



Research article

Hyperspectral reflectance imaging for visualizing reducing sugar content, moisture, and hollow rate in red ginseng

Xueyuan Bai¹, Yuting You¹, Hairui Wang, Daqing Zhao, Jiawen Wang^{**}, Wei Zhang^{*}

Northeast Asia Research Institute of Traditional Chinese Medicine, Changchun University of Chinese Medicine, 1035 Boshuo Road, Changchun, 130117, China

ARTICLE INFO

Keywords:

Hyperspectral imaging
Non-destructive analysis
Red ginseng
Visualization
Quality assessment
Spectral angle mapping
Principal component analysis
Traditional Chinese medicine

ABSTRACT

Red ginseng (RG) has been traditionally valued in Northeast Asia for its health-enhancing properties. Recent advancements in hyperspectral imaging (HSI) offer a non-destructive, efficient, and reliable method to assess critical quality indicators of RG, such as reducing sugar content (RSC), water content (WC), and hollow rate (HR). This study developed predictive models using HSI technology to monitor these quality indicators over the spectral range of 400–1700 nm. Image features were enhanced using Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF), followed by classification through Spectral Angle Mapping (SAM). The best-performing model for RSC achieved an R^2 value of 0.6198 and a root mean square error (RMSE) of 0.013. For WC, the optimal model obtained an R^2 value of 0.6555 and an RMSE of 0.014. The spatial distribution of RSC, WC, and HR was effectively visualized, demonstrating the potential of HSI for on-site quality control of RG. This study provides a foundation for real-time, non-invasive monitoring of RG quality, addressing industry needs for rapid and reliable assessment methods.

1. Introduction

Red ginseng (RG), derived from the root of *Panax ginseng* C.A. Meyer, has been traditionally used since the 17th century, widely recognized for its health benefits such as boosting immunity, enhancing mental and physical well-being, and exhibiting antioxidative properties [1]. The mild nature and slightly sweet taste of RG have contributed to its popularity. The rising global demand for RG has prompted the expansion of *P. ginseng* cultivation across diverse climates, significantly influencing the quality of RG due to variations in temperature, soil, and growth conditions [2]. To standardize RG quality assessment, regulations in China dictate that water content (WC) must not exceed 12 %, and reducing sugar content (RSC) must not exceed 30 %. Traditionally, quality evaluation has depended on destructive biochemical methods, which, while reliable, are time-consuming, labor-intensive, and costly [3]. Recent advancements aim to shift towards non-destructive methods that align with the industry's needs for efficiency.

Hyperspectral imaging (HSI) has emerged as a revolutionary tool for non-destructive quality evaluation, offering high-resolution, wide spectral coverage, and continuous spectral bands that surpass conventional methods in capturing both chemical and physical

* Corresponding author.

** Corresponding author.

E-mail addresses: wangjiawen1229@163.com (J. Wang), weizcaas@126.com (W. Zhang).

¹ Xueyuan Bai and Yuting You have made equal contributions as the first authors.

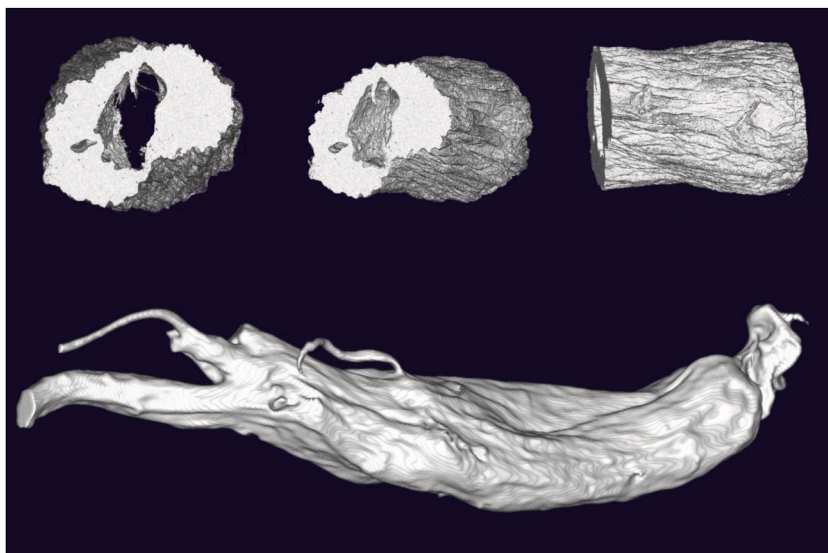


Fig. 1. Microscopic CT imaging of red ginseng hollow under different angles of view. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

information of RG [4–6]. Utilizing hyperspectral cameras, high-resolution images can be acquired in the visible, near-infrared (VIS-NIR), and short-wave infrared (SWIR) regions, providing comprehensive analysis capabilities. Recent advancements in HSI have expanded its applications from traditional uses to innovative monitoring systems, such as flexible Vis-NIR wireless sensors for agricultural monitoring, demonstrating its adaptability and potential for rapid quality assessment in diverse settings [7]. Advanced methods, such as dynamic index evaluation in remote sensing, have shown effectiveness in enhancing data accuracy and reliability under varying environmental conditions, supporting their applicability in real-time RG quality assessments [8].

While HSI has been effectively utilized in agriculture and food industries for tasks such as crop identification [9], pesticide detection [10], and freshness assessment [11], its application in traditional Chinese medicine (TCM) is still developing, mainly focusing on authenticity verification, content detection, and origin identification [12–19]. Notably, few studies address the simultaneous monitoring of RSC, WC, and hardness (HR) in RG, highlighting a significant research gap [19]. In assessing the quality of traditional medicinal plants, it is crucial to validate innovative approaches that improve efficiency and cost-effectiveness. Recent studies highlight the importance of integrating robust machine learning algorithms, such as generative adversarial networks (GANs), which enhance spectral data processing by addressing challenges like noise and interference, thereby significantly improving model accuracy and robustness [20].

To ensure robust quality control of RG, it is crucial to investigate and validate innovative methods that emphasize efficiency and cost-effectiveness. Comparing our HSI-based method with traditional biochemical analyses, we highlight substantial improvements in time, labor, and operational costs, positioning HSI as a transformative approach for real-time RG quality assessment [21]. Although the initial investment in HSI technology is higher, it proves cost-effective in the long run by reducing waste, operational costs, and allowing repeated analyses on the same samples [22]. This approach not only enhances efficiency but also aligns with the industry's shift towards sustainable and non-destructive quality assessment methods.

In the era of AI and big data, integrating "dirty data" from traditional experiments for data augmentation is becoming increasingly important. Recent advancements in machine learning, particularly in convolutional neural networks (CNN) and autoencoders, have demonstrated substantial accuracy improvements in remote sensing and hyperspectral data analysis [23,24]. Advanced image processing techniques, such as Spectral Angle Mapping (SAM), Principal Component Analysis (PCA), and Minimum Noise Fraction (MNF), were selected to handle HSI data due to their specific advantages and suitability for the analysis, significantly improving model robustness and prediction accuracy [25,26]. SAM evaluates spectral similarity by calculating the angle between spectra, making it robust against illumination variations and effective in capturing key spectral features related to quality indicators of red ginseng, while being well-suited for high-dimensional data analysis. PCA is employed for dimensionality reduction and feature extraction, transforming correlated variables into principal components that retain the most significant information, simplifying the modeling process and enhancing prediction accuracy. MNF focuses on optimizing the signal-to-noise ratio, effectively separating signal from noise, which improves image quality and allows clearer visualization of quality indicators. The combination of these three algorithms addresses the challenges of noise reduction, feature enhancement, and classification in hyperspectral data, ensuring the robustness, accuracy, and efficiency of the models, supporting non-destructive and real-time quality control of red ginseng.

