

Improving SSC detection accuracy of cherry tomatoes by feature synergy and complementary spectral bands combination



Yuanhao Zheng^{a,b}, Penghui Liu^{a,b}, Yingjie Zheng^{a,b}, Lijuan Xie^{a,b,*}

^a College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, PR China

^b Key Laboratory of Intelligent Equipment and Robotics for Agriculture of Zhejiang Province, Hangzhou 310058, PR China

ARTICLE INFO

Keywords:

Ultraviolet/Visible/Near-infrared spectra
Universal strategy
Characteristic absorption
Spectral bands combination

ABSTRACT

The visible/near-infrared (VIS/NIR) spectroscopy technique has been extensively employed for the non-destructive detection of soluble solids content (SSC) in fruit. However, some existing algorithms and modeling optimization methods for improving the detection accuracy of SSC may be applicable to specific conditions and not universal. To break through this bottleneck problem, we propose a potentially universal strategy based on feature synergy and spectral bands combination of multiple detection modes. Firstly, compared with the commonly used halogen lamp, the xenon lamp was utilized to cover the ultraviolet (UV) range and match the characteristic information of different sugars (glucose, fructose, and sucrose). Furthermore, the spectral fusion method was employed to combine the reflectance and transmittance spectra to improve the cherry tomatoes SSC prediction accuracy with comprehensive data in the UV/VIS/NIR region. Finally, the best results were achieved by the partial least square regression (PLSR) model with spectral bands fusion, yielding the results of R_p^2 , RMSEP, and RPD as 0.9653, 0.1998%, and 5.31, respectively. The same conclusion can also be verified when predicting the SSC in strawberries. Overall, this strategy is potentially universal to improve the prediction accuracy of SSC in fruit by matching light source spectra with characteristic absorption, as well as utilizing comprehensive spectral information in the 200–1100 nm range, providing valuable insights for the practical application of nondestructive detection of fruit internal quality.

1. Introduction

The visible/near-infrared (VIS/NIR) spectroscopy technique is regarded as one of the most suitable and efficient methods to evaluate the internal quality of various fruit (Cortés et al., 2019). It has been successfully applied in scientific research and practical commercial sorting lines (Huang et al., 2008; Porep et al., 2015) to detect indicators such as soluble solids content (SSC), acidity, ripeness, internal diseases, and phytochemicals in fruit. Among these indicators, SSC is a crucial parameter in fruit quality assessment, as it is related to maturity and serves as a key factor influencing the taste and flavor of fruit (Tian et al., 2023). Improving its prediction accuracy could enhance product supply quality and meet consumer demands.

SSC primarily contains soluble sugars such as glucose, fructose, and sucrose (Desnoes et al., 2014), as well as trace quantities of organic acids, minerals, vitamins, and other soluble compounds (Michailidis et al., 2024). Many researchers have dedicated substantial efforts to improving the SSC prediction accuracy based on VIS/NIR technology

(Pandiselvam et al., 2022). These studies have systematically addressed and enhanced the SSC detection accuracy from three primary perspectives (Xu et al., 2019): system construction and optimization (Xie et al., 2016), innovations in algorithms and modeling methods (Zareef et al., 2020), and mitigation of influence factors such as temperature and size (Ren et al., 2019; Jiang et al., 2022). However, diverse conclusions may arise based on distinct experimental conditions regarding modeling methods and reducing influence factors, which means that the algorithm optimization may not be universal. For example, in terms of modeling methods, Wang et al. (2022) achieved the optimal prediction model for apple SSC using the CARS-PLS combination method, with effective wavelengths around 690 nm, 733 nm, 840 nm, and 888 nm, while Zheng et al. (2023) delivered the best result with different wavelength ranges around 735 nm, 755 nm, 780 nm, and 820 nm by the same method. When it comes to the correction of influencing factors, the highest correlation exhibited between the diameter and absorbance values at 825 nm with the diameter correction method (DCM) to reduce the impact of size on the orange spectra (Tian et al., 2022), which was

* Corresponding author at: College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, PR China.
E-mail address: ljxie@zju.edu.cn (L. Xie).

different from the highly correlated wavelength 844 nm identified by Kawano et al. (1993). These inconsistent conclusions arise because some algorithms and models concentrate on chemometric aspects, the modeling process with a ‘black-box’ nature (Zhang et al., 2021). In essence, the interaction of the characteristic absorption and spectra is closely related to the Lambert-Beer law, which is one of the theoretical foundations for chemometric models in the VIS/NIR spectroscopy technique (Qi et al., 2023). To develop a potentially universal method to improve the detection accuracy of SSC, understanding the absorption feature of each primary component of SSC, mainly soluble sugars (Dai et al., 2016), and effectively obtaining the corresponding characteristic signal is essential. For this purpose, we propose a strategy from the perspective of feature synergy and spectral bands combination, considering both the characteristic absorption of sugars in theory and the acquisition of spectral signals in practical applications.

Specifically, on the one hand, feature synergy means that the light source spectrum must cover all the characteristic bands of the objective substances, and the spectrum peaks align optimally with these features. In this way, the complete information of the target can be obtained, and the signal-to-noise ratio will be relatively high due to the synergy (Zhang et al., 2014). In practice, low-cost halogen tungsten lamp covering the VIS and NIR regions is the most common choice (Lee et al., 2023), while its radiation energy is relatively low within 200–400 nm, leading to the loss of some valuable information, such as the weak absorption of the carbonyl group in the primary SSC components (glucose, fructose, and sucrose) (Pacsu, 1948). As for the characteristic bands, using a narrowband light source could capture partial target information (Zhu et al., 2021), however, relative spectral data may lose due to the complex overlapped nature of NIR spectra (Sun et al., 2022), which affects detection accuracy. Therefore, it is essential to consider selecting a broadband light source that covers the UV/VIS/NIR region and matches the characteristic regions to obtain comprehensive spectral information.

On the other hand, the spectral bands combination means that the complementary reflectance and transmittance spectra are integrated for complete spectral information. Generally, there are three modes available for application: transmittance, reflectance, and interactance mode (Nicolaï et al., 2007). Although the transmittance mode enables the acquisition of more internal information of samples, especially in the NIR region, the strong absorption of fruit in 200–600 nm still prevents the information collection in the UV region. For the reflectance and non-contact interaction modes, these two similar modes can obtain information in the UV region due to their limited penetration depth. However, they also introduce more noise information such as specular reflection (Mollazade et al., 2012). Thus, considering the existence of characteristic signals in the UV region and the spectral quality in the NIR region, the combination of reflectance and transmittance spectral data could be employed to improve the accuracy of non-destructive SSC detection.

The purpose of this study is to validate the feasibility of the proposed strategy based on feature synergy and spectral bands combination, taking the improvement of prediction performance for cherry tomatoes SSC as an illustrative case and further confirming the obtained conclusion using strawberry samples. The specific contents are to: (1) elucidate the synergy relationship between the characteristic absorption of soluble sugars and the spectrum of the xenon lamp, (2) compare and explain the performance of the prediction models for cherry tomato SSC under different light sources and detection modes, and (3) further improve SSC prediction performance using raw-spectral fusion with complementary spectral bands.

