

Lycopene detection in cherry tomatoes with feature enhancement and data fusion



Yuanhao Zheng^{a,b}, Xuan Luo^{a,c}, Yuan Gao^{a,b,c}, Zhizhong Sun^d, Kang Huang^e, Weilu Gao^f, Huirong Xu^{a,b,c}, Lijuan Xie^{a,b,*}

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

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

^c Key Laboratory of On-Site Processing Equipment for Agricultural Products, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, PR China

^d College of Chemistry and Materials Engineering, Zhejiang A&F University, Hangzhou 311300, PR China

^e Department of Biological Systems Engineering, Washington State University, Pullman, WA 99164, USA

^f Department of Electrical and Computer Engineering, The University of Utah, 201 Presidents' Cir, Salt Lake City, UT 84112, USA

ARTICLE INFO

Keywords:

Lycopene content
UV/Vis/NIR spectroscopy
Machine vision
Spectra enhancement
Data fusion

ABSTRACT

Lycopene, a biologically active phytochemical with health benefits, is a key quality indicator for cherry tomatoes. While ultraviolet/visible/near-infrared (UV/Vis/NIR) spectroscopy holds promise for large-scale online lycopene detection, capturing its characteristic signals is challenging due to the low lycopene concentration in cherry tomatoes. This study improved the prediction accuracy of lycopene by supplementing spectral data with image information through spectral feature enhancement and spectra-image fusion. The feasibility of using UV/Vis/NIR spectra and image features to predict lycopene content was validated. By enhancing spectral bands corresponding to colors correlated with lycopene, the performance of the spectral model was improved. Additionally, direct spectra-image fusion further enhanced the prediction accuracy, achieving R_p^2 , RMSEP, and RPD as 0.95, 8.96 mg/kg, and 4.25, respectively. Overall, this research offers valuable insights into supplementing spectral data with image information to improve the accuracy of non-destructive lycopene detection, providing practical implications for online fruit quality prediction.

1. Introduction

As health awareness grows, people are becoming more conscious of the positive effects of diet on their well-being (Granato et al., 2020). Lycopene, a natural carotenoid, has garnered attention for its beneficial impacts on human health, such as reducing reactive oxygen species and lowering the risk of cardiovascular diseases (Li et al., 2021). Lycopene is a pigment found in fruits such as tomatoes, watermelons, and pink guavas, accumulating as these fruits ripen and giving them their distinctive red color (Guerra et al., 2021). In particular, cherry tomatoes are rich in lycopene, which makes lycopene content a key indicator for

evaluating internal quality. However, current chemical analysis methods, such as high-performance liquid chromatography (HPLC), are cumbersome, time-consuming, and unsuitable for large-scale, online detection (Hussain et al., 2019). Therefore, there is an unmet need to develop a rapid and non-destructive method for determining lycopene content for real-world applications.

Ultraviolet/visible/near-infrared (UV/Vis/NIR) spectroscopy is considered one of the most effective methods for online evaluation of fruit quality. Over the years, it has achieved promising results in quantitatively detecting fruit quality parameters (Cortés et al., 2019) and has demonstrated significant potential for determining lycopene

Abbreviations: BHT, 2,6-Di-tert-butyl-4-methylphenol; CQI, comprehensive quality index; GLCM, the gray-level co-occurrence matrix; HPLC, high-performance liquid chromatography; HSI, hyperspectral imaging; LYC, lycopene; LVs, latent variables; PLS, partial least squares; r, the correlation coefficient.; R_C^2 , the coefficient of determination for the calibration set; R_p^2 , the coefficient of determination for the prediction set; RMSEC, the root mean square error for the calibration set; RMSEP, the root mean square error for the prediction set; ROI, the region of interest; RPD, the ratio of the standard deviation of the prediction data to RMSEP; RS, Raman spectroscopy; SNR, the signal-to-noise ratio; SNV, standard normal variate; UV/Vis/NIR, ultraviolet/visible/near-infrared spectroscopy; R, red; G, green; B, blue; H, hue; S, saturation; V, value; L*, lightness; a*, the color position between green and red; b*, the color position between blue and yellow.

* 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).

<https://doi.org/10.1016/j.foodchem.2024.141183>

Received 17 July 2024; Received in revised form 15 August 2024; Accepted 5 September 2024

Available online 12 September 2024

0308-8146/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

content in cherry tomatoes (Li et al., 2024; Sheng et al., 2019). Effective UV/Vis/NIR information primarily arises from the absorption of stretching and bending vibrations of hydrogen-containing groups (X—H, such as C—H, O—H, and N—H) within molecules (Workman & Weyer, 2007). However, due to its low selectivity and the relatively low concentration of lycopene presented in food samples, its characteristic signals can be easily masked by water, soluble solids, and other components (Afara et al., 2021). Consequently, directly identifying distinctive lycopene absorption features in UV/Vis/NIR spectra is challenging. To further enhance the accuracy of UV/Vis/NIR-based online lycopene prediction, it is beneficial to integrate other methods that can provide multidimensional specific lycopene characteristics as a supplement (Song et al., 2024). Techniques such as Raman spectroscopy (RS) and fluorescence imaging have been employed for non-destructive lycopene detection. Although Raman spectroscopy can capture the Raman shifts around 1150 and 1510 cm^{-1} associated with lycopene (Hara et al., 2021), its current applicability for online detection is limited. Similarly, fluorescence imaging faces challenges due to its limited linear intensity and the low quantum yields of lycopene (Hussain et al., 2018; Konagaya et al., 2020).

Several studies have demonstrated the feasibility of non-destructive lycopene detection using color information (Goisser et al., 2020; Tilahun et al., 2018). This occurs because, as cherry tomatoes ripen, chlorophyll gradually degrades while lycopene accumulates, causing the fruit to change color from green to red and its surface to become more uniform (Carvalho et al., 2020). Since lycopene concentration is highest in the peel of cherry tomatoes (Strati & Oreopoulou, 2014), the resulting color change can be effectively monitored. For industrial agricultural product sorting systems, machine vision is typically used to capture images of samples to analyze surface color and texture features. This image-based information can therefore complement UV/Vis/NIR spectroscopy in lycopene detection (Zhang et al., 2018). Considering that color information can also be captured within the visible spectrum of UV/Vis/NIR, integrating image information with spectral data can be approached in two potential ways. Firstly, characteristic spectral color bands can be enhanced by indirectly utilizing color information. The feature enhancement relies on methods such as increasing resolution and averaging multiple samples (Mishra et al., 2020), which can potentially improve spectral quality. Similar approaches can enhance spectral color bands that are highly correlated with lycopene content, thereby improving the prediction accuracy. Secondly, surface color and texture features obtained from images can be directly fused with spectral data to create a more comprehensive dataset. This fusion also facilitates the development of robust models for predicting lycopene content. Although spectra-image fusion has been successfully applied to assess moisture content in tea (Sheng et al., 2023), nitrogen content in cotton

(Qin et al., 2024), and soil texture (Vakilzadeh Ebrahimi et al., 2023), there is no similar research for the online detection of lycopene content.

This study aims to enhance the prediction accuracy of lycopene content in cherry tomatoes by utilizing feature enhancement and spectra-image fusion, with image information supplementing spectra data. The specific objectives of this study are to (1) validate the feasibility of UV/Vis/NIR spectroscopy and machine vision for detecting lycopene content, (2) improve the performance of the spectra-based model through characteristic bands enhancement, and (3) integrate comprehensive data through spectra-image fusion to further enhance the prediction accuracy of lycopene content in cherry tomatoes.

2. Materials and methods

2.1. Samples

In this research, cherry tomato samples were harvested from a local orchard in Hangzhou, China. A total of 115 visually intact and uniform 'Zheyifeng' cherry tomatoes were used in the experiment (Fig. S1). According to the standard for tomato grades by the United States Department of Agriculture (2008), 'light red' and 'red' cherry tomatoes were selected as research samples to align with practical post-harvest sorting requirements. After destemming, washing, and numbering, these samples were stored at room temperature (25 °C) for 12 h to ensure uniform internal properties, such as temperature.