The study integrates advanced image processing algorithms with hyperspectral data ranging from 400 to 1700 nm, employing SAM to classify prominent characteristics and identify independent variables, aiming to develop predictive models that accurately assess RG's key quality indicators, such as RSC, WC, and HR. The main objectives are to evaluate HSI technology's feasibility for rapid, non-invasive monitoring and develop robust predictive models to enhance RG quality control processes.

2. Materials and methods

2.1. Sample preparation

P. ginseng samples were collected in Jingyu County, Baishan City, Jilin Province, China. In order to simulate the red ginseng produced under different steaming temperature conditions in the market, the steaming process consisted of 6 cycles, with each cycle gradually increasing the steaming time. Specifically, the steaming times for each cycle were as follows: 2 h, 2.5 h, 3 h, 3.5 h, 4 h, and 4.5 h. During each steaming process, a total of 9 pieces were extracted. After undergoing the steaming process, the RG is subjected to 2 rounds of drying, first at a temperature of 50 °C for a duration of 12 h, followed by an additional 10 h at 55 °C. Subsequently, the dried RG is placed in a location with ample sunlight and proper ventilation until it reaches a hardened texture. The RSC and WC of each RG sample were measured 4 times simultaneously, resulting in a total of 216 data. However, due to the limitations of computer tomography (CT) scanning (Fig. 1), only a single measurement of HR data was obtained, resulting in a smaller sample size of 54 data.

2.2. Basic data acquisition

2.2.1. Determination of hollow rate (HR) of RG

Two primary factors influence the HR in RG. Firstly, during the growth phase, certain cultivators employ oxytocin to expedite the attainment of sales standards and maximize profits. This rapid expansion of *P. ginseng* volume, however, surpasses the capacity for nutrient absorption, resulting in the formation of hollow spaces within the *P. ginseng*. Another contributing factor is the phenomenon of supersaturation of WC in *P. ginseng* during the steaming process, followed by high temperature and prolonged drying. This rapid evaporation of water leads to incongruent rates of internal and external contraction in *P. ginseng*, ultimately resulting in internal hollowness. CT technology is employed for the X-ray tomography of the entire RG, offering the advantages of swift scanning duration and high image clarity [27]. The CT scanning parameters encompass a scanning field acquisition of 72 mm, reconstruction visual field of 72 mm, standard resolution whole body scanning mode, and a scanning time of 6 min.

2.2.2. Determination of RSC in RG

The determination of RSC in RG was conducted in accordance with the regulations set by the China Drug and Food Administration, which stipulates that the content of reducing sugar should not exceed 30 %, using the method of alkaline copper tartrate titration.

2.2.3. Determination of WC in RG

Based on the most recent standard outlined in the Chinese Pharmacopoeia, the permissible WC in RG should not surpass 12 %. In this investigation, the WC was assessed using a moisture detector. A 2g sample of RG powder was utilized, and the detection process lasted for a duration of 2 min.

2.2.4. Hyperspectral data acquisition

HSI of RG was acquired within the 400–1700 nm range using a hyperspectral CCD camera (GEV-B1923M-TC000, IMPERX). The HSI system comprises a hyperspectral camera, 4 halogen lamps, and a data processing center. The entire system is enveloped in an opaque black cloth to prevent disturbing influence of stray-light. To ensure image clarity, the distance between the hyperspectral camera lens and the platform is set at 30 cm, the electric platform moves at a speed of 3 mm/s, the integration time is 3 ms, and the frame rate is 20 frames per second.

2.3. Relative reflectance correction and preprocessing of hyperspectral data

The original spectral data obtained in this study exhibit various issues, including scattering effect, random noise, and system noise, which can be attributed to the influence of hyperspectral acquisition instruments and environmental factors [28]. To mitigate the background noise and eliminate environmental interference, the approach proposed by Yu Liu [29] is employed to correct the reflectance of original spectra with following equation. The relative reflectance (R) of the corrected hyperspectral is determined by the DN value of the original spectrum (S). Additionally, the DN value obtained by covering the lens with a lens cover (D) and the DN value obtained by using a white reference plate (W), which is approximately 99.99 %, also contribute to the determination of R (Appendix Eq. (1)).

The RG hyperspectral image was processed using ENVI5.3 software from Research Systems Company in Boulder, CO, USA. The image was cropped and embedded to produce the primary hyperspectral image file. 4 regions of interest were identified for each RG sample using the ROI tool. The average spectrum of each ROI was extracted, and the relative reflectivity spectrum was subsequently preprocessed using curve smoothing techniques. This preprocessing step aimed to reduce noise and prevent any alteration or loss of critical information [30].

2.4. Feature enhancement and band selection of hyperspectral images

PCA and MNF are widely used techniques for enhancing features in hyperspectral data. PCA achieves this by transforming the original variables, which may exhibit correlation, into a set of orthogonal variables that are unrelated. The amount of information contained in these variables is typically measured by their variance. Consequently, the principal component with the highest variance,

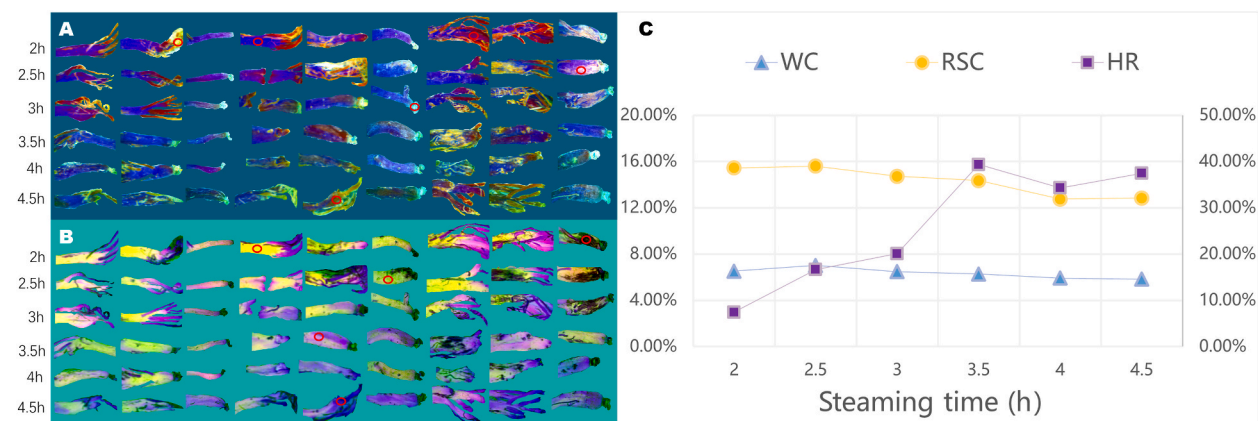


Fig. 2. PCA (A) and MNF (B) contrast-enhanced images (red circles represent the relevant alone features), changes in different measured components at different steaming time points (C). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1

Statistic data of calibration set and validation set.

| | Maximum | | Minimum | | Average | | Standard deviation | |
|-----|-------------|------------|-------------|------------|-------------|------------|--------------------|------------|
| | Calibration | Validation | Calibration | Validation | Calibration | Validation | Calibration | Validation |
| HR | 88.54 % | | 0.00 % | | 25.88 % | | 0.25 | |
| WC | 11.96 % | 11.85 % | 0.00 % | 0.00 % | 6.17 % | 6.70 % | 0.03 | 0.03 |
| RSC | 21.50 % | 21.30 % | 3.71 % | 3.77 % | 7.98 % | 8.02 % | 0.03 | 0.03 |

referred to as PC1, contains the most information [31]. In contrast, MNF demonstrates efficacy in the segregation of signal and noise through the augmentation of signal-to-noise ratio across all bands, thereby ensuring the preservation of spectral and spatial resolution [32].