2. Materials and methods

2.1. Samples

In this study, to determine the absorption spectra of carbohydrates, 400 mg of pure glucose, fructose, and sucrose (RHAWN, China) were

weighed using an electronic balance (SQP, Sartorius, Germany) and compressed with a tablet press (YP-24 T, Jinfulun, China) at 10 Mpa for 3 min, respectively. Three sheets of each substance were prepared for repeated spectroscopic determination.

In this research, we chose cherries tomatoes and strawberries as illustrative cases to cover a wider SSC range and to minimize the impact of fruit diameter on the prediction model. A total of 208 fully ripe ‘Millennium’ cherry tomatoes (*Solanum lycopersicum* var. *cerasiforme*) were purchased from the local supermarket in Hangzhou, China. All samples were free of damage and externally uniform, with diameters ranging from 25.82 mm to 32.66 mm and masses ranging from 11.22 g to 19.81 g. Before spectra collection, these samples were numbered and maintained at 25 °C for 12 h to ensure uniformity.

Strawberries were also selected as the detection subjects to verify the applicability of the proposed strategy. A total of 199 strawberries were purchased from the local supermarket, and all these samples were free of damage and externally uniform. These strawberry samples were also numbered and maintained at 25 °C for 12 h before the collection of spectra. The same experimental methods as those used for cherry tomatoes were employed for strawberry data collection and processing.

2.2. Experiments

2.2.1. Spectra acquisition

The transmittance and diffuse reflectance spectra of the samples were collected using a self-built spectral acquisition system, as depicted in Fig. 1. The xenon lamp, with spectral coverage in VIS and NIR regions and significant emission in the UV range (Gaudin et al., 2006), is considered as the alternative light source. The light source in the system comprised either a xenon lamp (ProSp-Xe300-UV, ProSp, China) or a halogen lamp (HL-2000-HP, Ocean Insight, USA). The light was propagated through an optical fiber, interacted with the sample, and was received by a spectrometer (Maya2000 Pro, Ocean Optics, USA) via another bifurcated optical fiber for detection.

The carbohydrate tablet was positioned on the sample holder, and diffuse reflectance spectra were collected under two different light sources with a uniformly textured polytetrafluoroethylene (PTFE) sphere served as the reference. Subsequently, the absorbance spectra were calculated for each type of sugar, which facilitated a visual analysis of the characteristic absorption features of the compounds. Following this, transmittance and reflectance spectra of cherry tomato samples were collected using the halogen and xenon lamp as the light source, respectively. To ensure comparability across different light sources, the integration times for the halogen and xenon lamps were set to 50 ms and 30 ms, respectively. Each spectrum was collected 5 times and smoothed to obtain high-quality spectra and minimize the impact of noise. All spectra were collected within a dark box to minimize the interference from ambient light. Additionally, the spectra of each sample were collected three times, and the average was taken as the final spectra with a more accurate representation.

2.2.2. SSC measurement

After acquiring the spectral data, the SSC of each cherry tomato was promptly determined immediately (Cao et al., 2023). Each sample was washed, chopped, crushed into the uniform mixture, and filtered using a sieve to obtain clear juice. The juice was then measured for SSC with a refractometer (PR-101α, Atago, Japan) calibrated with distilled water. Each sample was measured three times, and the average value was calculated as the result.

2.3. Chemometric methods

2.3.1. Preprocessing method

During the spectral acquisition process, the raw spectral data not only contains the characteristic absorption of samples but also includes high-frequency random noise, baseline drift, light scattering, and other

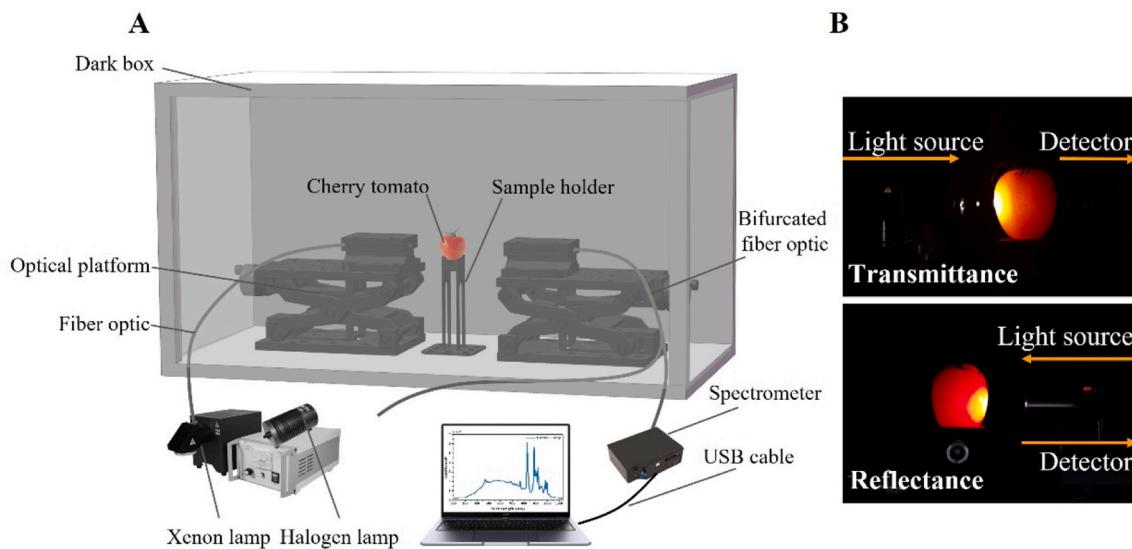


Fig. 1. Spectral acquisition system: (A) Schematic diagram of the UV/VIS/NIR spectra acquisition system, (B) Schematic diagram of transmittance and reflectance detection modes.

similar factors. Therefore, the preprocessing of the spectra is very important to achieve the purpose of noise reduction and mitigate other interference. In this study, the spectra were automatically smoothed during the spectra data collection. Standard normal variate (SNV) transform was chosen as the preprocessing method for scattering correction (Oliveri et al., 2019).

2.3.2. PLSR modeling

Partial least squares regression (PLSR) is a multivariate statistical technique commonly employed for modeling the relationship between predictor variables (spectra data) and a response value (SSC). PLSR has been proven effective in handling high-dimensional datasets and addressing collinearity issues in fields such as chemometrics and spectroscopy (Wold et al., 2001). It was used to develop the SSC prediction models in this study. In the modeling process, the number of latent variables (LVs) was determined through 10-fold cross-validation, and the maximum value of the LVs was set to 20 to avoid overfitting. The optimal number of LVs was chosen as the point where the root mean square error of cross-validation (RMSECV) no longer exhibited a significant decrease. Then, to assess the effective information of the spectra, all samples were randomly divided into a calibration set and a prediction set at a ratio of 4:1 before modeling. The modeling process was performed 10 times, and the results were averaged as the model result.

2.3.3. Fusion method

Multispectral data fusion has the potential to enhance model accuracy and stability. In this study, raw-level fusion (Cocchi, 2019) was adopted based on the modeling result and application scenarios to take advantage of the complementarity of different spectral bands. After preprocessed with SNV, the complementary spectral bands from different detection modes were directly concatenated into a new matrix, and a quantitative PLSR model of SSC was established with this matrix (Wu et al., 2016).

