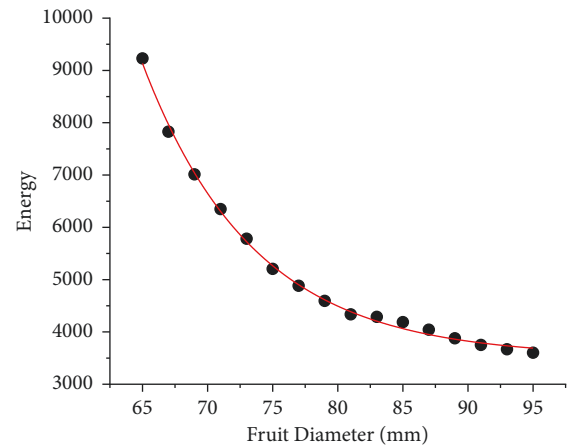


FIGURE 5: Diagram of apple fruit diameter versus spectrum.

assays; intermediate fusion is the screening of the effective variables of each assay and then data fusion and modeling; advanced fusion is the modeling of each assay independently and then the decision making after considering the results of each model. In recent years, data fusion techniques have been applied in several fields, such as metabolomics [24, 25], artwork identification [26, 27], dye classification [28], and food testing [29, 30]. In this study, a preliminary analysis of the effect of apple fruit diameter on the visible/near-infrared spectrum fruit nondestructive inspection model was conducted using a mid-level data fusion technique, the technical flow chart shown in Figure 7.

CARS and SPA were used to select the spectrum variables of the apple modeling set in the fruit diameter of 75 mm–85 mm to eliminate useless variables, further optimize the prediction model's performance, and improve the detection speed of the model. PLS modeling was performed in two cases: (1) PLS models of apple SSC were established with the wavelength variables selected by CARS or SPA. (2) The wavelength variables selected by CARS or SPA were fused with their corresponding apple size data and normalized to establish the PLS model of apple SSC. The selected spectrum wavelength points of CARS and SPA are shown in Figures 8(a) and 8(b), and the results of the established PLS model of apple SSC are shown in Table 6. The results showed that most of the wavelength points selected by CARS and SPA were located at 650 nm–850 nm. Most of them were at the peaks and valleys, indicating a large amount of information on the SSC of apples in this spectrum region. The poor performance of the SSC model built with the screened wavelengths after using SPA wavelength screening was caused by the fact that the SSC has multiple representations on the spectrum, and the wavelength selection was performed to remove a lot of useful information, which led to the poor performance of the model. Among them, the R_p of the model prediction set established by SPA for fruit diameter of 65 mm–75 mm was 0.744 and RMSEP was 1.340%; the R_p of the model prediction set for fruit diameter of 85 mm–95 mm was 0.665 and RMSEP was 1.942%. Using PLS to model the wavelength variables selected by CARS, the R_p of the model prediction is set for fruit diameter of 65 mm–75 mm. The number of wavelength variables used in

TABLE 4: PLS model results for SSC of mixed apple sizes.

Modeling set sample size	Prediction set sample size	LVs	R_c	Rmsec/%	R_p	RMSEP/%
120	40	11	0.932	0.869	0.853	0.911

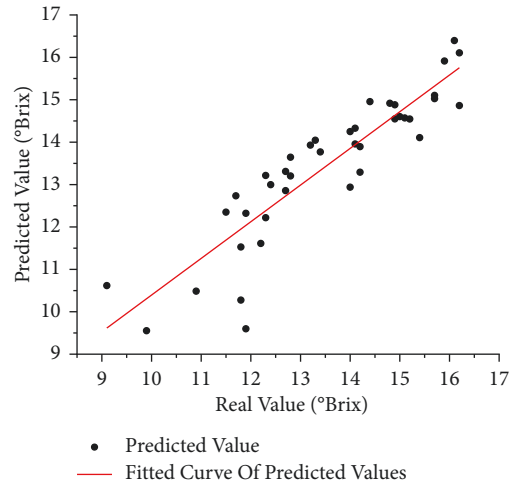


FIGURE 6: Scatter diagram of mixed fruit diameter prediction model.

TABLE 5: PLS model results for SSC of mixed apple sizes after adding size variables.

