



Monitoring quality changes in green tea during storage: A hyperspectral imaging method

Feilan Li¹, Jingfei Shen¹, Qianfeng Yang, Yongning Wei, Yifan Zuo, Yujie Wang, Jingming Ning, Luqing Li^{*}

State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China

ARTICLE INFO

Keywords:

Green tea
Storage period
Hyperspectral imaging technique
Multivariate analysis
Visualization

ABSTRACT

The quality of green tea deteriorates the longer it is stored. However, there is a lack of accurate and rapid methods for determining the storage period of tea. In this study, hyperspectral imaging (HSI) was used to determine the storage period of green teas stored at 4 °C (set 1) and 25 °C (set 2), and to quantify and visualize the main chemical components (e.g. catechins). In this study, three prediction algorithms were compared, in which partial least squares discriminant analysis outperformed the other models in qualitative discrimination, with 98% and 96% correct discrimination for two sets, respectively. Moreover, quantitative models for ester catechins, simple catechins, and total catechins were developed with $R_p > 0.90$ and $RPD > 1.0$, indicating that the models were reliable. Further, a more intuitive visualization of catechin content was achieved. In conclusion, the HSI provides a rapid, non-destructive method to determine the freshness of stored green tea.

1. Introduction

Green tea, with pleasant flavor and fresh taste, is the most produced and marketed type of tea in China. This popularity is also due to its health benefits. Scientific researches support the health benefits of drinking green tea, mainly referring to the catechins in green tea (Mangels & Mohler 3rd., 2017). Due to the absence of fermentation in tea processing, green tea contains significantly higher content of catechins than fermented teas such as black and oolong tea (Musial, Kuban-Jankowska, & Gorska-Ponikowska, 2020). In addition, as the main flavor component, caffeine also provide a crucial role in eliminating fatigue and invigorating the mind (Faudone, Arifi, & Merk, 2021).

In 2022, total annual tea production was 3.34 million tons, of which 1.85 million tons were green tea. Additionally, total tea sales reached 2.40 million tons in 2022, of which 1.31 million tons were sold domestically and 314 kt were exported (China Tea Distribution Association, 2023). Far more tea is produced than can be sold or exported, and this excess tea grows stale when stored for long periods. Moreover, unscrupulous merchants often mix stale tea with new tea to increase profit margins, harming consumers' rights and health and damaging the tea industry's reputation.

When catechins degrade, the total quantities of epigallocatechin

(EGC), epigallocatechin gallate (EGCG) and epicatechin gallate (ECG), which are positively correlated with the quality of green tea, substantially decrease. By contrast, quantities of gallic catechin (GC), epicatechin (EC), and catechin (C), which are negatively correlated with quality, dramatically increase degrading the quality of the tea (Baek, Kim, Duncan, Leitch, & O'Keefe, 2021). Studying the relationship between storage period and quality of tea can inspire the preservation methods of tea in daily life and also predict the shelf life of tea. Guo et al. studied the effect of various factors, including storage time, on the physicochemical composition of Longjing green tea and predicted shelf life using a partial least squares discriminant analysis (PLS-DA) model (Guo et al., 2022). Therefore, it is important to find a fast and effective method to detect the storage period of tea.

The traditional methods for determining tea quality are sensory review (Kochman, Jakubczyk, Antoniewicz, Mruk, & Janda, 2021; Wang et al., 2022a, 2022b) and chemical component analysis (Donlao & Ogawa, 2019; Y. Wei et al., 2021). Sensory review lacks objectivity and accuracy, and it is difficult to adapt to the current needs of quality judgment in the tea industry. The chemical analysis process is complicated, time-consuming and labor-intensive, and can only analyze a single content, which is also not applicable to the online detection and rapid evaluation of tea quality. More rapid methods combining chemical

* Corresponding author.

E-mail address: lilq@ahau.edu.cn (L. Li).

¹ These authors contributed equally to this work

and sensory observations include electronic nose (Yuan et al., 2019) and electronic tongue (Ren, Li, Wei, Ning, & Zhang, 2021; Y. Zhu et al., 2021). However, the sensitivity and immunity of both methods need to be improved.