To enhance the efficiency of subsequent modeling, the utilization of PCA and MNF algorithms is employed to enhance the entirety of spectral data within the 400–1700 nm range. The outcome of the PCA and MNF processing is visually presented in Fig. 2A and B. The ROI spectral curves pertaining to various feature categories are extracted, followed by the application of the SAM algorithm to classify the aforementioned selected features. Feature bands exhibiting high standard deviation are chosen, while bands with low standard deviation are disregarded as they are deemed invalid, thereby reducing the computational burden associated with subsequent modeling.

2.5. Establishment and assessment of regression model

The models for HR, WC, and RSC of RG within the VIS-NIR (400–1000 nm) and SWIR (900–1700 nm) range was established using linear regression. The samples were divided into a calibration set and a validation set in a 7:3 proportion. Based on Zornoza's report [33], it is suggested that R^2 values ranging from 0.61 to 0.8 signify the suitability of these models for prediction purposes, while R^2 values between 0.81 and 0.9 indicate commendable model performance. Conversely, a smaller Root Mean Square Error (RMSE) value that approaches 0 signifies enhanced accuracy of the model. Additionally, the Mean Absolute Percentage Error (MAPE) value should be below 20 % and above 0.2 %. However, a large MAPE value suggests a high average fluctuation level within the overall data and an uneven data distribution.

3. Results and discussion

3.1. Actual measured data from HR, WC and RSC in RG

In this study, we examined the alterations in HR, WC, and RSC of RG samples during the steaming process (Fig. 2C). The measured results revealed that both RSC and WC exhibited a similar pattern of change, reaching their peak at the second steaming for a duration of 2.5 h. As the steaming time was prolonged, there was a gradual decrease in both RSC and WC. Conversely, the HR of RG experienced a sharp increase between 2 h and 3.5 h of steaming, reaching its maximum at 3.5 h. Subsequently, the HR remained relatively stable with minimal fluctuations between 3.5 h and 4.5 h.

Based on the Kennard-Stone (KS) algorithm, as implemented by Wei [34], a total of 216 sample data points were partitioned into a calibration set and a validation set, maintaining a 7:3 ratio. It is essential for the calibration set to encompass the content range of the

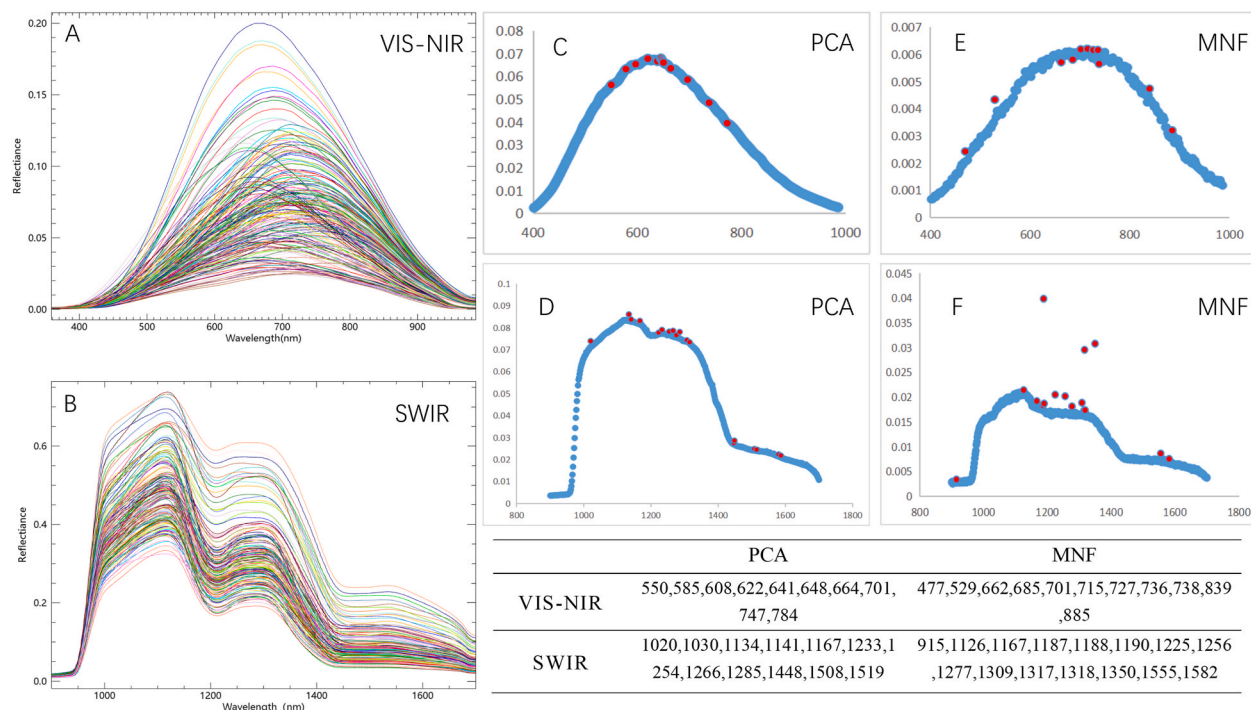


Fig. 3. Average spectra of RG in VIS-NIR (A) and SWIR (B) range; wavelength selection based on different preprocessed methods VIS-NIR-PCA (C), SWIR-PCA (D), VIS-NIR-MNF (E) and SWIR-MNF (F).

validation set. Due to the limited availability of HR data samples, no dataset partitioning was conducted for this variable. The statistical summary of the calibration and validation sets, including maximum, minimum, average, and standard deviation values, can be found in Table 1. The standard deviation for WC and RSC is 0.03, while the standard deviation for hollow fraction is 0.25. From the analysis of standard deviation data, it is evident that there is a significant variation in the HR data among different RG samples, whereas the WC and RSC exhibit relatively consistent levels.

Table 1 presents the statistical distribution of the calibration and validation datasets used in the study, including the minimum, maximum, mean, and standard deviation values for key quality indicators such as reducing sugar content, water content, and hollow rate. This detailed analysis provides a clear understanding of the data's variability and distribution, underscoring the dataset's representativeness and robustness. The information conveyed by Table 1 is essential as it demonstrates that the calibration and validation sets encompass a wide range of conditions, simulating real-world variability. This ensures that the developed models are not only validated on consistent data but are also capable of accurately predicting quality indicators across diverse sample conditions, reinforcing the models' reliability and practical applicability in red ginseng quality assessment.

3.2. Basic spectral analysis of hyperspectral data

Fig. 3A and B depict the outcomes of applying relative reflectivity correction and curve smoothing to the original spectra. However, the preprocessed spectral superposition image exhibits a noticeable reduction in noise and improved stability. To eliminate noise at the band's extremities, spectral data within the 500–900 nm and 1000–1600 nm ranges were selected for subsequent analysis. In Fig. 3A, the spectral reflectance within the wavelength range of 500–900 nm exhibits an initial increase followed by a subsequent decrease, with a notable absorption peak observed at 680 nm. Conversely, Fig. 3B displays 2 distinct peaks (1118 and 1300 nm) and 2 troughs (1200 and 1420 nm) within the wavelength range of 1000–1600 nm. The peak at 1118 nm is potentially associated with the stretching vibration of the C-H bond [35], while the peaks at 1300 nm are indicative of the stretching vibration of the O-H bond [36]. In contrast, the troughs observed at 1200 nm and 1420 nm primarily arise from the plane bending of the O-H bond [37] and the stretching vibration of the N-H bond [38], respectively.