Modeling set sample size	Prediction set sample size	LVs	R_c	Rmsec/%	R_p	RMSEP/%
120	40	11	0.943	0.711	0.862	0.822

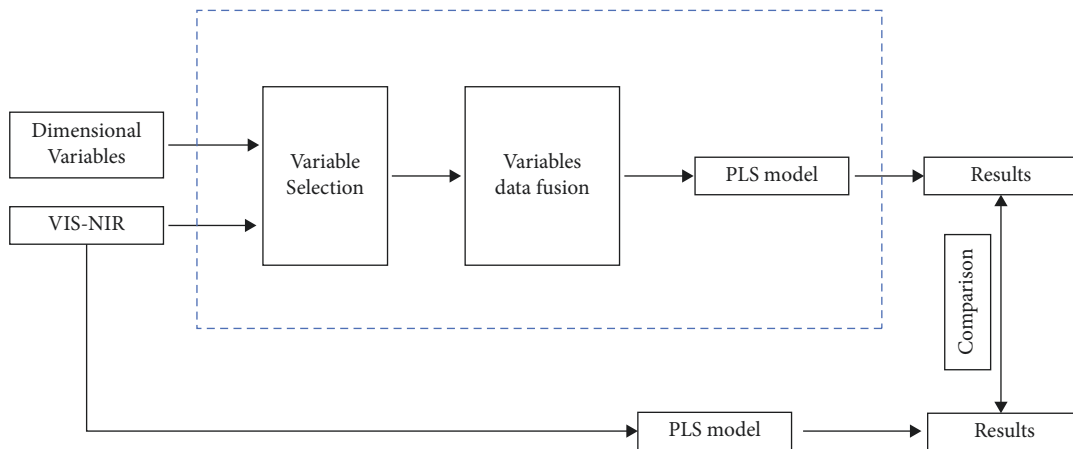


FIGURE 7: Flow diagram of data fusion technology.

the model decreases from 1044 to 61, effectively simplifying the model and improving its stability.

The apple size variables were added to the wavelength variables selected by CARS, and the data were normalized because the apple size data and spectrum data units do not coincide [31]. The effect of data normalization is to eliminate the impact of data dimensionality and make the data metrics comparable, which is essential for model building [32]. A PLS model of apple SSC with the intermediate level fusion of spectrum and apple fruit diameter data after normalization was established to show the model performance in Table 7. The model scatter diagrams are shown in Figures 9(a) and 9(b).

As shown in Table 7, compared with the PLS model of apple SSC built with CARS selected wavelength variables, the developed size-compensated intermediate fusion model had an improved prediction set R_p from 0.854 to 0.886 and a reduced RMSEP from 0.611% to 0.536% for fruit diameter of 65 mm–75 mm, and its prediction set R_p from 0.863 to 0.913 and a reduced RMSEP from 0.586% to 0.497% for fruit diameter of 85 mm–95 mm improved to 0.913 and RMSEP decreased from 0.586% to 0.497%. Compared with the mixed fruit size model, its prediction set R_p had a significant improvement and RMSEP had a considerable decrease. The results indicate that apple size influences the

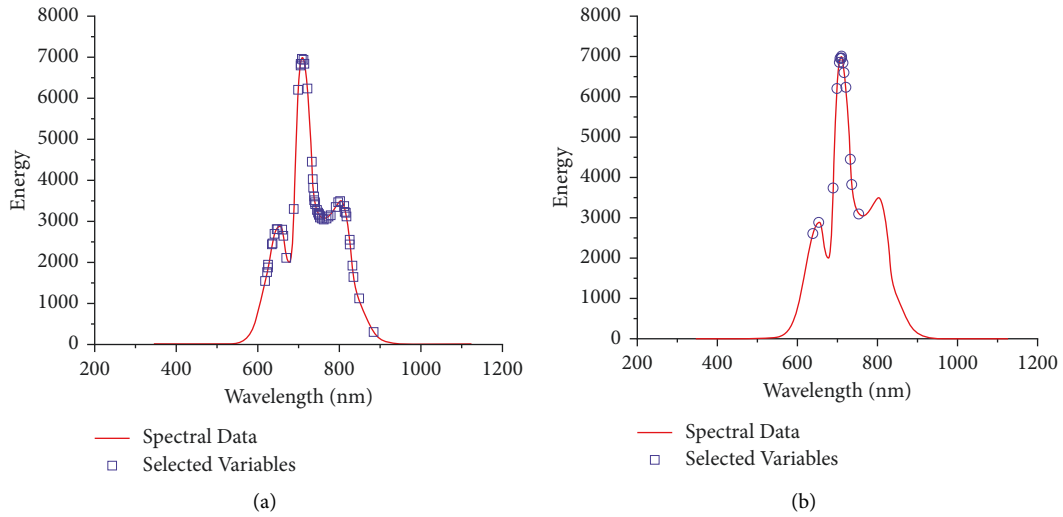


FIGURE 8: Distribution of the selected wavelength points. (a) CARS and (b) SPA.