2.2. Experiments

2.2.1. Spectra collection

Transmittance and reflectance spectra of cherry tomatoes were collected using a self-built spectral acquisition system (Fig. 1A). The system includes an illumination device with a xenon lamp and its regulated power supply (ProSp-Xe300-UV, ProSp, China), a sample holder, a spectrometer for spectrum collection (Maya2000 Pro, Ocean Optics, USA), a dark box to mitigate external interference, and a computer for data collection and storage. Both the xenon lamp and the spectrometer cover a spectral range of 200–1100 nm, so the collected spectra also span this entire range. To ensure stable spectra acquisition, the light source and spectrometer were preheated for 30 min. The spectral integration time was 60 ms under transmission mode and 15 ms under reflection mode, with each spectrum being averaged 10 times. To minimize random errors and improve representativeness, five spectra were collected for each sample at different positions along the equator and averaged as the final spectrum (Fig. S2). All spectra were collected within the dark box and prepared for subsequent data analysis and model building.

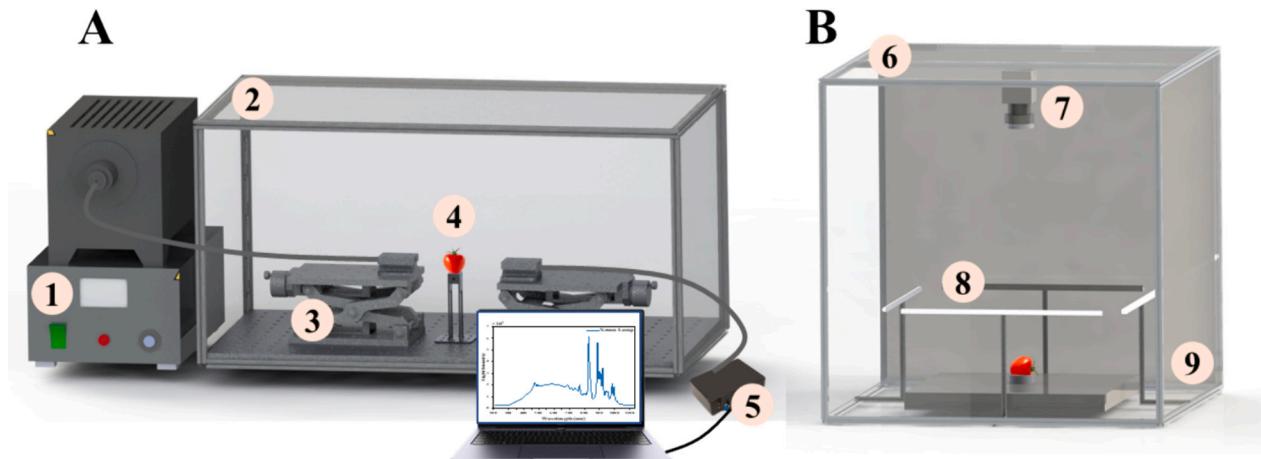


Fig. 1. Schematic diagram of information acquisition systems: (A) Spectral acquisition system (1. Xenon lamp, 2. Dark box, 3. Optical platform, 4. Cherry tomato sample, 5. Spectrometer), and (B) images acquisition system (6. Dark box, 7. Camera, 8. Light source, 9. Reflective fabric).

2.2.2. Image acquisition

Fig. 1B illustrates the RGB image acquisition setup, which consists of an industrial-grade camera (Dahua, A7200/CG30, 1920 × 1200 pixels) with a 16 mm focal length lens (Dahua, MH1220M) for high-resolution image capture, four LED light tubes (OPPLE, MX575C-D8Z) with corresponding reflective fabric, a black sample tray and a computer for images collection and storage. To ensure stable image acquisition, the camera and light source were preheated for 30 min. To achieve uniform illumination, four light tubes were positioned at 45° angles, directing light to reflect off the fabric, thereby providing consistent illumination for the imaging system (Gao et al., 2024). The sample tray was positioned in the center beneath the camera with a black background to aid in subsequent image processing. Each cherry tomato was imaged five times around its equator plane to ensure comprehensive coverage of all sides. After image acquisition, the cherry tomatoes were immediately frozen using liquid nitrogen and stored in a -80 °C refrigerator to prevent any changes in lycopene content.

2.2.3. Lycopene content measurement

The lycopene content was determined following the method described by Li et al. (2024), with some modifications. After thawing, approximately 3.5 g of cherry tomato samples were weighed for analysis. The extraction solvent was a 1:1 (v/v) mixture of acetone and petroleum ether (SCR, China), with 0.1 % (m/v) 2,6-Di-tert-butyl-4-methylphenol (BHT, Macklin, China) added as an antioxidant to prevent lycopene oxidation. The weighed sample and 100 mL of the solvent were placed in a high-speed homogenizer (H6, Hehui, China) and stirred at 31,000 rpm for 3 min until the tomato pulp turned white. Subsequently, the extract was filtered through a glass funnel with filter paper and transferred to a 100 mL conical flask with a sealed cap. The lycopene extract was then reconstituted to a final volume of 100 mL and mixed thoroughly. The absorbance of the extract was measured using a UV–Vis spectrophotometer (Evolution 220, Thermo Fisher Scientific, USA), and

the lycopene content (mg/kg) in cherry tomatoes was calculated using a standard curve, which was prepared as follows. Firstly, 5 mg of lycopene (Aladdin, China) was dissolved in 100 mL extraction solvent to prepare a 50 mg/L stock solution. Then, working solutions with concentrations of 1, 2, 3, 4, and 5 mg/L were prepared by diluting the stock solution. The absorbance of each solution at 504 nm was measured using the UV–Vis spectrophotometer, and the absorbance-concentration standard curve was constructed (Popescu et al., 2022). To ensure precision and accuracy, the preparation of each lycopene standard solution and the detection of lycopene content in each sample were performed in triplicate.

2.3. Preprocessing of spectra and image data

2.3.1. Spectra preprocessing and combination

During spectral acquisition, factors such as light scattering, sample size, temperature variations, and instrument instability can introduce artifacts like random noise, baseline shift, and multiplicative effects. Spectral preprocessing helps remove irrelevant variables or effects, thereby retaining useful information for modeling (Barra et al., 2021). Spectral preprocessing involved two main steps: smoothing and standard normal variate (SNV). The spectra of cherry tomatoes were automatically smoothed during collection to improve the signal-to-noise ratio (SNR). Within the spectra suite, the spectral acquisition settings included an averaging number of 10 and a smoothing parameter of 3. Then, transmittance and reflectance spectra were subjected to SNV processing separately to reduce uncontrolled variations caused by path length and light scattering. SNV involved standardizing each data point by calculating the mean and standard deviation at each wavelength. After SNV preprocessing, the spectral data was recentered to follow a normal distribution (**Fig. 2A~D**).