3.3. Characteristic wavelength selection of RG

This study conducts a comparative analysis of 2 distinct feature enhancement techniques, namely PCA and MNF, for the purpose of feature enhancement. Subsequently, the standard deviation index of SAM classification outcomes is employed to identify the wavelengths associated with the extracted features. The findings of the extraction process reveal that PCA yields 10 characteristic wavelengths in the VIS-NIR region and 12 in the SWIR region, while MNF produces 11 characteristic wavelengths in the VIS-NIR region and

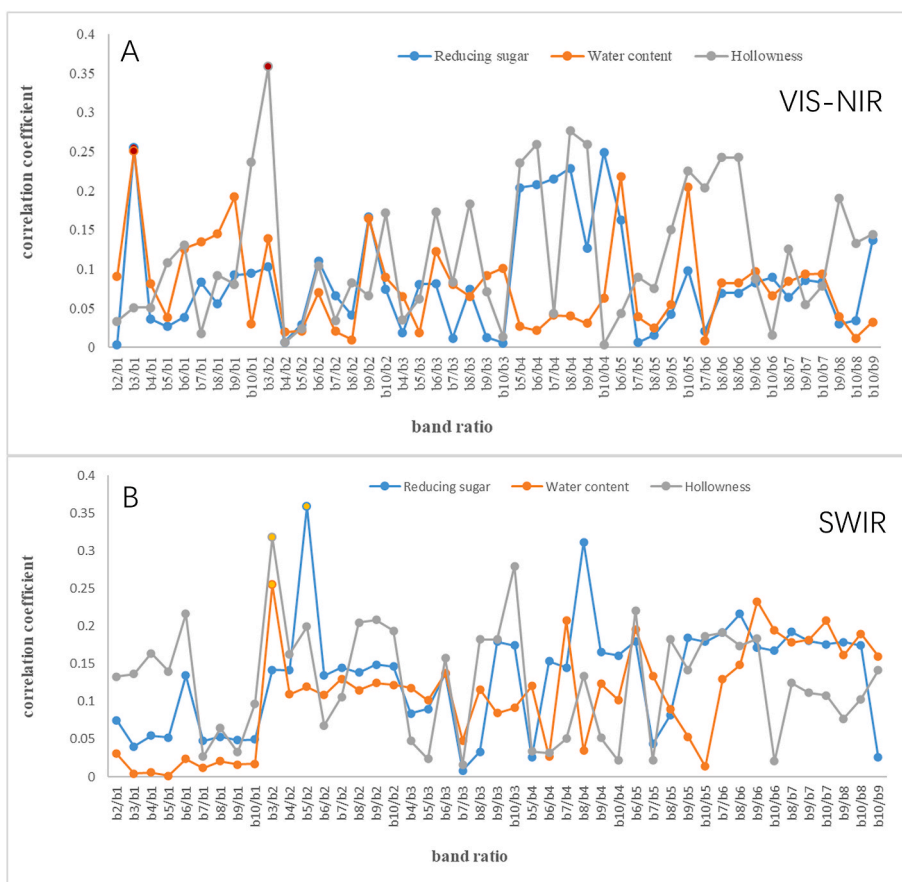


Fig. 4. Correlation coefficient between PCA band ratio and RG components.

15 in the SWIR region (refer to Fig. 3C–F). In comparison to MNF, PCA selects a reduced number of wavelengths, indicating that the characteristic wavelengths chosen using MNF may encompass more noise band information, resulting in a relatively lower accuracy of the model. Numerous studies have demonstrated that models incorporating effective wavelengths outperform those incorporating noise wavelengths. For instance, Xiao [39] employed PCA to forecast the geographical origin of Radix Astragali using both full wavelength and feature wavelength modeling, and the findings clearly indicate that the model based on characteristic wavelengths significantly outperforms the prediction outcome of the full wavelength model.

Fig. 3 displays the characteristic wavelengths of various spectral ranges. Notably, the characteristic wavelengths chosen within the VIS-NIR range predominantly fall within the 500–750 nm range. Similarly, the characteristic wavelengths selected within the SWIR range are primarily concentrated in the 1100–1300 nm range. This can be attributed to the presence of C-H bonds in carbohydrates and O-H bonds in water, as previously noted.

3.4. PCA bands selection for the modeling of HR, WC and RSC in RG

Initially, we conducted a pairwise comparison of the feature wavelengths chosen by the SAM classification algorithm. Subsequently, we examined the relationship between the ratio outcomes and the values of RSC, HR, and WC. The wavelength ratio exhibiting stronger correlation coefficients was chosen as the independent variable for constructing a linear regression model. Nevertheless, the model failed to achieve the intended discrimination effect due to substantial noise present in the image visualization analysis. Hence, the initial selection of the first 10 principal component bands identified through PCA was made for the purpose of conducting linear correlation analysis between the ratio of each band and the component content. Similarly, the band ratio exhibiting stronger correlation coefficients was chosen as an additional independent variable to construct a linear regression model.

Fig. 4 displays the correlation coefficients between the first 10 PCA bands subsequent to pairwise ratio and each component. A higher correlation indicates a greater accuracy in subsequent modeling. In the VIS-NIR region (Fig. 4A), the correlation coefficients of b3/b1 exhibit the highest values for RSC and WC, measuring 0.26 and 0.25, respectively. As for HR, the band ratio b3/b2 demonstrates the highest correlation coefficient of 0.36. Consequently, within the VIS-NIR region, we ultimately selected b3/b1 as the independent variable for the linear regression model of RSC and WC, while b3/b2 was chosen as the independent variable for the linear regression model of HR.

Table 2
Evaluation results of RG related component models.

| Spectral range | Content | Model evaluation | PCA | | MNF | | |
|----------------|---------|------------------|-----------------------------|------------------|---------------------------|------------------|--------|
| | | | Characteristic wavelength | PCA bands (1–10) | Characteristic wavelength | MNF bands (1–10) | |
| VIS-NIR | HR | R ² | 0.2865 | 0.3028 | 0.4554 | 0.1696 | |
| | | WC | R _c ² | 0.5979 | 0.6015 | 0.6183 | 0.3145 |
| | | | R _v ² | 0.643 | 0.6555 | 0.4573 | 0.3151 |
| | | | RMSE | 0.012 | 0.014 | 0.018 | 0.013 |
| | | | MAPE | 16.6 | 14.9 | 18.5 | 18.7 |
| | RSC | | R _c ² | 0.476 | 0.508 | 0.5225 | 0.1773 |
| | | | R _v ² | 0.5325 | 0.4896 | 0.4749 | 0.3879 |
| | | RMSE | 0.015 | 0.016 | 0.016 | 0.017 | |
| | | MAPE | 16.5 | 16.5 | 17.8 | 17.7 | |
| SWIR | HR | R ² | 0.4354 | 0.6184 | 0.5626 | 0.5151 | |
| | | WC | R _c ² | 0.4836 | 0.4206 | 0.5745 | 0.4776 |
| | | | R _v ² | 0.404 | 0.3831 | 0.5823 | 0.2877 |
| | | | RMSE | 0.015 | 0.014 | 0.012 | 0.016 |
| | | | MAPE | 18.4 | 18 | 19.3 | 17.4 |
| | RSC | | R _c ² | 0.4884 | 0.5225 | 0.5337 | 0.5003 |
| | | | R _v ² | 0.5687 | 0.6198 | 0.3213 | 0.28 |
| | | | RMSE | 0.016 | 0.013 | 0.014 | 0.016 |
| | | | MAPE | 14.7 | 16.2 | 16.3 | 23.9 |