TABLE 6: Results of the models built by CARS and SPA.

Sample selection algorithm	Modeling set fruit diameter	Predicted fruit set diameter	Number of variables	R_c	Rmsec/%	R_p	RMSEP/%
CARS	75 mm–85 mm	65 mm–75 mm	61	0.932	0.568	0.854	0.611
		85 mm–95 mm		0.932		0.863	0.586
SPA	75 mm–85 mm	65 mm–75 mm	16	0.815	0.942	0.744	1.340
		85 mm–95 mm		0.815		0.665	1.942

TABLE 7: PLS model for SSC of apples with intermediate data fusion.

Sample selection algorithm	Modeling set fruit diameter	Predicted fruit set diameter	Number of variables	R_c	Rmsec/%	R_p	RMSEP/%
CARS	75 mm–85 mm	65 mm–75 mm	61	0.967	0.440	0.886	0.536
		85 mm–95 mm		0.967		0.913	0.497

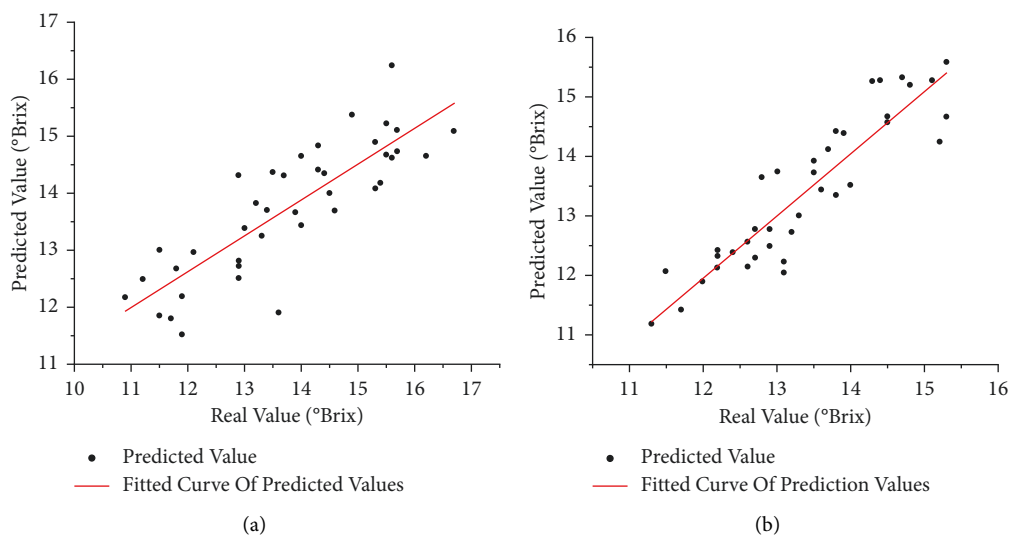


FIGURE 9: Intermediate fusion model scatter diagram. (a) Fruit diameter 75 mm–85 mm predicted fruit diameter 65 mm–75 mm, and (b) fruit diameter 75 mm–85 mm predicted fruit diameter 85 mm–95 mm.

performance of the apple SSC model, and adding fruit diameter variables to establish a size-compensated model of apple SSC can improve the prediction performance of the model.

4. Conclusion

This paper investigated the effect of apple fruit diameter differences on its SSC prediction model. The results showed that apple size differences will have an impact on its spectrum, and apple size and its spectrum light intensity satisfy the relationship of the logarithmic function, which will eventually have an effect on the prediction performance of the PLS model of apple SSC established by it. For this reason, the solution methods of different size differences and different preprocessing models, and apple fruit size were studied. We found that SNV is a pretreatment method that can eliminate sample particle size, surface scattering, and light range variation. It can solve the poor performance of the prediction model of SSC due to the difference in apple size to a certain extent. The correlation coefficient R_p was significantly improved; the root means the square error of the prediction set RMSEP was reduced considerably. The stability of the model was dramatically enhanced, which could reduce the influence of apple size on the model of apple SSC. The prediction set R_p for fruit diameter 65 mm–75 mm is 0.886 and RMSEP is 0.536%, and the prediction set R_p for fruit diameter 85 mm–95 mm is 0.913 and RMSEP is 0.497%, which is the best model performance. Therefore, adding the fruit diameter variable to establish the size compensation model of apple SSC can improve the model's prediction performance and meet the requirements of online detection of the SSC of apples with different fruit diameters.