Due to the strong absorption of the samples and the longer path length under transmission mode, the transmittance spectra within the

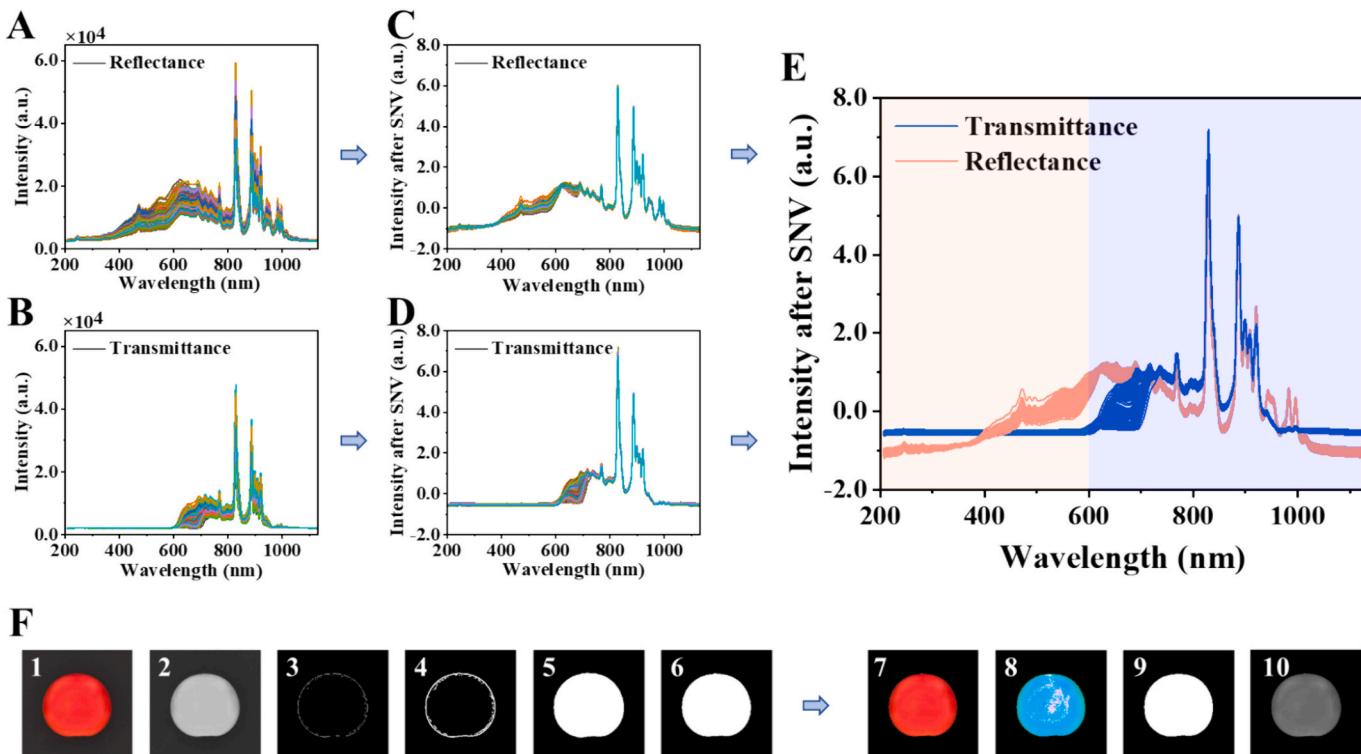


Fig. 2. Workflow of spectra and image processing: (A) Raw reflectance spectra, (B) raw transmittance spectra, (C) SNV-preprocessed reflectance spectra, (D) SNV-preprocessed transmittance spectra, (E) spectral combination of reflectance (200–600 nm) and transmittance (600–1100 nm), and (F) steps for extracting image color and texture features (1. original image, 2. grayscale, 3. binary gradient mask, 4. dilated gradient mask, 5. internal filling, 6. boundary smoothing, 7. RGB image, 8. HSV image, 9. L*a*b* image, 10. grayscale image).

200–600 nm range typically contain less useful information. In contrast, the spectral information within 200–600 nm can be obtained under reflection mode, where light interacts only with the shallow tissues of the sample and experiences less absorption. However, the reflection spectra may lack representativeness of the sample's overall information. To acquire complete spectra across 200–1100 nm and ensure representativeness, the reflectance and transmittance spectra were combined to predict lycopene content (Zheng et al., 2024). After separately preprocessing the transmittance and reflectance spectra, the reflectance spectra (200–600 nm) and transmittance spectra (600–1100 nm) were concatenated to form a new combined spectra dataset. This approach leveraged the advantageous bands from both single-mode spectra. The steps for spectral preprocessing and combination are depicted in Fig. 2A~E.

2.3.2. Image color and texture features extraction

In this study, image information was extracted through three main steps: cherry tomato region segmentation, image format conversion, and feature extraction. For cherry tomato region segmentation, the Sobel operator was used to accurately delineate the sample boundaries. Firstly, the acquired color images were converted to grayscale using weights of 0.7, 0.2, and 0.1 for the R, G, and B channels, respectively, to enhance contrast with the dark background. Next, a binary gradient mask was generated using the Sobel operator, and the detected boundaries were refined by edge dilation to achieve a continuous outline. Finally, the region of interest (ROI) was obtained by internal filling and boundary smoothing (Fig. 2F, 1–6).

The R (red), G (green), and B (blue) values were directly extracted from the cherry tomato color images. HSV and CIELAB images were derived from the RGB images, and the corresponding color variables were calculated, including H (hue), S (saturation), V (value), L* (lightness), a* (the color position between green and red), and b* (the color position between blue and yellow). Additionally, the gray-level co-occurrence matrix (GLCM) was employed to extract four texture features (contrast, correlation, energy, and homogeneity) from the grayscale images (Wu et al., 2023). In this way, a total of 9 color metrics and 4 texture features were utilized to characterize the image information. To reduce noise and enhance representativeness, surface color and texture features for each cherry tomato were averaged over five images. The image information extraction process described above is illustrated in Fig. 2F.

2.4. Spectra enhancement and spectra-image fusion

2.4.1. Spectra enhancement

Spectra enhancement is based on improving spectral quality. In general, improving spectral resolution enhances accurate identification and differentiation of subtle features in the spectrum (Fan et al., 2020). Multiple sampling and averaging can reduce spectral noise and enhance representativeness. A similar approach was applied to spectral enhancement in this study (Fig. 5A). Firstly, spectral bands corresponding to colors highly correlated with lycopene were selected as characteristic wavelengths. Subsequently, N spectral points within the characteristic bands were averaged to generate new spectral values. These newly obtained spectral points were inserted back into the original spectrum at their respective N-point positions to enhance the overall spectral quality. Finally, the prediction model results of enhanced spectra based on different step widths N were compared to determine the optimal value of N.

2.4.2. Spectra-image fusion

Data fusion can integrate multiple data sources from different perspectives to characterize the physical and chemical properties of

samples, providing comprehensive information for model building. Generally, data fusion is classified into three levels based on the type of data: low-level fusion (raw-data level), mid-level fusion (feature level), and high-level fusion (decision level) (Zhou et al., 2020). This study adopts a low-level fusion strategy to fully utilize the raw data by integrating spectral data and image information. Specifically, after normalizing the spectral data and image features, they were concatenated to create a new data matrix for lycopene prediction. In addition, both the original and enhanced spectra were fused with the image information to compare their prediction accuracy.

2.5. Chemometric methods

2.5.1. Partial least squares regression

Partial least squares (PLS) regression, a multivariate statistical analysis technique used to characterize the quantitative relationship between predictor and response variables, has been successfully applied to develop NIR models in many studies. In this study, PLS regression was used to develop the lycopene content prediction model. Before modeling, the dataset was divided into calibration and prediction sets at a 2:1 ratio. To ensure the representativeness of the calibration and prediction set, the descending-order selection method was used to partition the dataset. The lycopene content in the calibration set ranged from 27.7 to 183.4 mg/kg, while in the prediction set it ranged from 32.8 to 172.5 mg/kg, providing a solid foundation for constructing a robust prediction model. Moreover, the optimal number of latent variables (LVs) was determined through 10-fold cross-validation, with a maximum limit of 10 LVs to prevent overfitting (Yun et al., 2019).