2.4. Model evaluation

The following parameters were used to evaluate the model performance of different spectra on SSC: determination coefficient of calibration (R_c^2) and prediction (R_p^2), root mean square error of calibration (RMSEC) and prediction (RMSEP), and ratio of the prediction data standard deviation to RMSEP (RPD). These parameters are calculated as

follows:

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

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

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

Where n is the number of samples, y is the reference SSC values, y_i is the reference SSC value of the i^{th} sample, \bar{y} is the mean value of reference SSC and \hat{y}_i is the SSC value of i^{th} sample predicted by the model. SD refers to the standard deviation of the reference SSC values in the prediction set. Typically, superior model performance is indicated by higher R^2 , RPD values, and lower RMSE. Additionally, the RMSEC and RMSEP should be as close as possible, with an RMSEP/RMSEC ratio below 1.2, and an RPD value greater than 2 (Ma et al., 2021).

All spectra data were obtained using spectrum suite (Ocean Optics, USA). The data processing and modeling procedure was conducted using MATLAB (R2021b) software with libPLS toolbox (Li et al., 2018).

3. Results and discussion

3.1. Spectral data analysis

It is essential to identify the characteristic absorption of the fundamental components of SSC to improve the spectral interpretability and model universality. For this purpose, the absorption information of soluble sugars (glucose, fructose, and sucrose) under different light sources (halogen lamp, xenon lamp) was obtained. The spectral information of the two light sources were also compared, as shown in Fig. 2.

The diffuse reflectance absorbance spectra of glucose, fructose, and sucrose tablets under halogen lamp illumination are depicted in Fig. 2A. In the NIR region, glucose exhibits two absorption peaks at 922 nm and 1029 nm, fructose displays two absorption peaks at 918 nm and 1000 nm, and sucrose demonstrates two absorption peaks at 920 nm and 985 nm. These absorption peaks are associated with the vibrational absorption band of O-H (Weyer et al., 2007). However, the absorbance value sharply decreases around 400 nm, and only relatively weak

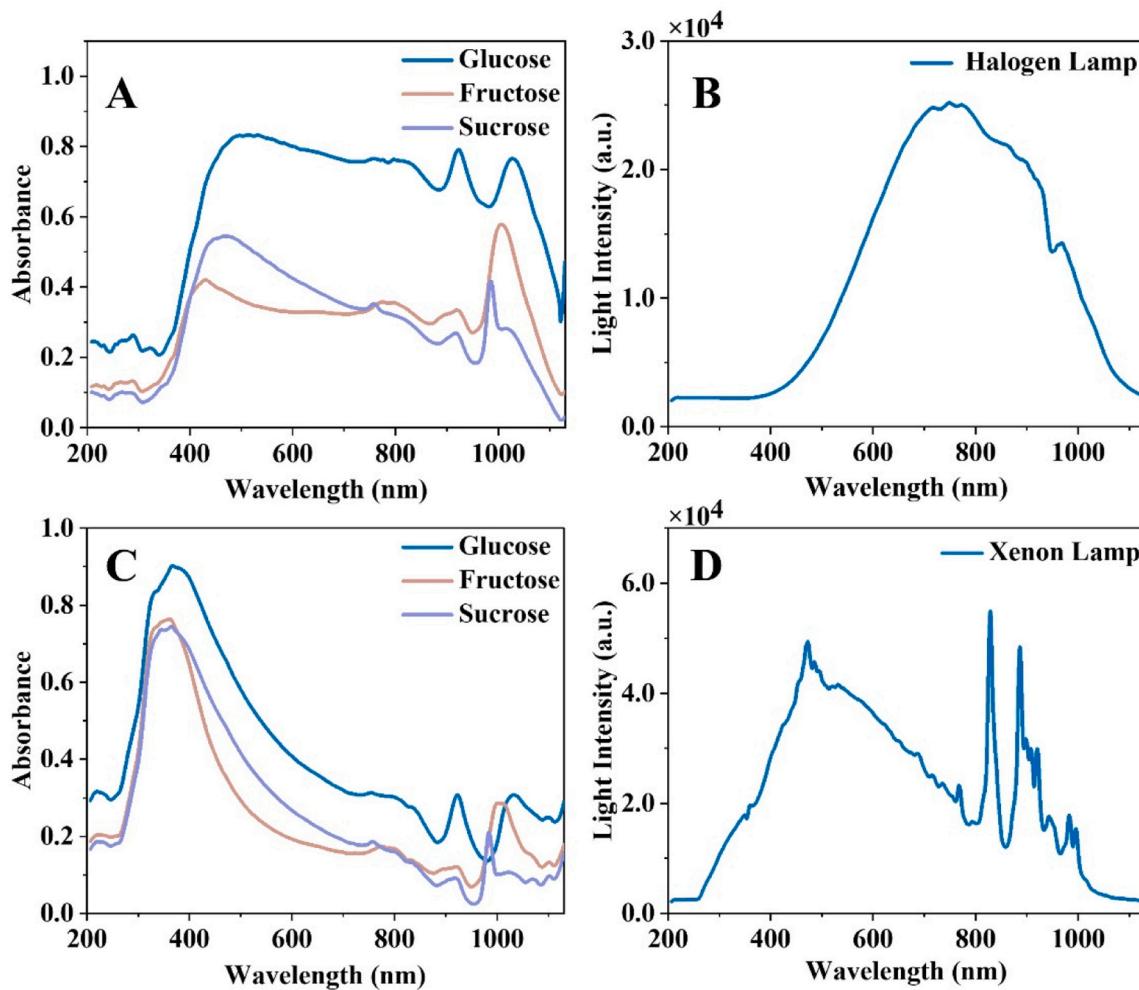


Fig. 2. Absorption data of soluble sugars and spectral information of different light sources: The absorption spectra of sugar tablets under the halogen lamp (A) and the xenon lamp (C), the light source spectra of the halogen (B) and xenon lamp (D).

information was observed in 200–400 nm. This absence is attributed to the weak light intensity generated by the halogen lamp in the corresponding wavelength range. As depicted in Fig. 2B, the light generated by the halogen lamp in the UV region is relatively limited.

To confirm the presence of absorption information of fruit soluble sugars in 200–400 nm, diffuse reflectance spectra of these tablets were collected using a xenon lamp as the light source. Its spectrum is depicted in Fig. 2D, and it could generate light across a broader band starting from 260 nm (Esen et al., 2017). Fig. 2C depicts the absorbance information under the xenon lamp. In the NIR spectral region, glucose, fructose, and sucrose exhibit absorption peaks in the exactly same wavelength when different light sources are used. However, these sugars display prominent absorption in the 260–400 nm range compared with halogen lamps when a xenon lamp is employed as the light source. This absorption is related to the carbonyl groups in sugars (Akash and Rehman, 2020), which indicates that the commonly used halogen lamp could miss valuable information in the UV band in determining fruit SSC. Due to the broader wavelength range of the xenon lamp, it could cover the ultraviolet region and allow for the acquisition of more comprehensive raw data. On the other hand, the xenon lamp exhibits spectral peaks with relatively higher intensity at the wavelengths corresponding to the absorption peaks of soluble sugars around 920 nm and 1000 nm. This correlation might impact the accuracy of soluble solids detection for the high signal-to-noise ratio (Gaudin et al., 2006) when using a xenon lamp as the light source. To validate this, the subsequent discussion will compare the prediction performance of SSC in cherry

tomato using the xenon and halogen lamp as the light source, respectively.