Hyperspectral imaging (HSI) offers an alternative; it is fast, accurate, and does not damage the tea. HSI has been applied to food analysis (Lu, Sun, Mao, Wu, & Gao, 2017; Y. Wei et al., 2019), damage detection (Kang et al., 2022; J. Xiong et al., 2018), and variety identification (Sun et al., 2018; S. Zhu et al., 2019). In the field of tea production and preservation, Wang et al. (Y. Wang et al., 2024) used HSI to image tea leaves, classifying them using multibranch kernel attention (MBKA). However, their study did not quantitatively analyze tea quality. Tang et al. used HSI to detect the quality of tea (Tang et al., 2023); they collected spectral information on green tea samples utilizing first derivative principal component analysis support vector machine (FD-PCA-SVM) discriminative model to identify tea grades with an accuracy of 93.8% in a training set and 98.2% in a testing set. Nevertheless, Tang et al.'s analysis also lacked quantitative analysis, and the visualization of tea quality was not carried out on the contents of the tea. Based on the shortcomings of the existing studies, we conducted a more in-depth study on green tea storage period using HSI.

The present study analyzed the chemical composition of stored green tea using HSI and quantitative analyses to determine the freshness of green tea. Our methodology proceeded as follows. (1) The collected tea samples were stored under refrigerated and room temperature for 360

days, and a total of 270 samples were fingerprinted based on HSI. (2) Principal component analysis (PCA) results were used to demonstrate the differences between green tea samples at different storage temperatures and storage stages. (3) Three modeling methods, SVM, PLS-DA and random forest (RF), were compared to screen out the best method for qualitatively discriminating the storage period of tea. (4) Then the hyperspectral data were further used to quantify catechins and caffeine, and the distribution of catechins was visualized.

2. Materials and methods

2.1. Sample preparation

Fresh Lu'an Guapian green tea was purchased from Huiliu Tea Ltd., Co. (Lu'an, China). Chemical index analysis and hyperspectral data collection were performed on 30 tea samples. The remaining 240 samples were placed in sealed aluminum foil bags and stored separately at either 4 °C (sample set 1) or 25 °C (sample set 2) for further analysis. For the next 360 days (i.e., 90, 180, 270, 360 days), some samples were taken out of the refrigerator every 90 days for experiment. Fig. 1 illustrates the qualitative and quantitative evaluation of green tea storage quality.