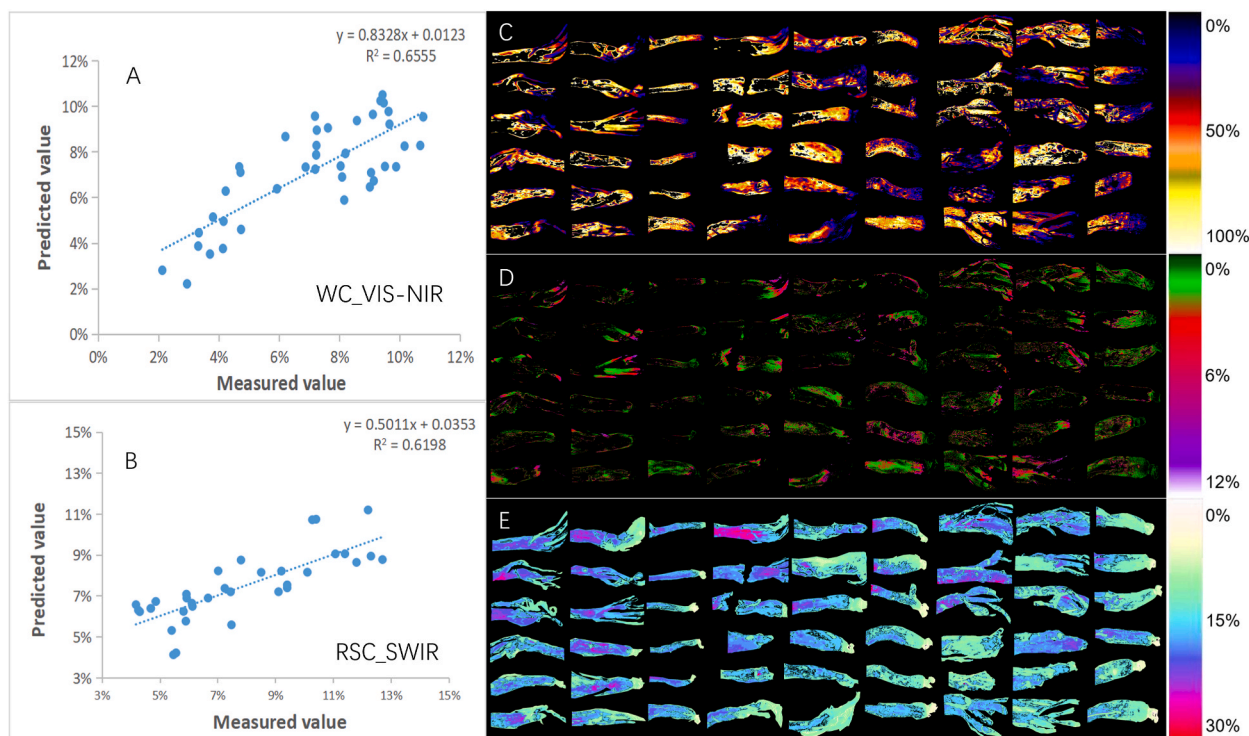


Fig. 5. Scatter plots of predicted/measured values of WC (A) and RSC (B), visualization of HR (C), WC (D) and RSC (E).

In the SWIR region, the band ratios of b3/b2 exhibit the highest correlation coefficients for WC and HR, with values of 0.25 and 0.31, respectively. Conversely, for RSC, the band ratio of b5/b2 demonstrates the highest correlation coefficient, measuring 0.35. Consequently, we have selected b3/b2 as the independent variable for the linear regression model of WC and HR, while b5/b2 serves as the independent variable for the linear regression model of RSC.

3.5. Construction and assessment of linear regression models

Several linear regression models were developed to analyze the relevant components of RG by utilizing characteristic wavelength and PCA band as independent variables. The evaluation results of the model calibration set and validation set can be found in Table 2.

In terms of predicting HR and RSC, the SWIR region was identified as the optimal spectral range, and the most accurate prediction models were established using the PCA bands. Notably, the R^2 value for the top-performing HR model was determined to be 0.6184, while the R^2 , RMSE, and MAPE values for the best RSC model (Fig. 5B) were found to be 0.6198, 0.013, and 16.2 (Table 2), respectively.

According to the optimal spectral range for WC prediction for RG (Fig. 5A), VIS-NIR is the optimal range, and the PCA band is used as an independent variable to determine the optimal model. The evaluation parameters R^2 , RMSE, and MRE have values of 0.6555, 0.014, and 14.9 (Table 2), respectively. Based on comparison, we can conclude that the prediction model established in the MNF band performs the worst.

Table 2 presents the evaluation results of the models using PCA and MNF feature selection methods across the VIS-NIR and SWIR spectral ranges. Overall, models based on PCA characteristic wavelengths performed better than those based on MNF characteristic wavelengths, with varying performance observed between PCA and MNF bands models for different quality indicators. For RSC, the PCA characteristic wavelength model achieved an R^2 value of 0.5325 in the VIS-NIR range, slightly higher than the MNF characteristic wavelength's 0.4749, indicating that PCA is more effective in capturing RSC variations. However, in the SWIR range, the MNF characteristic wavelength model had a lower RMSE of 0.014, demonstrating MNF's superior noise reduction and stability in high-noise data environments. Regarding WC, the PCA characteristic wavelength models showed significantly higher R^2 values (0.643 in VIS-NIR) compared to MNF characteristic wavelengths, indicating that PCA is more sensitive to spectral features related to water content. Additionally, PCA bands models maintained lower RMSE values of 0.014 in VIS-NIR and SWIR ranges, highlighting PCA's consistent performance in water content prediction. For HR, although the R^2 value of the PCA characteristic wavelength model in the VIS-NIR range was relatively low (0.2865), its PCA bands model performance improved significantly, reaching an R^2 of 0.6184, demonstrating PCA's advantage when combining multiple bands. Overall, PCA and MNF exhibit distinct strengths across different spectral ranges and feature selection methods. PCA characteristic wavelengths excel in prediction accuracy and model stability, while MNF shows unique advantages in noise reduction and handling complex spectral features. These results suggest that integrating PCA and MNF feature enhancement strategies could provide more reliable solutions for non-destructive detection of RG quality indicators and highlight directions for further model performance optimization.

In general, it can be stated that the MNF band possesses a greater amount of information, thereby implying that models constructed using the MNF band are expected to exhibit higher accuracy compared to those relying on characteristic wavelength and PCA bands. Nevertheless, it is important to note that the underlying principle of the MNF algorithm is to enhance the signal-to-noise ratio in a proportional manner. Consequently, as the signal amplifies, so does the noise, thereby potentially impacting the precision of model establishment.