2.5.2. Model evaluation

The model's performance is evaluated using the following metrics: the coefficient of determination for the calibration (R_C^2) and prediction (R_p^2) set, the root mean square error for the calibration (RMSEC) and prediction (RMSEP) set, and the ratio of the standard deviation of the prediction data to RMSEP (RPD). A favorable model exhibits R_C^2 and R_p^2 values closer to 1, higher RPD values, and lower RMSEC and RMSEP values. Generally, a model is considered to perform well when R^2 is greater than 0.8, the RMSEP/RMSEC ratio falls between 0.8 and 1.2, and the RPD value exceeds 2 (Li et al., 2023). These metrics are calculated using the following formulas:

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

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

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

Here, n is the number of samples, y_i is the reference lycopene content of the i^{th} sample, \bar{y} is the mean value of reference lycopene content and \hat{y}_i is the lycopene content of i^{th} sample predicted by the model. SD refers to the standard deviation of the reference values in the prediction set.

All spectra were collected using spectra suite (Ocean Optics, USA) and images were obtained using MVS (V4.3.2, Hikvision, China). Significance analysis (one-way analysis of variance, ANOVA) was conducted using SPSS (ver.25.0, IBM, USA). Spectral processing, image feature extraction, and model development were performed using MATLAB (R2021b, MathWorks, USA).

3. Results and discussion

3.1. Prediction results of UV/Vis/NIR spectra

Fig. 3A shows the transmittance and reflectance spectra of cherry tomatoes under a xenon lamp. The intensity of the transmittance spectra is lower than that of the reflectance spectra, primarily due to the longer optical path length and the interaction between light and the fruit tissue, even though the integration time in transmission mode is longer. The reduced intensity values around 680 nm and 980 nm in the transmittance spectra correspond to chlorophyll and water absorption, respectively (Fig. S3) (Song et al., 2021). However, spectral information within the 200–600 nm range is absent due to the strong absorption by cherry tomatoes. The inset in **Fig. 3A** shows lycopene absorption in the acetone–petroleum ether solution, with three absorption peaks at 446, 472, and 504 nm, suggesting that some characteristic lycopene absorption may be missing in the transmittance spectra. In contrast, reflectance spectra provide relatively complete spectral information across the 200–1100 nm range, with an absorption peak near 480 nm possibly associated with lycopene (Fig. S3). However, reflectance spectra generally offer a less comprehensive representation of overall fruit quality (Hong & Chia, 2021). Thus, the combination of reflectance spectra (200–600 nm) and transmittance spectra (600–1100 nm) can form a more comprehensive spectral dataset that covers the UV/Vis/NIR regions, providing a more representative analysis simultaneously.

The distribution histogram of lycopene content in cherry tomatoes is shown in **Fig. 3B**. The lycopene content of samples ranges from 28.44 to 183.36 mg/kg, with a mean of 94.7 and a standard deviation of 38.9 mg/kg. The wide range and distribution of lycopene content indicate that the selected samples are representative and suitable for model development.

Table 1 presents the prediction results of lycopene content based on

Table 1

Lycopene prediction results of reflectance, transmittance and their combined spectra.

Spectra	R _C ²	RMSEC (mg/kg)	R _P ²	RMSEP (mg/kg)	RPD
Reflectance	0.88	13.29	0.86	14.44	2.64
Transmittance	0.90	12.22	0.89	12.59	3.02
Combination¹	0.91	11.57	0.91	11.60	3.28

¹ The combination of reflectance (200–600 nm) and transmittance (600–1100 nm) spectra.

spectra from different detection modes. Predictions based on transmittance spectra outperform those from reflectance spectra, with the best performance achieved by combining complementary spectral bands. **Fig. 3C** visually compares these prediction results using a radar chart. Firstly, the R_P², RMSEP and RPD values for lycopene prediction based on reflectance spectra are 0.86, 14.44 mg/kg, and 2.64, while for transmittance spectra these values are 0.89, 12.59 mg/kg, and 3.02, respectively. The better performance of the transmittance model is attributed to photons passing through the entire sample in transmission mode, resulting in spectra that are more representative of the whole sample. In contrast, in reflection mode, light interacts only with the superficial tissue and is more susceptible to specular reflection, scattering, and other interference (Magwaza et al., 2012), leading to lower spectral quality (Fig. S2). Secondly, the R_P², RMSEP, and RPD values for the lycopene prediction model based on the combined spectra improved to 0.91, 11.60 mg/kg, and 3.28, respectively (**Fig. 3D**). Combining complementary bands utilizes the spectral regions in the reflectance spectra associated with characteristic lycopene absorption and the transmittance spectrum which is more representative of the entire fruit. This approach ensures both spectral quality and representativeness, while covering the entire UV/Vis/NIR range, effectively enhancing the accuracy of the lycopene prediction model. However, there remains

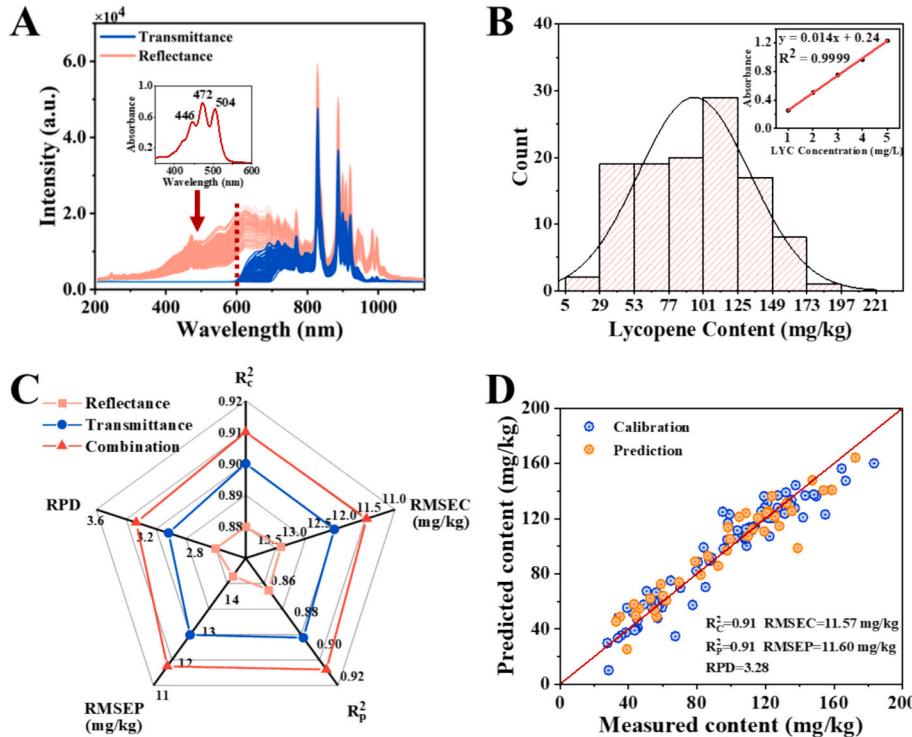


Fig. 3. UV/Vis/NIR spectra and the results of spectral prediction models: (A) The reflectance and transmittance spectra of cherry tomatoes under xenon lamp, with the embedded absorption spectrum of lycopene in the acetone–petroleum ether solution, (B) the distribution histogram of lycopene (LYC) content in cherry tomatoes, with the embedded standard curve of absorbance versus concentration for the lycopene solution, (C) the radar chart of model results under spectra from different detection modes, and (D) the prediction result of the combined spectra from reflectance (200–600 nm) and transmittance (600–1100 nm) spectra. (Detailed visualization of the embedded figures is shown in Fig. S4.)

potential for further improvement in the combined spectra prediction model by enhancing spectra or incorporating additional information related to lycopene.

3.2. Correlation between lycopene content and image features

To assess how variations in lycopene content affect surface color, visual color analyses were performed. Cherry tomatoes were categorized into low, medium, and high lycopene content groups. Representative RGB images and average RGB values for each group are displayed in Fig. 4A. As lycopene content increases, the color shifts from light orange to deep red, with a more uniform color distribution (Carvalho et al., 2020). Correspondingly, in the RGB images, the R-value remains high and gradually decreases, the G-value decreases significantly, while the B-value remains relatively stable.