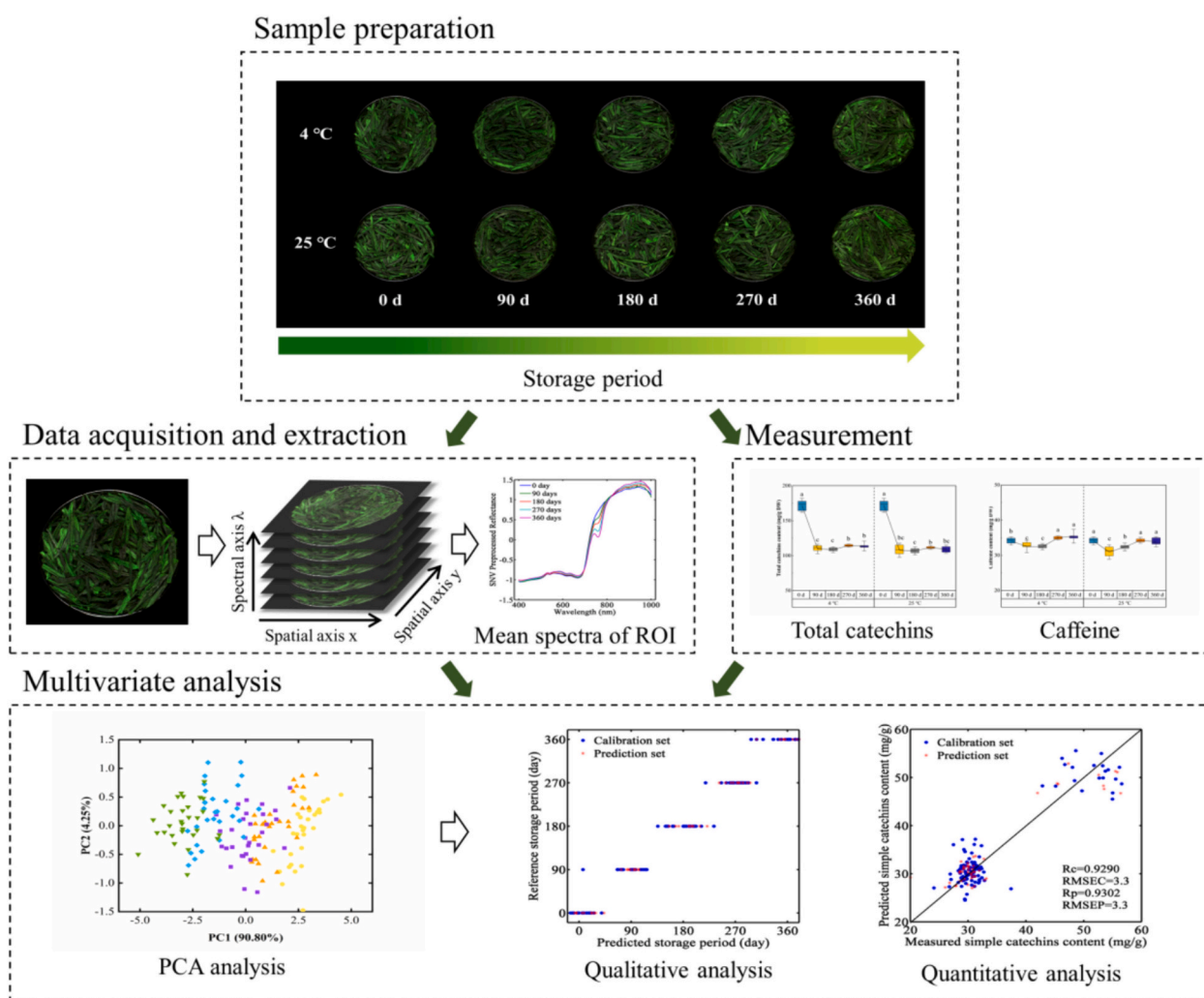


Fig. 1. Flowchart for qualitative and quantitative evaluation of green tea storage quality.

Note: ROI: region of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Chemical measurement of catechins and caffeine

The catechin and caffeine contents of the tea samples were determined at 0, 90, 180, 270 and 360 days according to the ISO 14502-2 standard (ISO, 2005). The chemical measurements of each sample were repeated thrice and averaged.

2.3. HSI technique

Hyperspectral images of tea samples were acquired in reflectance mode using a laboratory hyperspectrometer. The HSI system consisted of a Vis–near-infrared (Vis–NIR) spectrograph with a spectral range of 370.38–1036.54 nm and a spectral resolution of 2.8 nm (ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland), a charge-coupled device camera (GEV–B1621, Imperx, Boca Raton, USA), a set of 150 W halogen lamps (3900-ER, Illumination Technologies Inc., New York, USA), a moving platform with a step motor (Isuzu Optics, Hsinchu, Taiwan, China), and a computer with data acquisition and analysis software (Spectral Image Software, Isuzu Optics Corp., Taiwan, China). The hyperspectral spectral range was 370.38–1036.54 nm with a spectral resolution of 2.8 nm. A total of 1108 spectral bands in the range of 400 to 1000 nm were used due to the low signal-to-noise ratios at both ends of the spectral region. The speed of the conveyor belt, the exposure time of the camera, and the vertical distance between the lens and the sample were set to 1.15 mm s⁻¹, 8.5 ms, and 17.3 cm, respectively, to ensure that the images were clear. After image acquisition, the original hyperspectral image was corrected to the reference image. Specific collection methods are described in the program we have developed (Li et al., 2021). Fig. 1 illustrates the qualitative and quantitative evaluation of green tea storage quality.