The coefficient of determination (R^2) values for the WC and RSC prediction models established through this method are notably low. This can be attributed primarily to the instability of the selected ROI average spectrum resulting from the physical state (wrinkles) of the RG surface during ROI sample selection. Consequently, while the ROI spectrum captures the general chemical composition of red ginseng, it diminishes specificity. In addition, we attempted to concentrate our efforts by minimizing the quantity of ROI pixels; however, this strategy results in the loss of holistic information pertaining to the sample. Consequently, after thorough deliberation, we opted to compromise the model's accuracy in order to enhance its generalizability, thereby addressing a practical challenge encountered when employing this rudimentary algorithm for content visualization.

3.6. Visualization of RG grade indicators (HR, WC and RSC) distribution

The color discrimination capability of HSI technology confers a significant benefit. Through the extraction of spectral information at the pixel level, a predictive model for grade indicators is constructed, enabling the estimation of the content value associated with each pixel. Distinct content values are represented by diverse colored pixels, ultimately presenting the information to observers through an image distribution format [40]. The visualization inversion outcomes of RG's RSC, HR, and WC are depicted in Fig. 5. In Fig. 5C, the presence of voids in all RG is readily discernible, with brighter colors indicating higher HR. This suggests that voids are a prevalent occurrence during the initial processing of RG. Fig. 5D predominantly exhibits green and red hues, with the heavier RG samples more likely to conform to the WC standards. The RSC index (Fig. 5E) demonstrates that the image of RG is clearly visible, indicating that all RG samples meet the standard requirement of less than 30 %, even when the RSC in low weight RG is closer to the threshold.

Based on the overall trend of changes in the 3 indicator components, it can be inferred that the intrinsic quality of RG is not influenced by the number of steaming times (6 times) in the 6 steaming processes of RG processing. The visualization results suggest that HSI has the capability to predict the content of the target component in RG on-site and also enables non-destructive detection of the spatial structure of RG.

To compare with the latest existing research finding, we also referred to the study by M.R.H. Hossain et al. [41], who developed a multiple linear regression (MLR) model for estimating soil moisture, with an R^2 value of 0.60. This indicates that the predictive accuracy of the model established in this study for certain metrics is acceptable, but future research should focus on optimizing data preprocessing steps and improving the stability of the model. In addition, the application of Independent Component Analysis (ICA) techniques for spectral processing to reduce noise is a highly promising research direction. This approach can significantly improve the quality of spectral data processing, thereby enhancing the reliability of subsequent analyses.

Table 3
Comparison of predicted vs actual values and performance metrics for WC and RSC in external validation dataset.

| ID | Actual WC | Predicted WC | Actual RSC | Predicted RSC | R ² (WC) | RMSE (WC) | MAPE (WC) | R ² (RSC) | RMSE (RSC) | MAPE (RSC) |
|------------------------|-----------|--------------|------------|---------------|---------------------|-----------|-----------|----------------------|------------|------------|
| 1 | 0.147 | 0.171 | 0.299 | 0.318 | | | | | | |
| 2 | 0.189 | 0.198 | 0.340 | 0.315 | | | | | | |
| 3 | 0.127 | 0.113 | 0.263 | 0.269 | | | | | | |
| 4 | 0.155 | 0.179 | 0.274 | 0.297 | | | | | | |
| 5 | 0.197 | 0.206 | 0.283 | 0.303 | | | | | | |
| 6 | 0.104 | 0.094 | 0.347 | 0.370 | | | | | | |
| 7 | 0.166 | 0.149 | 0.266 | 0.280 | | | | | | |
| 8 | 0.132 | 0.147 | 0.345 | 0.359 | | | | | | |
| 9 | 0.153 | 0.165 | 0.276 | 0.287 | | | | | | |
| 10 | 0.125 | 0.114 | 0.289 | 0.317 | | | | | | |
| 11 | 0.136 | 0.148 | 0.254 | 0.237 | | | | | | |
| 12 | 0.178 | 0.164 | 0.335 | 0.311 | | | | | | |
| 13 | 0.141 | 0.131 | 0.271 | 0.262 | | | | | | |
| 14 | 0.191 | 0.206 | 0.330 | 0.360 | | | | | | |
| 15 | 0.150 | 0.157 | 0.305 | 0.287 | | | | | | |
| 16 | 0.120 | 0.097 | 0.281 | 0.305 | | | | | | |
| 17 | 0.138 | 0.128 | 0.275 | 0.259 | | | | | | |
| 18 | 0.173 | 0.190 | 0.324 | 0.336 | | | | | | |
| 19 | 0.121 | 0.136 | 0.257 | 0.265 | | | | | | |
| 20 | 0.165 | 0.177 | 0.332 | 0.348 | | | | | | |
| 21 | 0.146 | 0.157 | 0.297 | 0.282 | | | | | | |
| 22 | 0.186 | 0.171 | 0.336 | 0.319 | | | | | | |
| 23 | 0.140 | 0.150 | 0.265 | 0.253 | | | | | | |
| 24 | 0.168 | 0.179 | 0.344 | 0.323 | | | | | | |
| 25 | 0.152 | 0.138 | 0.293 | 0.312 | | | | | | |
| 26 | 0.127 | 0.112 | 0.272 | 0.288 | | | | | | |
| 27 | 0.176 | 0.185 | 0.321 | 0.313 | | | | | | |
| 28 | 0.141 | 0.127 | 0.285 | 0.304 | | | | | | |
| 29 | 0.131 | 0.118 | 0.269 | 0.257 | | | | | | |
| 30 | 0.180 | 0.186 | 0.346 | 0.354 | | | | | | |
| Overall Metrics | | | | | 0.659 | 0.014 | 9.02 | 0.66 | 0.018 | 5.64 |

3.7. Validation of predictive performance on an external RG dataset

In order to provide more experimental evidence demonstrating the effectiveness of the proposed method, we conducted additional experiments on an external dataset comprising 30 samples of RG. The experimental results, including actual and predicted values of WC and RSC, as well as performance metrics, are provided in Table 3. The model achieved R² values of 0.659 and 0.66 for WC and RSC, respectively, indicating acceptable predictive capability. Furthermore, the RMSE values were 0.014 for WC and 0.018 for RSC, while the MAPE values were 9.02 % and 5.64 %, respectively, demonstrating the accuracy and reliability of the proposed method. These additional results further validate the effectiveness of our approach in predicting the key quality attributes of RG, confirming its robustness and practical applicability across different sample sets.

3.8. Discussion of limitations

One of the primary limitations of this study is the initial high cost associated with HSI technology. The sophisticated hardware required for capturing high resolution spectral data, including hyperspectral cameras and advanced lighting systems, presents a significant financial barrier, especially for small scale or resource constrained environments. Although the long-term benefits of HIS, such as non-destructive testing and enhanced accuracy, can offset these costs, the initial investment remains a challenge for broader adoption. To mitigate this, future work could explore cost-effective hardware alternatives or the development of simplified HSI systems tailored to specific applications.

Additionally, the real-time application of HSI technology in quality control settings faces challenges related to data processing speed and algorithm optimization. The large volume of spectral data generated by HSI requires robust and efficient algorithms to process and analyze data rapidly. Currently, the processing speed may not meet the demands of real-time industrial applications, limiting the practical deployment of HSI for immediate decision-making. Enhancing the efficiency of data processing algorithms, such as through the use of machine learning and AI-driven optimization techniques, is crucial for overcoming these barriers. Future research should focus on refining these algorithms to reduce computational load and improve processing times, thereby facilitating real-time application in diverse settings.