Based on the above differences, the Pearson correlation coefficients (r) between these image features and lycopene content were calculated to analyze their correlations (Fig. 4B). The indices B, S, energy, and homogeneity show low correlation coefficient ($r < 0.3$), while all other indices exhibit significant correlations with lycopene content ($p < 0.01$). Since lycopene is most concentrated in the peel (Strati & Oreopoulou, 2014), its content directly affects the surface color of cherry tomatoes, resulting in a stronger correlation with the color indices. As image texture features are related to surface color distribution, texture indices could indirectly reflect lycopene content, resulting in relatively lower correlations.

Among these color indices, R, G, L*, and b* exhibit high correlations with lycopene content, as shown in the linear regression analysis in Fig. 4C. The negative correlations between lycopene content and R, G, and b* values can be attributed to carotenoid accumulation (primarily lycopene) and chlorophyll degradation, which cause the fruit color to shift from light orange to deep red. Lycopene accumulation also leads to a darker color and lower brightness, resulting in negative correlations between L* and V with lycopene content. These changes are further supported by the interrelationships among the color indices. For example, the strong positive correlations between R and V, G and L*, and b* and L* suggest the gradual deepening of red color and a reduction in

Table 2

Comparison of lycopene content prediction results based on UV/Vis/NIR spectroscopy and image features.

Data source	R _C ²	RMSEC (mg/kg)	R _P ²	RMSEP (mg/kg)	RPD
Spectral combination	0.91	11.57	0.91	11.60	3.28
Image features	0.88	13.34	0.87	14.00	2.72

brightness. Regarding texture features, higher lycopene content makes the surface of cherry tomatoes more uniform and less variable, which reduces contrast and correlation (Fig. S5) (Khojastehnazhand & Ramezani, 2020). Therefore, these findings indicate that color and texture features can be effectively used to predict lycopene content.

A PLS model was developed based on the color indices and texture features (excluding B, S, and homogeneity) to further evaluate the lycopene prediction performance. As shown in Table 2, the model achieved R_P², RMSEP, and RPD values of 0.87, 14.00 mg/kg, and 2.72, respectively. This indicates that the model has a reasonable prediction capability although it performs slightly worse than the spectral combination. Thus, color and texture information can serve as supplementary data to spectral data, to further improve lycopene prediction accuracy. This supplementation could be achieved by enhancing spectral bands related to color or through spectra-image fusion.

3.3. Spectral enhancement

One of the methods to utilize lycopene-related information is to enhance spectral bands associated with colors, especially red and green (Song et al., 2021). This color information is mainly reflected in the visible region of the UV/Vis/NIR spectrum. Specifically, the wavelength range for green light is around 495–570 nm, and the range for red light is approximately 620–750 nm (Wong et al., 2020). Due to the spectral overlap and similarities among adjacent bands in the visible ranges, spectra within the ranges of 490–560 nm and 610–760 nm were enhanced (Fig. 5B). Each original spectrum contains a total of 1900

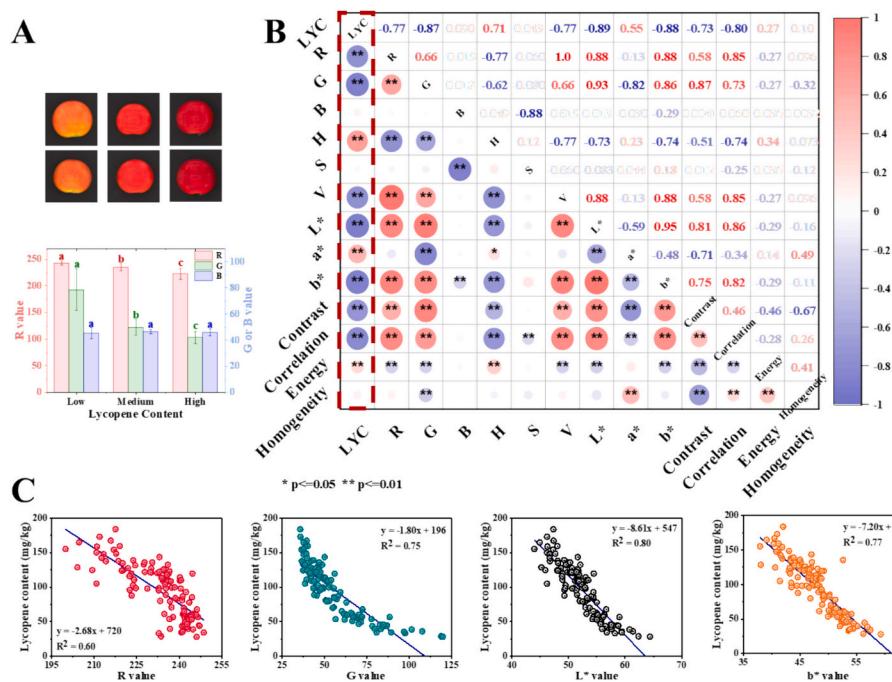


Fig. 4. The relationship between lycopene content and image features (color and texture): (A) Typical RGB images of cherry tomatoes with different lycopene contents and their average RGB values, (B) correlation analysis plots between image features (color, texture) and lycopene content (LYC), and (C) linear fitting results between color indices (R, G, L*, b*) and lycopene content of cherry tomatoes.

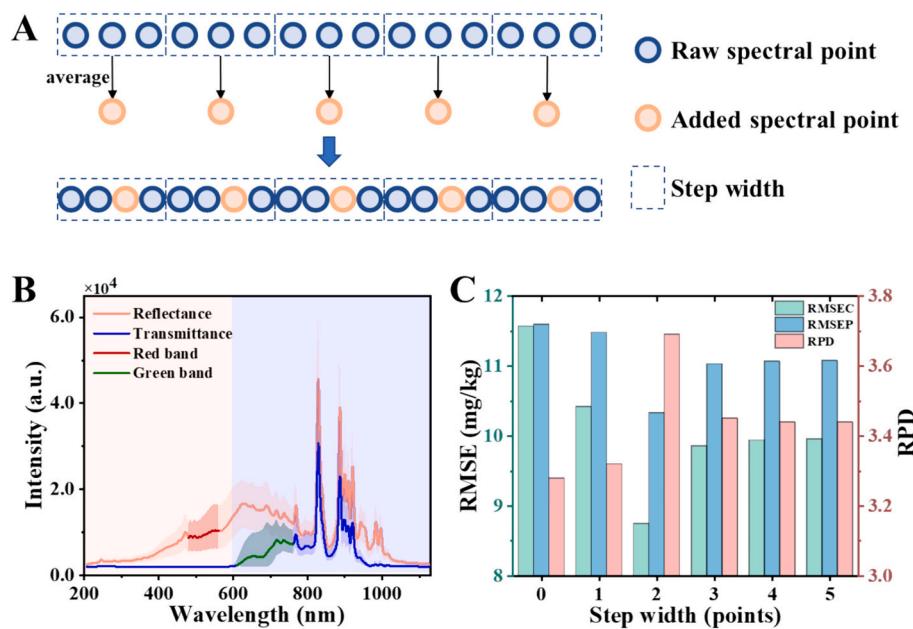


Fig. 5. Spectral feature enhancement: (A) Schematic diagram of spectral band enhancement, (B) enhanced spectral bands corresponding to colors related to lycopene content, and (C) comparison of prediction results for lycopene content based on raw (step width as 0) and enhanced spectra at different step widths.

spectral points, with 335 in the green band and 487 in the red band. An overview of the spectral enhancement method is shown in Fig. 5A. New spectral points are generated by averaging several original points within the target regions and then inserted into the original spectrum to create the enhanced spectrum. This process enhances spectral bands theoretically related to lycopene, which is expected to improve prediction accuracy.