2.4. Extraction and preprocessing of spectral data

After acquiring and correcting the hyperspectral images, a 100 × 100 pixel region at the center of the images was selected as the region of interest (ROI). The mean spectrum was extracted by averaging all pixels within the ROI and used as the spectrum for the entire sample. The standard normal variate algorithm (SNV) was employed to preprocess the spectra and reduce the influence of surface scattering and optical path change on the diffuse reflection spectra.

2.5. Supervised and unsupervised chemometrics

PCA involves transforming an initial set of variables into a collection of uncorrelated variables known as principal components (PCs) (Nobari Moghaddam, Tamiji, Akbari Lakeh, Khoshayand, & Haji Mahmoodi, 2022) that capture the maximum variation in the data. In the study, PCA was employed to downscale the hyperspectral data, then extract the main characteristic components of data.

Three machine learning models, namely, SVM, PLS-DA, and RF, were compared to determine the most accurate method for use in subsequent experiments. SVM is a supervised learning model that can be used for classification and regression analysis. These three models were evaluated by the discriminant rate. In our study, the radial basis function (RBF) is selected to convert linear non-fractional data into linear separable space by implicit mapping. The optimal penalty coefficient (*c*) and kernel function parameters (*g*) in the range of 2⁻¹⁰ to 2¹⁰ via a lattice search with a step size of one and a cross validation accuracy of five times to obtain a model with superior generalization ability (Nie, Zhu, & Li, 2020).

PLS is one of the most frequently used multivariate analysis methods for exploring the linear relationship between the independent variables (X) and dependent variables (Y). It is suitable to deal with large amount of spectral variables, especially when the number of spectral wavelengths is more than samples (Lee & Lee, 2021). PLS converts the raw spectral data into PCs which not only carry the variation information in

the raw spectral data as much as possible but also maximize the correlation with Y. In our study, the maximized number of PCs was set 15, and the optimal number of PCs was determined by minimizing predicted residual error sum of squares under a 5-fold cross-validation operation.

RF is a highly flexible algorithm with a wide scope of applicability in civil and geotechnical engineering (Zhou, Huang, Zhou, Armaghani, & Qiu, 2022). The number of trees in the forest determines the complexity and performance of the model and the number of features therein. The performance of the RF classifier was tested by gradually modifying the number of trees in the forest from 50 to 1000 with a step size of 50.

PLSR was introduced for further quantitative modeling. Correlation coefficient (R) and root mean square error (RMSE) were used to evaluate the performance of the model. Model performance was divided into the following three levels by the ratio of performance derivation (RPD): superior (RPD ≥ 2.0), excellent (1.8 ≤ RPD < 2.0), and fair (1.8 ≤ RPD < 2.0) (Munnaf et al., 2021).

2.6. Visualization of catechin content distribution

In this study, the spectra of each pixel of the hyperspectral image of the test data corresponding were fed into the established PLSR prediction model for catechin content and the catechin content of each pixel was calculated. This process was repeated for all pixels of the hyperspectral image to generate a prediction matrix. The predicted values for each pixel were then plotted using a linear color scale, resulting in a final hyperspectral visualization of the graphical effect.

2.7. Software

The ROI selection and spectral extraction procedure were implemented in ENVI 4.7 (ITT Visual Information Solutions, Boulder, CO, USA). The preprocess of spectral data preprocessing, model building and visualization were run on MATLAB R2014a (Version 2014b, MathWorks, Natick, MA, USA).