4. Conclusion

By leveraging HSI's visualization capabilities, this study effectively maps the spatial distribution of RSC, WC, and HR in RG, offering a comprehensive approach for real-time quality monitoring. Moreover, the study emphasizes the critical impact of feature extraction techniques on the development of predictive models for HSI data, specifically for RG. Our comparative analysis of PCA and MNF demonstrates that PCA significantly enhances the accuracy of models predicting key quality indicators such as HR, RSC, and WC. Optimal performance for HR and RSC was achieved using spectral intervals within the SWIR region, while WC showed best results in the VIS-NIR region.

Despite the promising outcomes, the study has limitations. The current experimental setup relies on sophisticated and expensive HSI equipment, which restricts broader application in agricultural and pharmacological settings. Additionally, the analysis was limited to specific spectral regions and quality indicators, suggesting that future work could explore broader spectral ranges and more diverse indicators.

Future research will focus on addressing these limitations by developing portable HSI devices and cost-effective HSI systems. Specifically, we aim to create mobile HSI solutions that integrate with smartphones or tablets, enabling on-site assessments of RG quality. Additionally, efforts will be made to streamline the design of HSI systems, using cost-effective materials and open-source software to reduce costs without compromising performance. These initiatives are expected to broaden the accessibility and practical use of HSI technology in real world applications, especially for small scale producers.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Xueyuan Bai: Investigation. **Yuting You:** Writing – original draft. **Hairui Wang:** Validation. **Daqing Zhao:** Conceptualization. **Jiawen Wang:** Visualization. **Wei Zhang:** Writing – review & editing, Project administration.

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.

Acknowledgements

This work was supported by the Science and Technology Development Plan Project of Jilin Province (20210204146YY), and the National Key Research and Development Program of Ministry of Science and Technology of China (2021YFD1600900, 2021YFD1600903-02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e37919>.

Appendix A. Formulae and equations

$$R = \frac{S - D}{W - D} \quad \text{Eq.1}$$

References

- [1] G. In, N.G. Ahn, B.S. Bae, M.W. Lee, H.W. Park, K.H. Jang, B.G. Cho, C.K. Han, C.K. Park, Y.S. Kwak, In situ analysis of chemical components induced by steaming between fresh ginseng, steamed ginseng, and red ginseng, *J. Ginseng Res.* 41 (3) (2017) 361–369, <https://doi.org/10.1016/j.jgr.2016.07.004>.
- [2] S.H. So, J.W. Lee, Y.S. Kim, S.H. Hyun, C.K. Han, Red ginseng monograph, *J. Ginseng Res.* 42 (4) (2018) 549–561, <https://doi.org/10.1016/j.jgr.2018.05.002>.
- [3] D. Wu, D.W. Sun, Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review - Part I: fundamentals, *Innov. Food Sci. Emerg. Technol.* 19 (2013) 1–14, <https://doi.org/10.1016/j.ifset.2013.04.014>.
- [4] S.H. Bai, I. Tahmasbian, J. Zhou, T. Nevenimo, G. Hannet, D. Walton, B. Randall, T. Gama, H.M. Wallace, A non-destructive determination of peroxide values, total nitrogen, and mineral nutrients in an edible tree nut using hyperspectral imaging, *Comput. Electron. Agric.* 151 (2018) 492–500, <https://doi.org/10.1016/j.compag.2018.06.029>.