3.2. Model results using different light sources

3.2.1. Spectral data analysis of cherry tomatoes

The spectra of cherry tomatoes were collected under different light sources to compare their SSC prediction performance. Furthermore, the spectra from different detection modes were also gathered in order to obtain the comprehensive information (Fig. 3).

Fig. 3A and Fig. 3C depict the transmittance spectra of cherry tomatoes under the halogen and xenon lamp, respectively. The 200–600 nm band spectra have high noise due to strong absorption. Notably, the xenon lamp exhibits intense spectral regions around 830 nm and 890 nm, leading to higher transmittance energy and, consequently, stronger spectral signals in this range. Compared to the transmittance mode, Fig. 3B and Fig. 3D display the reflectance spectra of cherry tomatoes under these two kinds of light sources, respectively, where spectra signals could be observed in 200–600 nm. Additionally, the reflectance spectra under the halogen lamp still lacks spectral information in 200–400 nm, due to the intrinsic spectral characteristics of the halogen lamp. In contrast, the reflectance spectra under the xenon lamp exhibit relatively complete spectra throughout the 200–600 nm range. Further attempts were made to combine the reflectance spectra (200–600 nm) with the transmittance spectra (600–1100 nm) to predict SSC using a more comprehensive UV/VIS/NIR spectral dataset in the

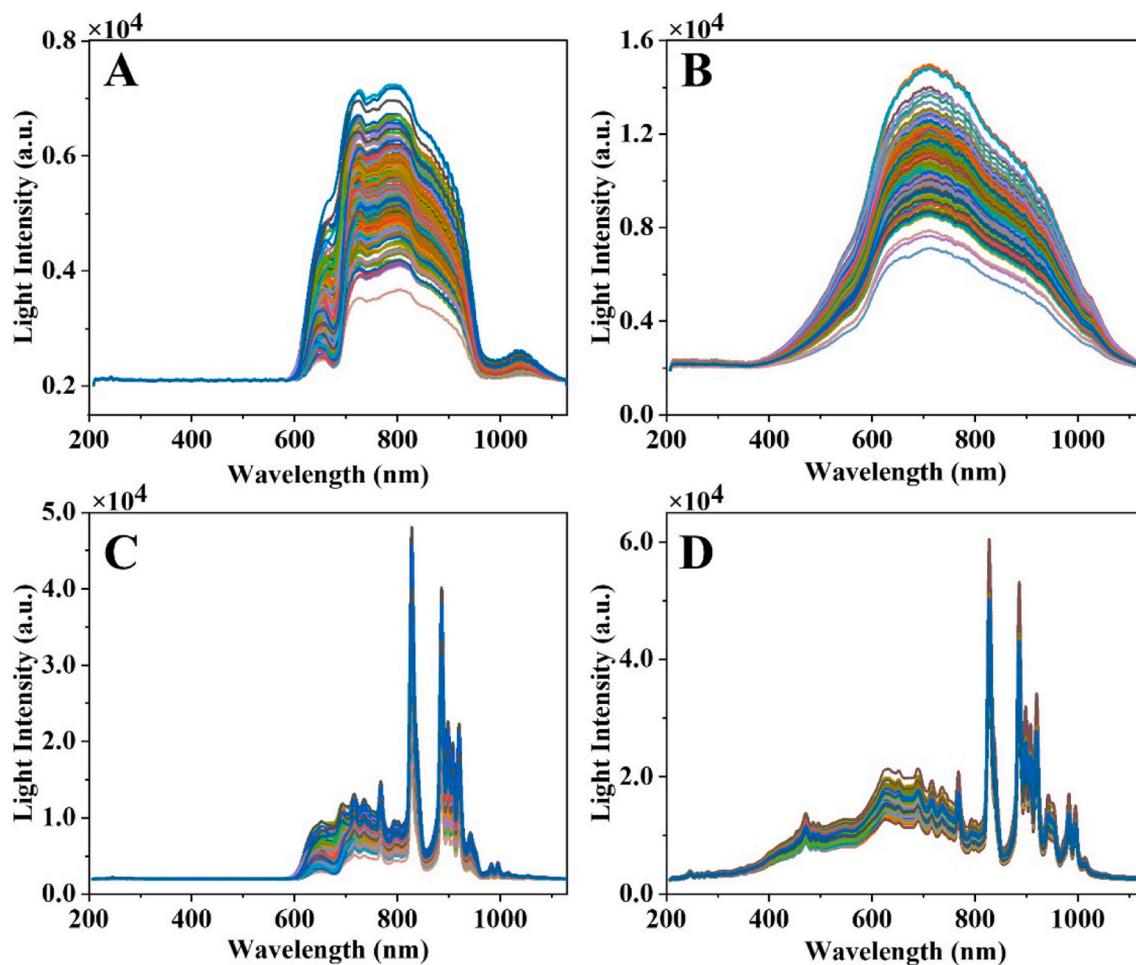


Fig. 3. Raw spectra of cherry tomato samples: The transmittance (A) and reflectance (B) spectra of cherry tomatoes under the halogen lamp, and the transmittance (C) and reflectance (D) spectra of cherry tomatoes under the xenon lamp.

modeling process.

3.2.2. Comparison and analysis of modeling results

The prediction model of SSC in cherry tomatoes was conducted using the raw spectra of different light sources and various detection modes to further compare the effective information. SNV was employed as the preprocessing method to eliminate the influence of scattering.

The results of modeling are presented in Table 1. Firstly, the prediction performance of the transmittance mode is superior to the reflectance mode under the same light source. The average values of R_p^2 , RMSEP, and RPD of the prediction model using xenon lamp under the transmittance mode are 0.9358, 0.2543%, and 3.92, respectively. In contrast, the prediction results using the xenon lamp under reflectance mode are 0.7424, 0.5563%, and 1.95, respectively. A similar conclusion is drawn when using a halogen lamp as the light source. This result is consistent with the findings of the previous researchers (Wang et al., 2014). This phenomenon could be attributed to the fact that the light

emitted by the light source can pass through the entire sample in the transmittance mode, carrying a more representative spectra of the entire fruit. On the contrary, for the reflectance mode, the light only interacts with part of the sample's surface, and this mode also incorporates invalid information due to the specular reflection, which results in less information for the reflectance spectra that can be used for the model prediction (Magwaza et al., 2012). As a result, the modeling effect of the transmittance spectra outperforms the reflectance spectra. Secondly, the modeling result using the xenon lamp spectra is better than that of the halogen lamp in the transmittance mode. Compared with the model result of the halogen lamp, the RMSEP value for modeling with xenon lamp spectra decreases by 0.0873%, and the RPD value increases by 0.94, indicating a notable improvement. The superior prediction performance of the xenon lamp may be related to (1) the synergy relationship between the xenon lamp spectrum and the characteristic absorption of soluble sugars, (2) the stronger light intensity of certain spectral regions, and (3) the broader spectral range and the more comprehensive spectral information.

Firstly, the synergy between the xenon lamp spectrum and the characteristic absorption of sugars could contribute to the improvement of SSC prediction performance. Due to the inherent emission characteristics of xenon lamps, their spectra include strong regions at specific wavelengths. If these specific ranges match the absorption peaks of the target substance (in this study, SSC), the light source will provide higher energy at these absorption peaks. This means that more light energy will be absorbed under this matching pattern, thereby improving the signal-to-noise ratio and detection sensitivity (Zhang et al., 2014). On the one

Table 1

The PLSR model results for SSC with different light sources under various detection modes.