Data Availability

The raw data in our research cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors report no conflicts of interest.

Authors' Contributions

Xiaogang Jiang carried out investigation, validation, and formal analysis. Mingwang Zhu did writing of original draft, data curation, and methodology. Jinliang Yao provided validation and software. Yuxiang Zhang provided resources. Yande Liu contributed with resources.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number: 31760344) and the Science and Technology Research Project of Education Department of Jiangxi Province (grant number: GJJ200615).

References

- [1] X. Xu, H. Xu, L. Xie, and Y. Ying, "Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system," *Journal of Food Measurement and Characterization*, vol. 13, no. 1, pp. 506–512, 2018.
- [2] H. Chen, Z. Liu, K. Cai, L. Xu, and A. Chen, "Grid search parametric optimization for FT-NIR quantitative analysis of solid soluble content in strawberry samples," *Vibrational Spectroscopy*, vol. 94, pp. 7–15, 2018.
- [3] H. Lee, M. S. Kim, H. S. Lim, E. Park, W. H. Lee, and B. K. Cho, "Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: application to seeds of the "Sambok Honey" cultivar," *Biosystems Engineering*, vol. 148, pp. 138–147, 2016.
- [4] T. Sun, X. X. Mo, and M. H. Liu, "Effect of fruit peel on the accuracy of visible/near infrared detection of soluble solids in navel oranges," *Spectroscopy and Spectral Analysis*, vol. 38, no. 5, p. 1406, 2018.
- [5] Z. M. Guo, W. Q. Huang, Q. S. Chen, Q. Y. Wang, C. Zhang, and W. J. Zhao, "Design and experiment of transmission spectroscopy online detection system for apple heart rot disease," *Journal of Agricultural Engineering*, vol. 32, no. 6, pp. 283–285, 2016.
- [6] Y. D. Liu, X. D. Sun, and A. G. Ouyang, "Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN," *Lebensmittel-Wissenschaft und -Technologie- Food Science and Technology*, vol. 43, no. 4, pp. 602–607, 2010.
- [7] L. Li, Y. K. Peng, and Y. Y. Li, "Design and testing of online non-destructive testing grading system for internal and external quality of apples," *Journal of Agricultural Engineering*, vol. 34, no. 9, p. 27, 2018.
- [8] D. H. Han, X. X. Liu, C. Lu, H. Wang, and X. D. Sun, "Study on optical non-destructive detection of internal browning in apples," *Journal of Agricultural Machinery*, vol. 37, no. 6, 2006.
- [9] X. X. Liu, D. H. Han, R. L. Tu, and C. H. Wen, "Non-destructive detection of changes in apple water heart disease during storage," *Journal of Agricultural Engineering*, vol. 20, no. 1, pp. 211–214, 2004.
- [10] X. Xu, J. C. Mo, L. J. Xie, and Y. Ying, "Influences of detection position and double detection regions on determining soluble solids content (SSC) for apples using on-line visible/near-infrared (vis/NIR) spectroscopy," *Food Analytical Methods*, vol. 12, no. 9, pp. 2078–2085, 2019.
- [11] Y. D. Liu, Y. Rao, X. D. Sun et al., "Influence of size difference on the detection model of navel orange soluble solids content by visible near infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 40, no. 10, p. 6, 2020.
- [12] S. Tian, J. Zhang, Z. Zhang, and H. Zhang, "Effective modification through transmission Vis/NIR spectra affected by fruit size to improve the prediction of moldy apple core," *Infrared Physics and Technology*, vol. 100, pp. 117–124, 2019.
- [13] V. A. McGlone, P. J. Martinsen, C. J. Clark, and R. B. Jordan, "On-line detection of Brownheart in Braeburn apples using near infrared transmission measurements," *Postharvest Biology and Technology*, vol. 37, no. 2, pp. 142–151, 2005.
- [14] J. Qin and R. Lu, "Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique," *Postharvest Biology and Technology*, vol. 49, no. 3, pp. 355–365, 2008.