- [5] L.J. Xu, Y.J. Chen, X.H. Wang, H. Chen, Z.L. Tang, X.S. Shi, X.Y. Chen, Y.C. Wang, Z.L. Kang, Z.Y. Zou, P. Huang, Y. He, N. Yang, Y.P. Zhao, Non-destructive detection of kiwifruit soluble solid content based on hyperspectral and fluorescence spectral imaging, *Front. Plant Sci.* 13 (2023) 1054, <https://doi.org/10.3389/fpls.2022.1075929>.
- [6] S. Heo, J.Y. Choi, J. Kim, K.D. Moon, Prediction of moisture content in steamed and dried purple sweet potato using hyperspectral imaging analysis, *Food Sci. Biotechnol.* 30 (6) (2021) 783–791, <https://doi.org/10.1007/s10068-021-00921-z>.
- [7] M. Wang, B. Wang, R. Zhang, Z. Wu, X. Xiao, Flexible Vis/NIR wireless sensing system for banana monitoring, *Food Qual. Saf.* 7 (2023) fyad025, <https://doi.org/10.1093/qsafe/fyad025>.
- [8] S. Jalayer, A. Sharifi, D. Abbasi-Moghadam, A. Tariq, S. Qin, Assessment of spatiotemporal characteristic of droughts using in situ and remote sensing-based drought indices, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 16 (2023) 1483–1502, <https://doi.org/10.1109/JSTARS.2023.3237380>.
- [9] X. Wu, W.Z. Zhang, J.F. Lu, Z.J. Qiu, Y. He, Study on visual identification of corn seeds based on hyperspectral imaging technology, *Spectrosc. Spect. Anal.* 36 (2) (2016) 511–514, [https://doi.org/10.3964/j.issn.1000-0593\(2016\)02-0511-04](https://doi.org/10.3964/j.issn.1000-0593(2016)02-0511-04).
- [10] W.X. Ye, T.Y. Yan, C. Zhang, L. Duan, W. Chen, H. Song, Y.F. Zhang, W. Xu, P. Gao, Detection of pesticide residue level in grape using hyperspectral imaging with machine learning, *Foods* 11 (11) (2022) 1532, <https://doi.org/10.3390/foods11111609>.
- [11] S.S. Zhu, L. Feng, C. Zhang, Y.D. Bao, Y. He, Identifying freshness of spinach leaves stored at different temperatures using hyperspectral imaging, *Foods* 8 (9) (2019) 350, <https://doi.org/10.3390/foods8090356>.
- [12] Y.X. Pan, H.X. Zhang, Y. Chen, X.C. Gong, J.Z. Yan, H. Zhang, Applications of hyperspectral imaging technology combined with machine learning in quality control of traditional Chinese medicine from the perspective of artificial intelligence: a review, *Crit. Rev. Anal. Chem.* (2023), <https://doi.org/10.1080/10408347.2023.2207652>.
- [13] S. Tankeu, I. Vermaak, W.Y. Chen, M. Sandasi, A. Viljoen, Differentiation between two "fang ji" herbal medicines, *Stephania tetrandra* and the nephrotoxic *Aristolochia fangchi*, using hyperspectral imaging, *Phytochemistry* 122 (2016) 213–222, <https://doi.org/10.1016/j.phytochem.2015.11.008>.
- [14] L.X. Huang, Y.B. Zhou, L.W. Meng, D. Wu, Y. He, Comparison of different CCD detectors and chemometrics for predicting total anthocyanin content and antioxidant activity of mulberry fruit using visible and near infrared hyperspectral imaging technique, *Food Chem.* 224 (2017) 1–10, <https://doi.org/10.1016/j.foodchem.2016.12.037>.
- [15] C. Zhang, W.Y. Wu, L. Zhou, H. Cheng, X.Q. Ye, Y. He, Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging, *Food Chem.* 319 (2020) 126551, <https://doi.org/10.1016/j.foodchem.2020.126536>.
- [16] F. Franch-Lage, J.M. Amigo, E. Skibsted, S. Maspocho, J. Coello, Fast assessment of the surface distribution of API and excipients in tablets using NIR-hyperspectral imaging, *Int. J. Pharm.* 411 (1–2) (2011) 27–35, <https://doi.org/10.1016/j.ijpharm.2011.03.012>.
- [17] J. Cruz, M. Blanco, Content uniformity studies in tablets by NIR-CL, *J. Pharm. Biomed. Anal.* 56 (2) (2011) 408–412, <https://doi.org/10.1016/j.jpba.2011.04.018>.
- [18] K. Awa, T. Okumura, H. Shinzawa, M. Otsuka, Y. Ozaki, Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets, *Anal. Chim. Acta* 619 (1) (2008) 81–86, <https://doi.org/10.1016/j.aca.2008.02.033>.
- [19] C.L. Ru, Z.H. Li, R.Z. Tang, A hyperspectral imaging approach for classifying geographical origins of rhizoma *atractylodis macrocephalae* using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI), *Sensors-Basel* 19 (9) (2019) 2072, <https://doi.org/10.3390/s19092045>.
- [20] J.S. Straub, M.S. Nowotarski, J. Lu, T. Sheth, S. Jiao, M.P.A. Fisher, M.S. Shell, M.E. Helgeson, A. Jerschow, S. Han, Phosphates form spectroscopically dark state assemblies in common aqueous solutions, *Proc. Natl. Acad. Sci. USA* 120 (2023) e2206765120, <https://doi.org/10.1073/pnas.2206765120>.
- [21] E.M. Rodrigues, E. Hemmer, Trends in hyperspectral imaging: from environmental and health sensing to structure-property and nano-bio interaction studies, *Anal. Bioanal. Chem.* 414 (2022) 4269–4279, <https://doi.org/10.1007/s00216-022-03959-y>.
- [22] M.B. Stuart, L.R. Stanger, M.J. Hobbs, et al., Low-cost hyperspectral imaging system: design and testing for laboratory-based environmental applications, *Sensors* 20 (11) (2020) 3293, <https://doi.org/10.3390/s20113293>.
- [23] M. Mohammadi, A. Sharifi, Evaluation of convolutional neural networks for urban mapping using satellite images, *J. Indian Soc. Remote Sens.* 49 (2021) 2031–2045, <https://doi.org/10.1007/s12524-021-01382-x>.
- [24] J. Lv, Y. Xu, L. Xu, L. Nie, Quantitative functional evaluation of liver fibrosis in mice with dynamic contrast-enhanced photoacoustic imaging, *Radiology* 300 (2021) 89–97, <https://doi.org/10.1148/radiol.2021204134>.
- [25] J. Chen, Y. Song, D. Li, X. Lin, S. Zhou, W. Xu, Specular removal of industrial metal objects without changing lighting configuration, *IEEE Trans. Ind. Inform.* 20 (3) (2024) 3144–3153, <https://doi.org/10.1109/TII.2023.3297613>.
- [26] H. Xu, Q. Li, J. Chen, Highlight removal from A single grayscale image using attentive gan, *Appl. Artif. Intell.* 36 (2022) e1988441, <https://doi.org/10.1080/08839514.2021.1988441>.
- [27] S. Nicolaou, M.F. Mohammed, Multienergy computed tomography: a new horizon in computed tomographic imaging preface, *Radiol. Clin. N. Am.* 56 (4) (2018) XV–XVI, <https://doi.org/10.1016/j.rcl.2018.04.001>.
- [28] Y.L. Mao, H. Li, Y. Wang, K. Fan, Y.J. Song, X. Han, J. Zhang, S.B. Ding, D.P. Song, H. Wang, Z.T. Ding, Prediction of tea polyphenols, free amino acids, and caffeine content in tea leaves during wilting and fermentation using hyperspectral imaging, *Foods* 11 (16) (2022) 2458, <https://doi.org/10.3390/foods11162537>.
- [29] Y. Liu, Y.B. Long, H.C. Liu, Y.B. Lan, T. Long, R. Kuang, Y.F. Wang, J. Zhao, Polysaccharide prediction in *Ganoderma lucidum* fruiting body by hyperspectral imaging, *Food Chem. X* 13 (2022) 100258, <https://doi.org/10.1016/j.fochx.2021.100199>.
- [30] E. Taghinezhad, A. Szumny, A. Figiel, The application of hyperspectral imaging technologies for the prediction and measurement of the moisture content of various agricultural crops during the drying process, *Molecules* 28 (7) (2023) 2836, <https://doi.org/10.3390/molecules28072930>.
- [31] H.M. Zhang, T.X. Wu, L.F. Zhang, P. Zhang, Development of a portable field imaging spectrometer: application for the identification of sun-dried and sulfur-fumigated Chinese herbals, *Appl. Spectrosc.* 70 (5) (2016) 879–887, <https://doi.org/10.1177/0003702816638293>.
- [32] B.H. Zhang, W.Q. Huang, J.B. Li, C.J. Zhao, C.L. Liu, D.F. Huang, L. Gong, Detection of slight bruises on apples based on hyperspectral imaging and MNF transform, *Spectrosc. Spect. Anal.* 34 (5) (2014) 1367–1372, [https://doi.org/10.3964/j.issn.1000-0593\(2014\)05-1367-06](https://doi.org/10.3964/j.issn.1000-0593(2014)05-1367-06).
- [33] R. Zornoza, C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui, J. Mataix-Beneyto, Near infrared spectroscopy for determination of various physical, chemical, and biochemical properties in Mediterranean soils, *Soil Biol. Biochem.* 40 (7) (2008) 1923–1930, <https://doi.org/10.1016/j.soilbio.2008.04.003>.
- [34] Y.Z. Wei, W.J. Hu, F.Y. Wu, Y. He, Polysaccharide determination and habitat classification for fresh *Dendrobium*s with hyperspectral imagery and modified RBFNN, *RSC Adv.* 12 (2) (2021) 1141–1148, <https://doi.org/10.1039/D1RA08577H>.
- [35] X.P. Feng, C. Peng, Y. Chen, X.D. Liu, X.J. Feng, Y. He, Discrimination of CRISPR/Cas9-induced mutants of rice seeds using near-infrared hyperspectral imaging, *Sci. Rep.* 7 (2017) 1728, <https://doi.org/10.1038/s41598-017-16254-z>.
- [36] S. Turker-Kaya, C.W. Huck, A review of mid-infrared and near-infrared imaging: principles, concepts, and applications in plant tissue analysis, *Molecules* 22 (1) (2017) 168, <https://doi.org/10.3390/molecules22010168>.
- [37] H.Y. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, *Trends Food Sci. Technol.* 18 (2) (2007) 72–83, <https://doi.org/10.1016/j.tifs.2006.09.003>.
- [38] R.K. Almajidy, K. Mankodiya, M. Abtahi, U.G. Hofmann, A newcomer's guide to functional near infrared spectroscopy experiments, *IEEE Rev. Biomed. Eng.* 13 (2020) 292–308, <https://doi.org/10.1109/RBME.2019.2944351>.

- [39] Q.L. Xiao, X.L. Bai, P. Gao, Y. He, Application of convolutional neural network-based feature extraction and data fusion for geographical origin identification of Radix Astragali by visible/short-wave near-infrared and near infrared hyperspectral imaging, *Sensors-Basel* 20 (17) (2020) 4911, <https://doi.org/10.3390/s20174940>.
- [40] J. He, L.D. Chen, B.Q. Chu, C. Zhang, Determination of total polysaccharides and total flavonoids in Chrysanthemum morifolium using near-infrared hyperspectral imaging and multivariate analysis, *Molecules* 23 (9) (2018) 2314, <https://doi.org/10.3390/molecules23092395>.
- [41] M.R.H. Hossain, M.A. Kabir, Machine learning techniques for estimating soil moisture from smartphone captured images, *Agriculture* 13 (2023) 574, <https://doi.org/10.3390/agriculture13030574>.