The effect of different averaging point numbers on spectral enhancement was investigated. Fig. 5C compares the lycopene prediction results for PLS models established before and after spectral enhancement. Compared to the original combined spectra (denoted as 0 in Fig. 5C), the prediction performance improves when the averaging step width is 2 or greater. However, no significant change is observed with a step width of 1. When the step width is 1, each original spectral point is merely duplicated without introducing new information, resulting in no effect on lycopene prediction performance. At a step width of 2, the prediction performance shows the most significant improvement, with the R_p^2 , RMSEP, and RPD values of 0.93, 10.33 mg/kg, and 3.69, respectively. The RMSEP value decreases by 10.95 % compared to the raw spectral combination, indicating a significant improvement. For step widths of 3, 4, and 5, the corresponding RMSEP values are 11.03, 11.07, and 11.08 mg/kg, respectively, showing less improvement to a step width of 2. The reduced effectiveness when averaging more than 2 spectral points may be due to the limited supplements to the raw spectral details.

As a spectral processing method, characteristic band enhancement improves spectral quality by adding averaged values of spectral points to the raw spectrum. Averaging spectral points reduces random noise and enhances SNR. In addition, inserting new spectral points increases resolution and helps smooth the spectrum (Mishra et al., 2020). As a result, the spectral quality of the green- and red-light bands related to lycopene is enhanced, leading to improved prediction performance for lycopene content. From a feature information utilization perspective, spectral band enhancement is an indirect method of using color information, though it has shown improved prediction performance for lycopene. To further leverage the image information, the direct fusion of spectral and

image data could be considered.

3.4. Spectra-image fusion

To directly utilize image information, spectral and image data (surface color and texture features) were concatenated into a single matrix. Before concatenation, the spectral and image data were separately normalized to mitigate the impact of scale differences (Fig. 6A) (Azcarate et al., 2021). Both raw and enhanced spectra were fused with image information, and the prediction results for lycopene content are presented in Table 3. Compared to models based solely on original and enhanced spectra, those using the fusion strategy achieved the best prediction performance (Fig. 6B). The R_p^2 , RMSEP, and RPD values of the PLS model based on the raw spectra and image fusion are 0.95, 8.96 mg/kg, and 4.25, respectively. The RMSEP value decreased by 13.3 %, and the RPD value increased by 15.2 % compared with the model results based on enhanced spectra. When compared to the prediction results using combined spectra, RMSEP decreased by 22.8 % and the RPD increased by 29.6 %. These improvements demonstrate the effectiveness of spectra-image data fusion in enhancing the prediction accuracy for lycopene content.

It is noteworthy that the fusion of spectral and image information achieved very similar prediction results both before and after spectral enhancement. This similarity may be due to the overlap between the enhanced spectral bands and the color information contained in the images. Since surface color and texture features can intuitively reflect the lycopene content in cherry tomatoes, the added value of indirect spectral enhancement may be limited. As shown in Fig. 6A, the PLS model exhibits relatively large regression coefficients near 550 nm (corresponding to the red-light band), 630, and 680 nm (green-light band) (Li et al., 2019). Additionally, surface color indices, especially R, G, H, and b*, also show large regression coefficients. Consequently, the enhancement of spectral characteristic bands has a minor impact on the prediction performance of the spectra-image fusion model, likely due to the overlap between the spectral and image information.

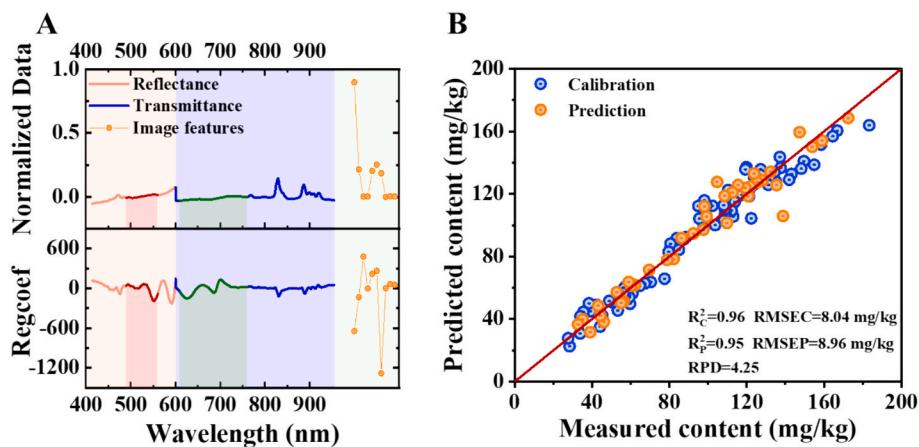


Fig. 6. Spectra-image fusion: (A) Schematic diagram of the concatenated normalized spectral data and image features (from left to right: R, G, H, V, L*, a*, b*, contrast, correlation, energy) and the regression coefficients of the PLS model, (B) the prediction results of the fusion of raw spectra-image features.

Table 3
Prediction results for lycopene before and after spectra-image fusion.

Data source	R^2_C	RMSEC (mg/kg)	R^2_P	RMSEP (mg/kg)	RPD
Spectral Combination	0.91	11.57	0.91	11.60	3.28
Image features	0.88	13.34	0.87	14.00	2.72
Spectra enhancement	0.95	8.75	0.93	10.33	3.69
Raw spectra-image	0.96	8.04	0.95	8.96	4.25
Enhanced spectra-image	0.96	7.98	0.95	8.99	4.24

3.5. Discussion

In this study, the prediction accuracy of lycopene content was improved by spectral feature enhancement and spectra-image fusion. Although the feature enhancement model did not perform as well as the direct spectra-image fusion approach (Table 3), it still holds intrinsic value. For instance, in scenarios where direct acquisition of image or color information is not feasible, such as with micro-portable devices (Egei et al., 2022), enhancing characteristic bands could improve prediction performance. Additionally, spectral enhancement can inform hardware design for detection systems (Lee et al., 2023), such as the development of light sources that target specific bands, thereby directly improving the quality of acquired spectra.

From the perspective of characteristic information supplementation, integrating image information related to lycopene with UV/Vis/NIR spectroscopy can further enhance prediction accuracy. Beyond surface color, other features such as RS, can also capture lycopene-specific information (Hara et al., 2021). In our previous research, RS was attempted as supplementary information to UV/Vis/NIR spectra, but the results were unsatisfactory (Fig. S6). This may be due to the point spectral characteristics of RS, which only captures partial surface information and may not be representative of the entire cherry tomato (Fu et al., 2016). Moreover, although RS is a non-destructive testing technique, it is currently unsuitable for online fruit quality detection.

Many conventional and novel methods have been applied to lycopene detection (Hussain et al., 2019). Cucu et al. (2012) confirmed a limit of quantitation (LOQ) of 60 ng/g for lycopene using HPLC, demonstrating high precision and accuracy, and successfully applied it to the lycopene content detection in various food products. However, the complex preprocessing and time-consuming nature of HPLC makes it unsuitable for large-scale, nondestructive detection of fruit quality. Hara et al. (2021) developed a quantitative prediction model for carotenoids in tomatoes using 785 nm-excited Raman spectra. However, weak Raman signals and the risk of thermal damage from laser pose

challenges for its application. Shao et al. (2022) employed hyperspectral imaging (HSI) to predict lycopene and other parameters in tomatoes and evaluated overall quality using the comprehensive quality index (CQI). Even though, the time-consuming data acquisition and redundant information associated with HSI limit its suitability for large-scale detection. Wang et al. (2018) used a ring light source to obtain transmittance Vis/NIR spectra of tomatoes and established a lycopene prediction model using a stepwise optimized SNV-UVE-CARS-PLS method. While this approach yielded high prediction accuracy, with R_p and RMSEP values of 0.9812 and 0.7071 mg/kg, respectively, the spectral range only covered 650–1100 nm, potentially missing the characteristic absorption of lycopene around 450–510 nm. Furthermore, these studies (Goisser et al., 2020; Li et al., 2024; Sheng et al., 2019) typically utilized a single technique (e.g., UV/Vis/NIR, RI, or colorimetry), resulting in limited characteristic information. This research integrates UV/Vis/NIR spectroscopy and machine vision to capture comprehensive internal and surface information on cherry tomatoes. Considering that commercial sorting lines typically use machine vision to detect surface defects and UV/Vis/NIR spectroscopy to assess internal quality simultaneously, this study provides valuable reference values for improving online detection systems.