Light Source	Detection Mode	R_c^2	RMSEC (%)	R_p^2	RMSEP (%)	RPD
Xenon Lamp	Transmittance	0.9579	0.2062	0.9358	0.2543	3.92
	Reflectance	0.7665	0.4758	0.7424	0.5563	1.95
Halogen Lamp	Transmittance	0.9235	0.2760	0.8895	0.3416	2.98
	Reflectance	0.7827	0.4639	0.7092	0.5518	1.85

hand, as mentioned in Section 3.1, glucose has apparent absorption around 923 nm and 1032 nm, fructose has notable absorption near 920 nm and 1000 nm, and for sucrose around 920 nm and 985 nm. Correspondingly, the xenon lamp exhibits strong spectral regions at 921 nm, 982 nm, 996 nm, and 1016 nm. This matching pattern of the xenon lamp spectra with the characteristic absorption of the sugar components may enhance the interaction between the xenon lamp and sugars, thereby improving the accuracy of SSC detection. On the other hand, NIR spectroscopy is based on the stretching, bending, or rotating vibrations of chemical bonds in molecular structures (Goodarzi et al., 2015). The overtones and combinations regions involving the stretching and bending vibrations of the X-H bond in hydrogen-containing groups are of particular significance (Polesello et al., 1983). In the NIR region, if characteristic frequency absorption and the stretching vibration bands of the functional groups match the peaks of the xenon lamp itself, it may also enhance the response of the light to substances containing those specific functional groups, finally improving the detection sensitivity of such substances. Glucose is a kind of polyhydroxy aldehyde, fructose is a kind of polyhydroxy ketone, and sucrose is formed by the glycosidic bonds linking one glucose and one fructose molecule (Qi and Tester, 2019). These sugar components contain hydrogen-containing groups, including C-H and O-H bonds, contributing to the absorption in the NIR region through combination and overtone vibrations. Correspondingly, the xenon lamp exhibits strong spectral regions around 768 nm, associated with the fifth stretching frequency of the C-H(5v) for methylene and the fourth stretching frequency of O-H(4v) for primary and secondary alcohols. Additionally, the xenon lamp displays strong spectral regions near 996 nm, corresponding to the stronger third stretching vibration of the O-H(3v) of primary and secondary alcohols (Weyer et al., 2007). The feature synergy is demonstrated in Fig. 4A, and the spectra-structure correlations are shown in Table 2. This matching pattern between the xenon lamp spectra and the characteristic absorption of substances containing specific functional groups may also enhance the xenon lamp's response to sugars, thereby improving the accuracy of SSC detection.

To further elucidate how feature synergy is manifested in the PLSR model, the regression coefficient is presented in Fig. 4B. Larger absolute regression coefficients are observed at 763 nm, 913 nm, and 999 nm, indicating significant contributions of the spectra around these wavelengths to the model (Liu et al., 2015). These wavelengths correspond to the absorption of sugars and the theoretical absorption of C-H and O-H groups in the NIR range, which supports the analysis of feature synergy. In contrast, the relatively low regression coefficients in the 200–600 nm range indicate that these spectra provide less effective information due to strong absorption.

Secondly, certain spectral bands with stronger light intensity of the

Table 2
Spectra-Structure Correlations for Near-Infrared. *

Wavelength (nm)	Functional group	Structure-Spectra	Spectral peak of xenon lamp (nm)
762	C-H of (-CH ₂)	C-H (5v)	768
767	O-H of (-CH ₂ -OH)	O-H (4v)	768
773	O-H of (-CH-OH)	O-H (4v)	768
996	O-H of (-CH ₂ -OH)	O-H (3v)	996
1004	O-H of (-CH-OH)	O-H (3v)	996

*With reference to Practical Guide to Interpretive Near-Infrared Spectroscopy (Weyer et al., 2007). The vibration of the hydrogen-containing group is represented by the symbol v.

xenon lamp spectrum means higher penetration. Owing to the inherent characteristics of the xenon lamp, it exhibits several high-intensity spectral regions in the NIR region. In the practical detection of fruit samples, these high-intensity spectral regions possess greater penetrability. They can carry the overtones and combination information of hydrogen-containing groups through the entire sample directly (Rady et al., 2020). This contributes to an increased signal-to-noise ratio, thereby improving the detection accuracy of SSC. In contrast, the spectrum of the halogen lamp is relatively smooth, lacking distinct high-penetrability bands. This difference in penetrability could result in varied modeling effects between the xenon and halogen lamp. Besides, the stronger penetrability of the xenon lamp means lower integration time, which can increase the detection speed in continuous detection on practical production lines (Cortés et al., 2019), providing a distinct advantage in the application.

Thirdly, the superior SSC prediction performance of the xenon lamp is also related to the broader spectral range. The xenon lamp spectrum covers UV, VIS, and NIR regions, which facilitates covering more vibration modes and absorption peaks, increasing the ability to utilize spectroscopy for the detection and analysis of specific compounds. Compared to commonly used halogen lamps, this broad spectral range is mainly reflected in the ultraviolet-visible band, allowing access to information regarding electronic transitions in this range (Qi et al., 2023). By determining the absorption spectra of pure sugar tablets, it is observed that these sugars exhibit absorption in the UV region, where the xenon lamp already generates sufficient intensity in the ultraviolet-visible range. If the information in this spectra region can be acquired, it may serve as a supplement to the NIR region spectra, enhancing the effective information related to SSC in the raw spectra. However, the strong absorption in the ultraviolet-visible spectra

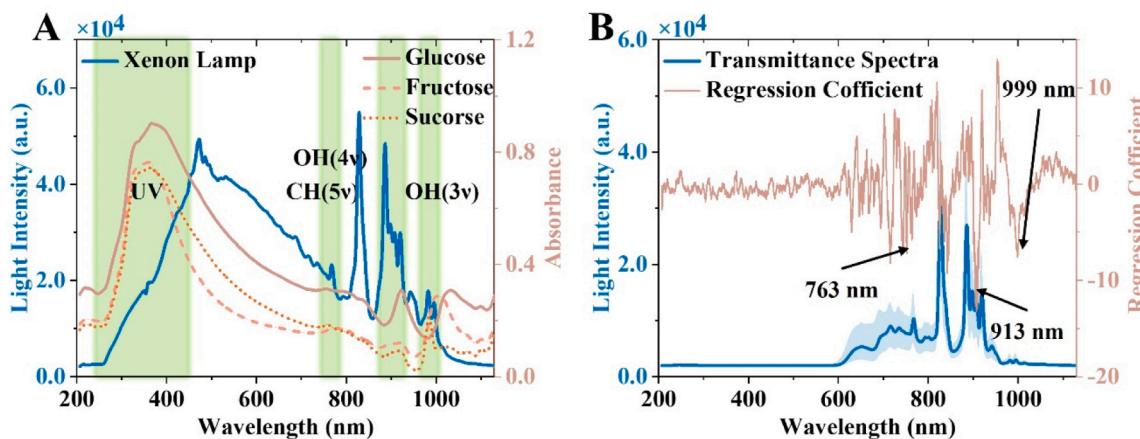


Fig. 4. Feature synergy and its manifestation in the PLSR model: (A) The synergy between characteristic absorption of sugars and xenon lamp spectrum, (B) The regression coefficient of PLSR model using xenon lamp under transmittance mode.