3. Results and discussion

3.1. Chemical measurement of catechin and caffeine

The changes in EGC, C, EC, EGCG, ECG, caffeine, simple catechins, ester catechins, and total catechin contents of the green tea samples are illustrated in Figs. 2. It reveals that the quantities of caffeine and C did not change substantially during storage at either 4 °C or 25 °C (Fig. 2. (b, f)). The range of caffeine in sample set 1 and 2 are 36.96–25.65, 37.23–23.15, and the standard deviation are 1.62 and 1.87, respectively. Dai et al. found that caffeine has a stable purine ring structure (Dai et al., 2020) and therefore does not degrade substantially over time, which may explain our results. The green tea samples at both temperatures exhibited a significant decrease in the contents of EGCs, ECs, EGCGs, ECGs, simple catechins, ester catechins, and total catechins during the first 90 days of storage. The variation range of catechin in sample set 1 was 187.27–84.71 and the standard deviation was 25.18. Meanwhile, the variation range of catechin in sample set 2 was 187.27–74.05, and the standard deviation was 26.37. This is consistent with the study of Lv et al. that demonstrated slower quality degradation of green tea at lower storage temperatures (Lv et al., 2023).

3.2. Spectral feature acquisition

Fig. 3 presents the spectral profiles of tea samples stored at 4 °C and 25 °C. The results show that the spectral reflectances of the stored samples are different. Compared with the spectral differences among Set 1 samples, the differences among Set 2 samples are more obvious. However, all the samples show similar patterns along the wavelength range (Fig. 3(c, f)). Additionally, samples stored at both temperatures exhibited a higher reflectance in the NIR region compared with the