4. Conclusion

This study demonstrates that supplementing spectral data with image information can significantly improve the prediction accuracy of lycopene content in cherry tomatoes through feature enhancement and data fusion. The combination of reflectance (200–600 nm) and transmittance (600–1100 nm) spectra provides comprehensive coverage across UV/Vis/NIR ranges, thus outperforming single-mode spectroscopy approaches for lycopene detection. Surface color and texture features, which exhibit a high correlation with lycopene content, can be effectively utilized for lycopene assessment. Moreover, indirectly leveraging color information to enhance spectral bands associated with lycopene, namely red (490–560 nm) and green (610–760 nm), improves the prediction performance of lycopene compared to the original spectral model. Ultimately, the prediction accuracy was further enhanced by direct spectra-image fusion, yielding R^2 , RMSEP, and RPD values of 0.95, 8.96 mg/kg, and 4.25, respectively. These findings underscore the potential of integrating spectral and image data to refine lycopene assessment by feature enhancement and data fusion, which aligns with the simultaneous use of spectroscopy and machine vision in some commercial sorting lines. Future research could focus on accelerating large-scale chemical testing and optimizing multi-source data fusion methods to better meet industrial demands.

CRediT authorship contribution statement

Yuanhao Zheng: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xuan Luo:** Writing – review & editing, Resources, Methodology, Conceptualization. **Yuan Gao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhizhong Sun:** Writing – review & editing, Validation, Conceptualization. **Kang Huang:** Writing – review & editing. **Weilu Gao:** Writing – review & editing. **Huirong Xu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Lijuan Xie:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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

This work is supported by the National Key R&D Program of China [2022YFD002204] and the National Key Laboratory of Agricultural Equipment Technology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2024.141183>.

References

- Afara, I. O., Shaikh, R., Nippolainen, E., Querido, W., Torniaisen, J., Sarin, J. K., ... Töyräs, J. (2021). Characterization of connective tissues using near-infrared spectroscopy and imaging. *Nature Protocols*, 16(2), 1297–1329. <https://doi.org/10.1038/s41596-020-00468-z>
- Azcarate, S. M., Ríos-Reina, R., Amigo, J. M., & Goicoechea, H. C. (2021). Data handling in data fusion: Methodologies and applications. *TrAC Trends in Analytical Chemistry*, 143, Article 116355. <https://doi.org/10.1016/j.trac.2021.116355>
- Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances – a review. *TrAC Trends in Analytical Chemistry*, 135, Article 116166. <https://doi.org/10.1016/j.trac.2020.116166>
- Carvalho, G. C., Sábio, R. M., & Chorilli, M. (2020). An overview of properties and analytical methods for lycopene in organic nanocarriers. *Critical Reviews in Analytical Chemistry*, 1–13. <https://doi.org/10.1080/10408347.2020.1763774>
- 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 in Food Science & Technology*, 85, 138–148. <https://doi.org/10.1016/j.tifs.2019.01.015>
- Cucu, T., Huvaere, K., Van Den Berghe, M.-A., Vinkx, C., & Van Loco, J. (2012). A simple and fast HPLC method to determine lycopene in foods. *Food Analytical Methods*, 5(5), 1221–1228. <https://doi.org/10.1007/s12161-011-9354-6>
- Égei, M., Takács, S., Palotás, G., Palotás, G., Szuvandzsiev, P., Daood, H. G., ... Pék, Z. (2022). Prediction of soluble solids and lycopene content of processing tomato cultivars by Vis-NIR spectroscopy. *Frontiers in Nutrition*, 9, Article 845317. <https://doi.org/10.3389/fnut.2022.845317>
- Fan, S., Wang, Q., Tian, X., Yang, G., Xia, Y., Li, J., & Huang, W. (2020). Non-destructive evaluation of soluble solids content of apples using a developed portable Vis/NIR device. *Biosystems Engineering*, 193, 138–148. <https://doi.org/10.1016/j.biosystemseng.2020.02.017>
- Fu, X., He, X., Xu, H., & Ying, Y. (2016). Nondestructive and rapid assessment of intact tomato freshness and lycopene content based on a miniaturized Raman spectroscopic system and colorimetry. *Food Analytical Methods*, 9(9), 2501–2508. <https://doi.org/10.1007/s12161-016-0440-7>
- Gao, Y., Wang, Q., Rao, X., Xie, L., & Ying, Y. (2024). OrangeStereo: A novel orange stereo matching network for 3D surface reconstruction. *Computers and Electronics in Agriculture*, 217, Article 108626. <https://doi.org/10.1016/j.compag.2024.108626>
- Goisser, S., Wittmann, S., Fernandes, M., Mempel, H., & Ulrichs, C. (2020). Comparison of colorimeter and different portable food-scanners for non-destructive prediction of lycopene content in tomato fruit. *Postharvest Biology and Technology*, 167, Article 111232. <https://doi.org/10.1016/j.postharvbio.2020.111232>
- Granato, D., Barba, F. J., Kováčević, D. B., Lorenzo, J. M., Cruz, A. G., & Putnik, P. (2020). Functional foods: Product development, technological trends, efficacy testing, and safety. *Annual Review of Food Science and Technology*, 11, 93–118. <https://doi.org/10.1146/annurev-food-032519-051708>
- Guerra, A. S., Hoyos, C. G., Molina-Ramírez, C., Velásquez-Cock, J., Vélez, L., Gañán, P., ... Zuluaga, R. (2021). Extraction and preservation of lycopene: A review of the advancements offered by the value chain of nanotechnology. *Trends in Food Science & Technology*, 116, 1120–1140. <https://doi.org/10.1016/j.tifs.2021.09.009>
- Hara, A., Ishigaki, M., Ozaki, Y., Ahamed, T., Noguchi, R., Miyamoto, A., & Genkawa, T. (2021). Effect of Raman exposure time on the quantitative and discriminant analyses of carotenoid concentrations in intact tomatoes. *Food Chemistry*, 360, Article 129896. <https://doi.org/10.1016/j.foodchem.2021.129896>
- Hong, F. W., & Chia, K. S. (2021). A review on recent near infrared spectroscopic measurement setups and their challenges. *Measurement*, 171, Article 108732. <https://doi.org/10.1016/j.measurement.2020.108732>
- Hussain, A., Pu, H., & Sun, D.-W. (2018). Innovative nondestructive imaging techniques for ripening and maturity of fruits – A review of recent applications. *Trends in Food Science & Technology*, 72, 144–152. <https://doi.org/10.1016/j.tifs.2017.12.010>
- Hussain, A., Pu, H., & Sun, D.-W. (2019). Measurements of lycopene contents in fruit: A review of recent developments in conventional and novel techniques. *Critical Reviews in Food Science and Nutrition*, 59(5), 758–769. <https://doi.org/10.1080/10408398.2018.1518896>
- Khajastehnazhand, M., & Ramezani, H. (2020). Machine vision system for classification of bulk raisins using texture features. *Journal of Food Engineering*, 271, Article 109864. <https://doi.org/10.1016/j.jfooodeng.2019.109864>
- Konagaya, K., Al Riza, D. F., Nie, S., Yoneda, M., Hirata, T., Takahashi, N., ... Kondo, N. (2020). Monitoring mature tomato (red stage) quality during storage using ultraviolet-induced visible fluorescence image. *Postharvest Biology and Technology*, 160, Article 111031. <https://doi.org/10.1016/j.postharvbio.2019.111031>
- Lee, H., Cho, S., Lim, J., Lee, A., Kim, G., Song, D.-J., ... Mo, C. (2023). Performance comparison of tungsten-halogen light and phosphor-converted NIR LED in soluble solid content estimation of apple. *Sensors*, 23(4), Article 4. <https://doi.org/10.3390/s23041961>
- Li, M., Shi, Y., Zhang, J., Wan, X., Fang, J., Wu, Y., ... Bian, Z. (2023). Rapid evaluation of Ziziphi Spinosae semen and its adulterants based on the combination of FT-NIR and multivariate algorithms. *Food Chemistry*, 20, Article 101022. <https://doi.org/10.1016/j.foodchx.2023.101022>
- Li, N., Wu, X., Zhuang, W., Xia, L., Chen, Y., Wu, C., Rao, Z., Du, L., Zhao, R., Yi, M., Wan, Q., & Zhou, Y. (2021). Tomato and lycopene and multiple health outcomes: Umbrella review. *Food Chemistry*, 343, Article 128396. <https://doi.org/10.1016/j.foodchem.2020.128396>
- Li, S., Wang, Q., Yang, X., Zhang, Q., Shi, R., & Li, J. (2024). Online detection of lycopene content in the two cultivars of tomatoes by multi-point full transmission Vis-NIR spectroscopy. *Postharvest Biology and Technology*, 211, Article 112813. <https://doi.org/10.1016/j.postharvbio.2024.112813>
- Li, X., Jin, J., Sun, C., Ye, D., & Liu, Y. (2019). Simultaneous determination of six main types of lipid-soluble pigments in green tea by visible and near-infrared spectroscopy. *Food Chemistry*, 270, 236–242. <https://doi.org/10.1016/j.foodchem.2018.07.039>
- 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 and Bioprocess Technology*, 5(2), 425–444. <https://doi.org/10.1007/s11947-011-0697-1>
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, 132, Article 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- Popescu, M., Iancu, P., Pleșu, V., Bîldean, C. S., & Todasca, C. M. (2022). Different spectrophotometric methods for simultaneous quantification of lycopene and β-carotene from a binary mixture. *LWT*, 160, Article 113238. <https://doi.org/10.1016/j.lwt.2022.113238>
- Qin, S., Ding, Y., Zhou, T., Zhai, M., Zhang, Z., Fan, M., Lv, X., Zhang, Z., & Zhang, L. (2024). “Image-Spectral” fusion monitoring of small cotton samples nitrogen content based on improved deep forest. *Computers and Electronics in Agriculture*, 221, Article 109002. <https://doi.org/10.1016/j.compag.2024.109002>
- Shao, Y., Shi, Y., Qin, Y., Xuan, G., Li, J., Li, Q., Yang, F., & Hu, Z. (2022). A new quantitative index for the assessment of tomato quality using Vis-NIR hyperspectral imaging. *Food Chemistry*, 386, Article 132864. <https://doi.org/10.1016/j.foodchem.2022.132864>
- Sheng, R., Cheng, W., 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 Biology and Technology*, 156, Article 110952. <https://doi.org/10.1016/j.postharvbio.2019.110952>
- Sheng, X., Zan, J., Jiang, Y., Shen, S., Li, L., & Yuan, H. (2023). Data fusion strategy for rapid prediction of moisture content during drying of black tea based on micro-NIR spectroscopy and machine vision. *Optik*, 276, Article 170645. <https://doi.org/10.1016/j.jleo.2023.170645>
- Song, D., Gao, D., Sun, H., Qiao, L., Zhao, R., Tang, W., & Li, M. (2021). Chlorophyll content estimation based on cascade spectral optimizations of interval and