(especially 200–600 nm) typically results in high noise levels under the transmittance detection mode (Schaare and Fraser, 2000) in the practical detection of fruit samples. To capture more effective information in this spectral range, it is necessary to adjust the light source arrangement and adopt spectral fusion methods to obtain information related to the target substance in the UV/VIS range.

3.3. Complementary spectral fusion

Cherry tomatoes exhibit strong absorption in the 200–600 nm region, and their transmittance spectra display high noise and limited effective information in this range. By contrast, the xenon lamp under reflectance mode could capture the reflection information in the 200–600 nm range, as shown in Fig. 5A. To make full use of more effective information from the original spectra for the prediction of SSC, we attempted to integrate the xenon lamp reflectance spectra (200–600 nm) with the transmittance spectra (600–1100 nm). The modeling results before and after the fusion were compared as follows.

Firstly, after excluding the high-noise bands in 200–600 nm, the modeling results using only the 600–1100 nm range of xenon lamp transmittance spectra are shown in Fig. 5C. The average values of R_p^2 , RMSEP, and RPD are 0.9494, 0.2338%, and 4.51, respectively. This indicates a better predictive performance for SSC than the model established using the complete spectra within 200–1100 nm. It further

illustrates that the spectra collected under the xenon lamp transmittance mode in 200–600 nm contain less effective information. Then, to compensate for the deficiency of transmittance spectra, the xenon lamp reflectance spectra in 200–600 nm and the transmittance spectra in 600–1100 nm are combined to predict the SSC in cherry tomatoes. The model results show the values of R_p^2 , RMSEP, and RPD are 0.9653, 0.1998%, and 5.31, respectively (Fig. 5D). The performance remains better than the result for the halogen lamp under transmittance mode (Fig. 5B), and also outperforms the modeling results before spectral fusion. This may be attributed to the complementation of effective information for the UV and VIS/NIR regions. Specifically, the combination of the reflectance and transmittance spectra could obtain information in the whole 200–1100 nm range for modeling. In the UV region, the energy absorption caused by the $n \rightarrow \pi^*$ electronic transition in the C=O bond is associated with the reflectance spectra (Akash and Rehman, 2020). In the NIR region, the transmittance spectra are related to the overtones of C-H and O-H in sugars. Through the spectral fusion in 200–1100 nm, the complete characteristic information about these CH, CH₂, OH, and C=O functional groups and chemical bonds present in glucose, fructose, and sucrose is obtained. As a result, the prediction performance for SSC has been significantly improved.

To further verify the applicability of the feature synergy and spectral bands combination method, strawberries are taken as the detection subject and the experimental procedures identical to those used for

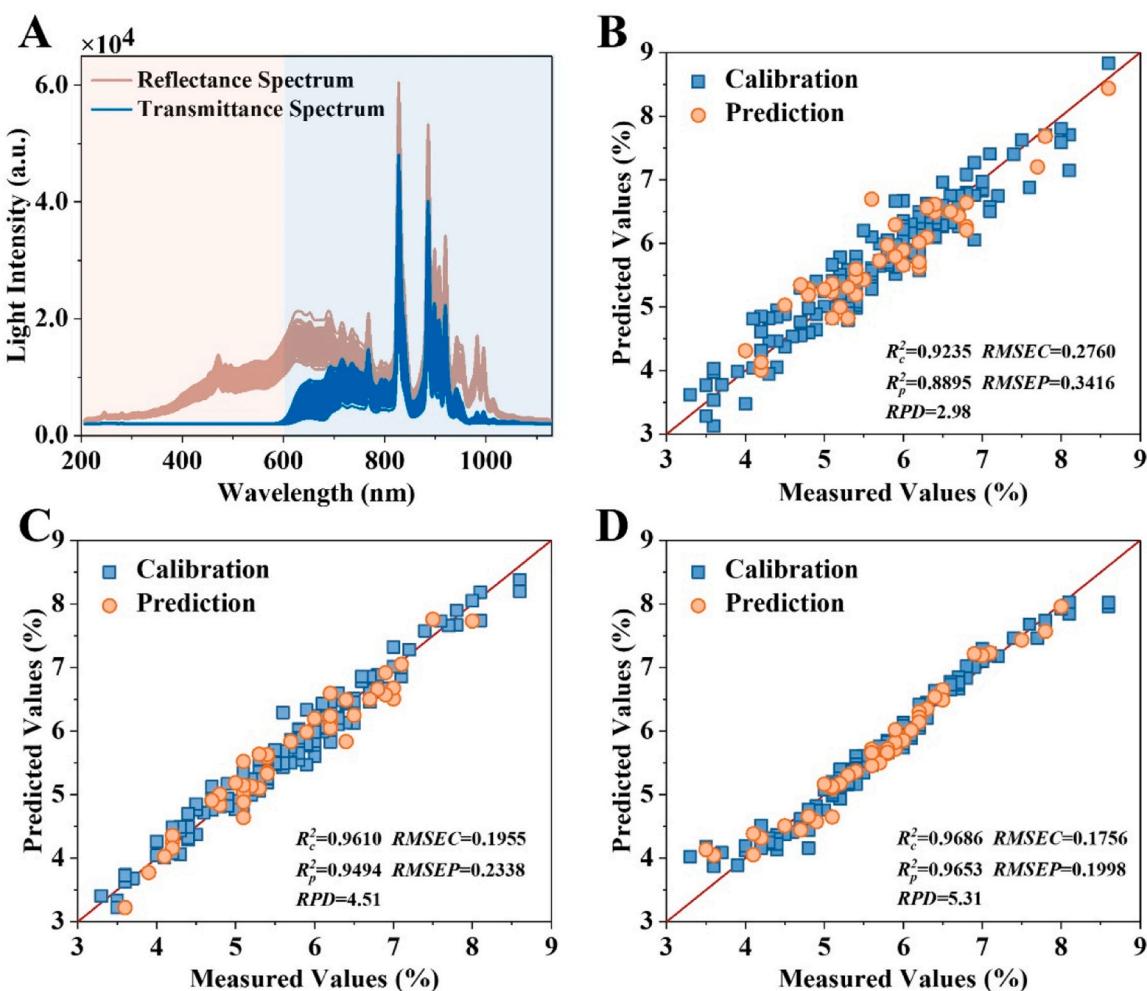


Fig. 5. The reflectance and transmittance spectra combination and PLSR model results for cherry tomatoes SSC. The spectral information of cherry tomatoes under reflectance and transmittance modes with a xenon lamp (A), and PLSR model results for cherry tomatoes SSC using: the 200–1100 nm range of halogen lamp under transmittance mode (B), the 600–1100 nm range of xenon lamp under transmittance mode (C), and the raw-level fusion of 200–600 nm reflectance spectra and 600–1100 nm transmittance spectra under xenon lamp (D).

cherry tomatoes (data acquisition, data processing, and model building) are employed to predict the SSC of strawberries. The same conclusions could be reached: the xenon lamp outperforms the halogen lamp in SSC prediction, and the combination of reflectance and transmittance spectra further enhances SSC prediction performance, as shown in Table. S1 and Fig. S2. The average values of R_p^2 , RMSEP, and RPD of the prediction model using combination spectra are 0.9249, 0.3421%, and 3.61, respectively. More details about the spectra and model results for strawberries can be found in the [supplementary material](#).