- [15] J. Qin and R. Lu, "Monte Carlo simulation for quantification of light transport features in apples," *Computers and Electronics in Agriculture*, vol. 68, no. 1, pp. 44–51, 2009.
- [16] Y. D. Liu, Y. Zhang, X. G. Jiang, and H. Liu, "Detection of the quality of juicy peach during storage by visible/near infrared spectroscopy," *Vibrational Spectroscopy*, vol. 111, Article ID 103152, 2020.
- [17] I. Murray and P. Williams, "Near infrared technology in the agricultural and food industries," *Chemical principles of near-infrared technology*, vol. 12, no. 2, pp. 17–34, 1987.
- [18] R. M. McDevitt, A. J. Gavin, S. Andrés, and I. Murray, "The ability of visible and near infrared reflectance spectroscopy to predict the chemical composition of ground chicken carcasses and to discriminate between carcasses from different genotypes," *Journal of Near Infrared Spectroscopy*, vol. 13, no. 3, pp. 109–117, 2005.
- [19] L. J. Zhou, H. Wu, J. T. Li, Z. Y. Wang, and L. Zhang, "Determination of fatty acids in broiler breast meat by near-infrared reflectance spectroscopy," *Meat Science*, vol. 90, no. 3, pp. 658–664, 2012.
- [20] N. Prieto, D. W. Ross, R. I. Navajas et al., "Online prediction of fatty acid profiles in crossbred Limousin and Aberdeen Angus beef cattle using near infrared reflectance spectroscopy," *Animal*, vol. 5, no. 1, pp. 155–165, 2011.
- [21] D. G. Fraser, R. Kunemeyer, V. A. McGlone, and R. B. Jordan, "Letter to the editor," *Postharvest Biology and Technology*, vol. 22, no. 3, pp. 191–194, 2001.
- [22] J. Li, S. K. Sun, L. W. Jiang, X. Liu, S. H. Ding, and P. Li, "Study on non-destructive identification method of green tea based on near infrared spectroscopy and chemometrics," *Journal of Analytical Testing*, vol. 39, no. 11, pp. 1344–1350, 2020.
- [23] E. Borrás, J. Ferré, R. Boque, M. Mestres, L. Acena, and O. Bustó, "Data fusion methodologies for food and beverage authentication and quality assessment—a review," *Analytica Chimica Acta*, vol. 891, 2015.
- [24] J. Boccard and D. N. Rutledge, "A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion," *Analytica Chimica Acta*, vol. 769, pp. 30–39, 2013.
- [25] B. Dieme, S. Mavel, H. Blasco et al., "Metabolomics study of urine in autism spectrum disorders using a multiplatform analytical methodology," *Journal of Proteome Research*, vol. 14, no. 12, pp. 5273–5282, 2015.
- [26] P. M. Ramos and I. Ruisánchez, "Data fusion and dual-domain classification analysis of pigments studied in works of art," *Analytica Chimica Acta*, vol. 558, pp. 274–282, 2006.
- [27] P. M. Ramos, M. P. Callao, and I. Ruisánchez, "Data fusion in the wavelet domain by means of fuzzy aggregation connectives," *Analytica Chimica Acta*, vol. 584, no. 2, pp. 360–369, 2007.
- [28] C. Fernández, M. Pilar Callao, and M. Soledad Larrechi, "UV-visible-DAD and ¹H-NMR spectroscopy data fusion for studying the photodegradation process of azo-dyes using MCR-ALS," *Talanta*, vol. 117, pp. 75–80, 2013.
- [29] M. S. Godinho, M. R. Blanco, F. F. Gambarra Neto et al., "Evaluation of transformer insulating oil quality using NIR, fluorescence, and NMR spectroscopic data fusion," *Talanta*, vol. 129, pp. 143–149, 2014.
- [30] E. Borrás, J. Ferré, R. Boqué et al., "Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA)," *Food Chemistry*, vol. 203, pp. 314–322, 2016.
- [31] Y. S. Wang, J. Li, B. Wang, Y. T. Zhang, and J. L. Geng, "Assessment of crude protein and moisture content in sorghum based on near infrared spectroscopy," *Journal of Animal Nutrition*, vol. 32, no. 03, pp. 1353–1361, 2020.
- [32] S. Rinnan, F. V. D. Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TRAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.