- wavelength characteristics. *Computers and Electronics in Agriculture*, 189, Article 106413. <https://doi.org/10.1016/j.compag.2021.106413>
- Song, J., Shi, X., Wang, H., Lv, X., Zhang, W., Wang, J., ... Li, W. (2024). Combination of feature selection and geographical stratification increases the soil total nitrogen estimation accuracy based on Vis-NIR and pXRF spectral fusion. *Computers and Electronics in Agriculture*, 218, Article 108636. <https://doi.org/10.1016/j.compag.2024.108636>
- Strati, I. F., & Oreopoulou, V. (2014). Recovery of carotenoids from tomato processing by-products – A review. *Food Research International*, 65, 311–321. <https://doi.org/10.1016/j.foodres.2014.09.032>
- Tilahun, S., Park, D. S., Seo, M. H., Hwang, I. G., Kim, S. H., Choi, H. R., & Jeong, C. S. (2018). Prediction of lycopene and β-carotene in tomatoes by portable chroma-meter and VIS/NIR spectra. *Postharvest Biology and Technology*, 136, 50–56. <https://doi.org/10.1016/j.postharvbio.2017.10.007>
- United States Department of Agriculture. (2008). *Tomato grades and standards*. U.S. Department of Agriculture, Agricultural Marketing Service. <https://www.ams.usda.gov/grades-standards/tomato-grades-and-standards>
- Vakilzadeh Ebrahimi, M. K., Lee, H., Won, J., Kim, S., & Park, S. S. (2023). Estimation of soil texture by fusion of near-infrared spectroscopy and image data based on convolutional neural network. *Computers and Electronics in Agriculture*, 212, Article 108117. <https://doi.org/10.1016/j.compag.2023.108117>
- Wang, F., Li, Y., Peng, Y., Sun, H., & Li, L. (2018). Nondestructive determination of lycopene content based on visible/near infrared transmission spectrum. *Chinese Journal of Analytical Chemistry*, 46(9), 1424–1431. <https://doi.org/10.11895/j.issn.0253-3820.181164>
- Wong, C. E., Teo, Z. W. N., Shen, L., & Yu, H. (2020). Seeing the lights for leafy greens in indoor vertical farming. *Trends in Food Science & Technology*, 106, 48–63. <https://doi.org/10.1016/j.tifs.2020.09.031>
- Workman, J., & Weyer, L. (2007). Introduction of near-infrared spectroscopy. In *Practical guide to interpretive near-infrared spectroscopy* (pp. 1–12). CRC Press. <https://doi.org/10.1201/9781420018318>.
- Wu, G., Fang, Y., Jiang, Q., Cui, M., Li, N., Ou, Y., Diao, Z., & Zhang, B. (2023). Early identification of strawberry leaves disease utilizing hyperspectral imaging combining with spectral features, multiple vegetation indices and textural features. *Computers and Electronics in Agriculture*, 204, Article 107553. <https://doi.org/10.1016/j.compag.2022.107553>
- Yun, G., Fang, Y., Liu, D.-L., Xu, L., Yan, T.-L., Cao, D.-S., & Xu, Q.-S. (2019). A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration. *Analytica Chimica Acta*, 1058, 58–69. <https://doi.org/10.1016/j.aca.2019.01.022>
- Zhang, B., Gu, B., Tian, G., Zhou, J., Huang, J., & Xiong, Y. (2018). Challenges and solutions of optical-based nondestructive quality inspection for robotic fruit and vegetable grading systems: A technical review. *Trends in Food Science & Technology*, 81, 213–231. <https://doi.org/10.1016/j.tifs.2018.09.018>
- Zheng, Y., Liu, P., Zheng, Y., & Xie, L. (2024). Improving SSC detection accuracy of cherry tomatoes by feature synergy and complementary spectral bands combination. *Postharvest Biology and Technology*, 213, Article 112922. <https://doi.org/10.1016/j.postharvbio.2024.112922>
- Zhou, L., Zhang, C., Qiu, Z., & He, Y. (2020). Information fusion of emerging non-destructive analytical techniques for food quality authentication: A survey. *TrAC Trends in Analytical Chemistry*, 127, Article 115901. <https://doi.org/10.1016/j.trac.2020.115901>