4. Conclusions

This study aims to develop a universal strategy to improve the prediction performance of fruit SSC based on UV/VIS/NIR spectroscopy technology. Compared to the commonly used halogen lamp, the xenon lamp could cover the UV region and match the characteristic absorption of primary SSC components. Based on this, our experiments validated that the xenon lamp performed better than the halogen lamp in predicting cherry tomato SSC. Subsequently, to overcome the limitation of the transmittance mode, raw-level fusion strategy was employed to combine the reflectance (200–600 nm) with transmittance (600–1100 nm) data to obtain more comprehensive spectral information for SSC prediction. In the end, the PLSR model with spectral fusion achieved the best results with the values of R_p^2 , RMSEP and RPD as 0.9653, 0.1998%, and 5.31, respectively. Also, the same conclusion can be drawn when predicting the SSC in strawberries using the same method.

Overall, to enhance the prediction performance of SSC, we proposed a potentially universal method involving feature synergy and complementary spectral bands combination. Specifically, feature synergy means matching the light source spectrum with the characteristic absorption, and spectral bands combination entails integrating spectra from different detection modes to obtain the comprehensive spectra. As for further application, a similar pattern could be applied to the non-destructive detection of other fruit indicators, such as acidity and phytochemicals. Moreover, validation of its detection effect on other fruit and conveyor sorting lines is still necessary for future implementation, since only cherry tomatoes and strawberries were used as illustrative cases in this study. Above all, we hope this study can help improve the accuracy and stability of fruit SSC prediction and provide some reference for developing universal methods in spectroscopy-based non-destructive fruit quality detection.

CRediT authorship contribution statement

Yuanhao Zheng: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Penghui Liu:** Methodology, Software, Supervision, Writing - review & editing. **Yingjie Zheng:** Visualization, Writing - review & editing. **Lijuan Xie:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge the financial support provided

by the National Key R&D Program of China (2022YFD2002204) and the support by the National Key Laboratory of Agricultural Equipment Technology.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.postharvbio.2024.112922](https://doi.org/10.1016/j.postharvbio.2024.112922).

References

- Akash, M.S.H., Rehman, K., 2020. Essentials of Pharmaceutical Analysis. Springer Nature Singapore, Singapore, pp. 29–33. https://doi.org/10.1007/978-981-15-1547-7_3. Ultraviolet-Visible (UV-VIS) Spectroscopy.
- Cao, Y., Xing, Z., Chen, M., Tian, S., Xie, L., 2023. Comparison of online quality prediction models of kiwifruit at different conveying speeds. *Food Meas.* 17, 686–693. <https://doi.org/10.1007/s11694-022-01645-2>.
- Cocchi, M., 2019. Data Fusion Methodology and Applications. Elsevier. 1. Introduction pp. 6–17. <https://doi.org/10.1016/B978-0-444-63984-4.00001-6>.
- Cortés, V., Blasco, J., Aleixos, N., Cubero, S., Talens, P., 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: a review. *Trends Food Sci. Technol.* 85, 138–148. <https://doi.org/10.1016/j.tifs.2019.01.015>.
- Dai, Z., Wu, H., Baldazzi, V., van Leeuwen, C., Bertin, N., Gautier, H., Wu, B., Duchêne, E., Gomès, E., Delrot, S., Lescourret, F., Génard, M., 2016. Inter-species comparative analysis of components of soluble sugar concentration in fleshy fruits. *Front. Plant Sci.* 7, 649. <https://doi.org/10.3389/fpls.2016.00649>.
- Desnoues, E., Gibon, Y., Baldazzi, V., Signoret, V., Génard, M., Quilot-Turion, B., 2014. Profiling sugar metabolism during fruit development in a peach progeny with different fructose-to-glucose ratios. *BMC Plant Biol.* 14, 336. <https://doi.org/10.1186/s12870-014-0336-x>.
- Esen, V., Sağlam, S., Oral, B., 2017. Light sources of solar simulators for photovoltaic devices: a review. *Renew. Sustain. Energy Rev.* 77, 1240–1250. <https://doi.org/10.1016/j.rser.2017.03.062>.
- Gaudin, K., Baillet, A., Chaminade, P., 2006. Application of a xenon arc lamp as a light source for evaporative light scattering detection. *Anal. Bioanal. Chem.* 384, 1302–1307. <https://doi.org/10.1007/s00216-006-0297-5>.
- Goodarzi, M., Sharma, S., Ramon, H., Saeyns, W., 2015. Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. *TrAC Trends Anal. Chem.* 67, 147–158. <https://doi.org/10.1016/j.trac.2014.12.005>.
- Huang, H., Yu, H., Xu, H., Ying, Y., 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *J. Food Eng.* 87, 303–313. <https://doi.org/10.1016/j.jfoodeng.2007.12.022>.
- Jiang, X., Zhu, M., Yao, J., Zhang, Y., Liu, Y., 2022. Calibration of near infrared spectroscopy of apples with different fruit sizes to improve soluble solids content model performance. *Foods* 11, 1923. <https://doi.org/10.3390/foods11131923>.
- Kawano, S., Fujiwara, T., Iwamoto, M., 1993. Nondestructive determination of sugar content in satsuma mandarine using near-infrared (NIR) transmittance. *J. Jpn. Soc. Hortic. Sci.* 62, 465–470. <https://doi.org/10.2503/jjshs.62.465>.
- Lee, H., Cho, S., Lim, J., Lee, A., Kim, G., Song, D.-J., Chun, S.-W., Kim, M.-J., Mo, C., 2023. Performance comparison of tungsten-halogen light and phosphor-converted NIR LED in soluble solid content estimation of apple. *Sensors* 23, 1961. <https://doi.org/10.3390/s23041961>.
- Li, H.-D., Xu, Q.-S., Liang, Y.-Z., 2018. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemom. Intell. Lab. Syst.* 176, 34–43. <https://doi.org/10.1016/j.chemolab.2018.03.003>.
- Liu, C., Yang, S.X., Deng, L., 2015. A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of navel oranges. *Expert Syst. Appl.* 42, 8497–8503. <https://doi.org/10.1016/j.eswa.2015.07.005>.
- Ma, F., Du, C., Zheng, S., Du, Y., 2021. In situ monitoring of nitrate content in leafy vegetables using attenuated total reflectance – fourier-transform mid-infrared spectroscopy coupled with machine learning algorithm. *Food Anal. Methods* 14, 2237–2248. <https://doi.org/10.1007/s12161-021-02048-7>.
- Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J.R., Saeyns, W., Nicolai, B., 2012. NIR Spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food Bioprocess Technol.* 5, 425–444. <https://doi.org/10.1007/s11947-011-0697-1>.
- Michailidis, M., Ziogas, V., Sarrou, E., Nasiopoulos, E., Stylianis Titeli, V., Skodra, C., Tanou, G., Ganopoulos, I., Martens, S., Molassisiotis, A., 2024. Screening the citrus greek national germplasm collection for fruit quality and metabolic footprint. *Food Chem.* 435, 137573. <https://doi.org/10.1016/j.foodchem.2023.137573>.
- Mollazade, K., Omid, M., Tab, F.A., Mohtasebi, S.S., 2012. Principles and applications of light backscattering imaging in quality evaluation of agro-food products: a review. *Food Bioprocess Technol.* 5, 1465–1485. <https://doi.org/10.1007/s11947-012-0821-x>.
- Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeyns, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol. Technol.* 46, 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>.
- Oliveri, P., Malegori, C., Simonetti, R., Casale, M., 2019. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Anal. Chim. Acta* 1058, 9–17. <https://doi.org/10.1016/j.aca.2018.10.055>.

- Pacsu, E., 1948. Open-Chain Sugars. I. Absorption in Ultraviolet of D-Glucose and L-Arabinose in Acid Solution. *J. Am. Chem. Soc.* 70 (2), 523–526. <https://doi.org/10.1021/ja01182a027>.
- Pandiselvam, R., Prithviraj, V., Manikantan, M.R., Kothakota, A., Rusu, A.V., Trif, M., Mousavi Khaneghah, A., 2022. Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: a comprehensive review. *Front. Nutr.* 9 <https://doi.org/10.3389/fnut.2022.973457>.
- Polesello, A., Giangiacomo, R., Dull, G.G., 1983. Application of near infrared spectrophotometry to the nondestructive analysis of foods: a review of experimental results. *C. R. C. Crit. Rev. Food Sci. Nutr.* 18, 203–230. <https://doi.org/10.1080/10408398309527363>.
- Porep, J.U., Kammerer, D.R., Carle, R., 2015. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* 46, 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>.
- Qi, X., Lian, Y., Xie, L., Wang, Y., Lu, Z., 2023. Water quality detection based on UV-Vis and NIR spectroscopy: a review. *Appl. Spectrosc. Rev.* 0 1–25. <https://doi.org/10.1080/05704928.2023.2294458>.
- Qi, X., Tester, R.F., 2019. Fructose, galactose and glucose – In health and disease. *Clin. Nutr. ESPEN* 33, 18–28. <https://doi.org/10.1016/j.clnesp.2019.07.004>.
- Rady, A., Fischer, J., Reeves, S., Logan, B., James Watson, N., 2020. The Effect of Light Intensity, sensor height, and spectral pre-processing methods when using NIR spectroscopy to identify different allergen-containing powdered foods. *Sensors* 20, 230. <https://doi.org/10.3390/s20010230>.
- Ren, S., Wu, C., Li, H., Ali, S., Akomeah Agyekum, A., Chen, Q., 2019. Model development for soluble solids and lycopene contents of cherry tomato at different temperatures using near-infrared spectroscopy. *Postharvest Biol. Technol.* 156, 110952 <https://doi.org/10.1016/j.postharvbio.2019.110952>.
- Schaare, P.N., Fraser, D.G., 2000. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biol. Technol.* 20, 175–184. [https://doi.org/10.1016/S0925-5214\(00\)00130-7](https://doi.org/10.1016/S0925-5214(00)00130-7).
- Sun, Y., Cai, W., Shao, X., 2022. Chemometrics: an excavator in temperature-dependent near-infrared spectroscopy. *Molecules* 27, 452. <https://doi.org/10.3390/molecules27020452>.
- Tian, S., Liu, W., Xu, H., 2023. Improving the prediction performance of soluble solids content (SSC) in kiwifruit by means of near-infrared spectroscopy using slope/bias correction and calibration updating. *Food Res. Int.* 170, 112988 <https://doi.org/10.1016/j.foodres.2023.112988>.
- Tian, S., Wang, S., Xu, H., 2022. Early detection of freezing damage in oranges by online Vis/NIR transmission coupled with diameter correction method and deep 1D-CNN. *Comput. Electron. Agric.* 193, 106638 <https://doi.org/10.1016/j.compag.2021.106638>.
- Wang, A., Hu, D., Xie, L., 2014. Comparison of detection modes in terms of the necessity of visible region (VIS) and influence of the peel on soluble solids content (SSC) determination of naval orange using VIS-SWNIR spectroscopy. *J. Food Eng.* 126, 126–132. <https://doi.org/10.1016/j.jfoodeng.2013.11.011>.
- Wang, J., Guo, Z., Zou, C., Jiang, S., El-Seedi, H.R., Zou, X., 2022. General model of multi-quality detection for apple from different origins by Vis/NIR transmittance spectroscopy. *Food Meas.* 16, 2582–2595. <https://doi.org/10.1007/s11694-022-01375-5>.
- Weyer, J.W.J., Lois, 2007. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press, Boca Raton. Appendix 4a, Spectra-Structure correlations for Near-Infrared. pp. 2–18, 79–81, 219–220. <https://doi.org/10.1201/9781420018318>.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst., PLS Methods* 58, 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- Wu, Z., Xu, E., Long, J., Pan, X., Xu, X., Jin, Z., Jiao, A., 2016. Comparison between ATR-IR, Raman, concatenated ATR-IR and Raman spectroscopy for the determination of total antioxidant capacity and total phenolic content of Chinese rice wine. *Food Chem.* 194, 671–679. <https://doi.org/10.1016/j.foodchem.2015.08.071>.
- Xie, L.J., Wang, A.C., Xu, H.R., Fu, X.P., Ying, Y.B., 2016. Applications of near-infrared systems for quality evaluation of fruits: a review. *Trans. ASABE* 59, 399–419. <https://doi.org/10.13031/trans.59.10655>.
- Xu, X., Xie, L., Ying, Y., 2019. Factors influencing near infrared spectroscopy analysis of agro-products: a review. *Front. Agr. Sci. Eng.* 6, 105–115. <https://doi.org/10.15302/J-FASE-2019255>.
- Zareef, M., Chen, Q., Hassan, M.M., Arslan, M., Hashim, M.M., Ahmad, W., Kutsaniedzie, F.Y.H., Agyekum, A.A., 2020. An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis. *Food Eng. Rev.* 12, 173–190. <https://doi.org/10.1007/s12393-020-09210-7>.
- Zhang, L., Xu, H., Gu, M., 2014. Use of signal to noise ratio and area change rate of spectra to evaluate the Visible/NIR spectral system for fruit internal quality detection. *J. Food Eng.* 139, 19–23. <https://doi.org/10.1016/j.jfoodeng.2014.04.009>.
- Zhang, X., Yang, J., Lin, T., Ying, Y., 2021. Food and agro-product quality evaluation based on spectroscopy and deep learning: a review. *Trends Food Sci. Technol.* 112, 431–441. <https://doi.org/10.1016/j.tifs.2021.04.008>.
- Zheng, Y., Cao, Y., Yang, J., Xie, L., 2023. Enhancing model robustness through different optimization methods and 1-D CNN to eliminate the variations in size and detection position for apple SSC determination. *Postharvest Biol. Technol.* 205, 112513. <https://doi.org/10.1016/j.postharvbio.2023.112513>.
- Zhu, Xingyue, Yu, K., Zhu, Xiaofan, Wu, C., 2021. Design and application of a nitrate measurement system based on a narrowband tunable ultraviolet light source. *IEEE Access* 9, 168268–168282. <https://doi.org/10.1109/ACCESS.2021.3135959>.