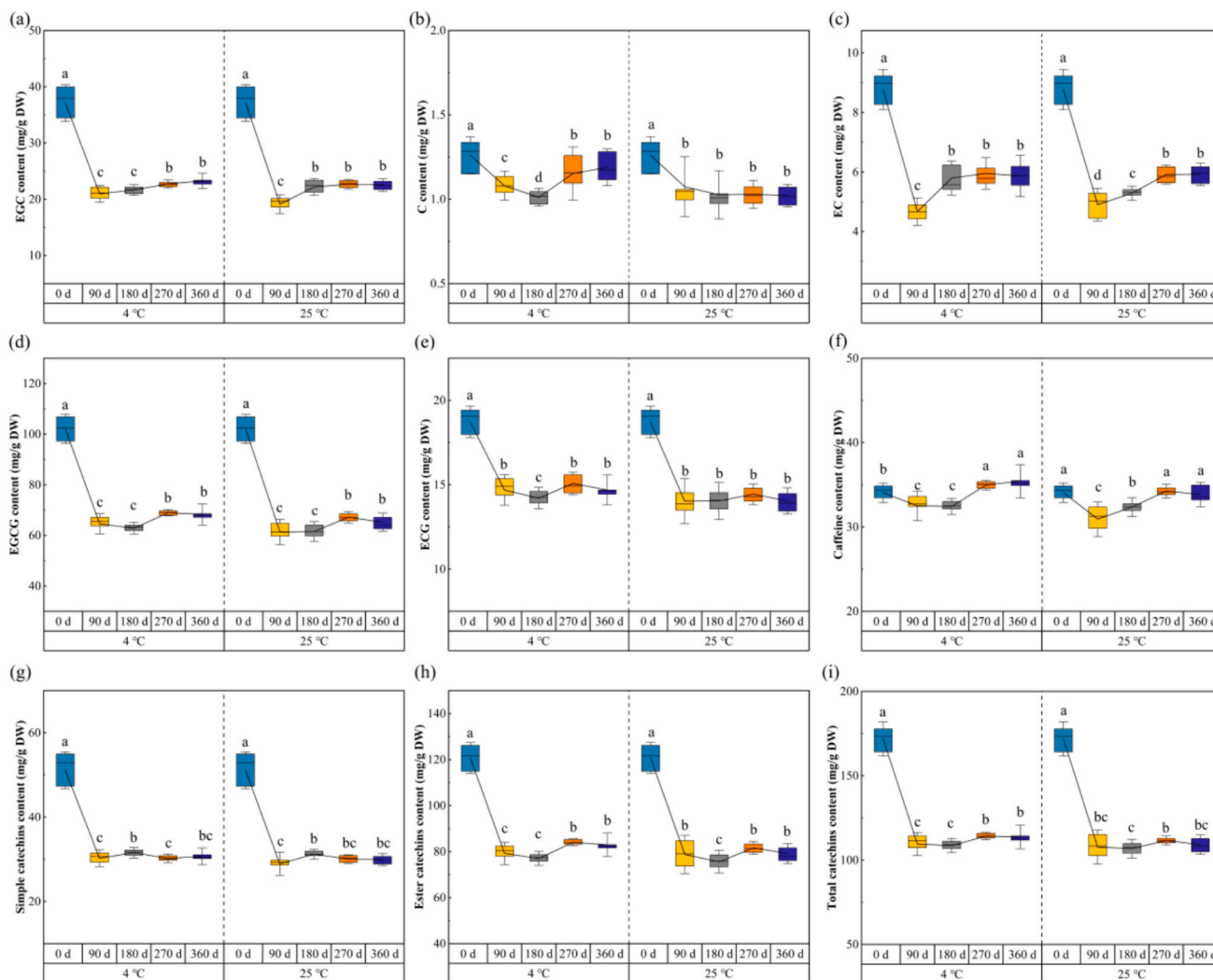


Fig. 2. Boxplot of data on EGC (a), C (b), EC (c), EGCG (d), ECG (e), caffeine (f), simple catechin (g), ester catechin (h), and total catechin (i) contents of green tea sample stored for 360 days under two different temperatures.

Note: EGC: epigallocatechin; C: catechin; EC: epicatechin; EGCG: epigallocatechin gallate; ECG: epicatechin gallate; CAFF: caffeine.

Lines in the upper and lower box represent SDs; values at the top and bottom of the box represent the upper quartile and lower quartile, respectively; the lines in the center of the box represent the median values, whereas different lowercase letters represent significant differences at $P = 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

visible region. Reflection troughs were detected between 650 and 700 nm in the sample spectra, which may have arisen because the chlorophyll in tea mainly absorbs red light. A similar phenomenon was observed in the studies of Huang and Kang (Sanaeifar et al., 2020), with absorption troughs in the red light region at 680 nm.

It is worth noting that a difference in spectral reflectance from 780 to 850 nm was observed in Fig.S1(c, f). The order of reflectance from highest to lowest is: day 0, day 90, day 180, day 270, day 360. The catechins in green tea contain carbonyl groups and phenol OH, which are responsible for the 780-nm reflectance. This finding supports Tang et al.'s observation that catechins are highly reactive between 650 and 800 nm (Tang et al., 2023). Therefore, it is hypothesized that these observed spectral differences are due to changes in the catechin content of the tea. Increased storage time leads to catechin oxidation, hydroxyl dehydrogenation and ketone formation. Additionally, this change was more pronounced in the spectra of the room temperature storage samples (Fig.S1(f)). This is due to the fact that higher temperatures lead to accelerated degradation of catechins, which is exhibited spectroscopically as a difference in reflectance at 780 nm.

3.3. Qualitative analysis of the storage period

3.3.1. PCA analysis

As illustrated in Fig. 3, over 90% of the variance could be explained by the PC1 and PC2, and exhibited superior downscaling. When Figs. 3 (a)(c) and Figs. 3(b)(d) were compared, the aggregation of the PCA score maps of spectra from 4 °C storage conditions was observed to be denser than that from 25 °C conditions. This difference in spectra between samples suggests that the tea samples stored under 4 °C conditions experienced minor degradation that was undetectable through HSI. However, throughout the 360-day storage period, an overlap in the distribution of samples from consecutive 90-day storage intervals was observed, especially between the samples collected on days 0 and 90. Additionally, the PCA utilizing SNV preprocessed data (Figs. 3(b)(d)) differed significantly from the PCA utilizing the raw data (Figs. 3(a)(c)). The observed difference between Figs. 3(a)(c) and Figs. 3(b)(d) may be due to the SNV's ability to make substantial individual corrections in the case of large differences between sample groups (Fig. 1). Moreover, as indicated in Fig. 3(d), the SNV algorithm differentiated the samples and corrected the over-aggregation of the PCA score plot for each storage

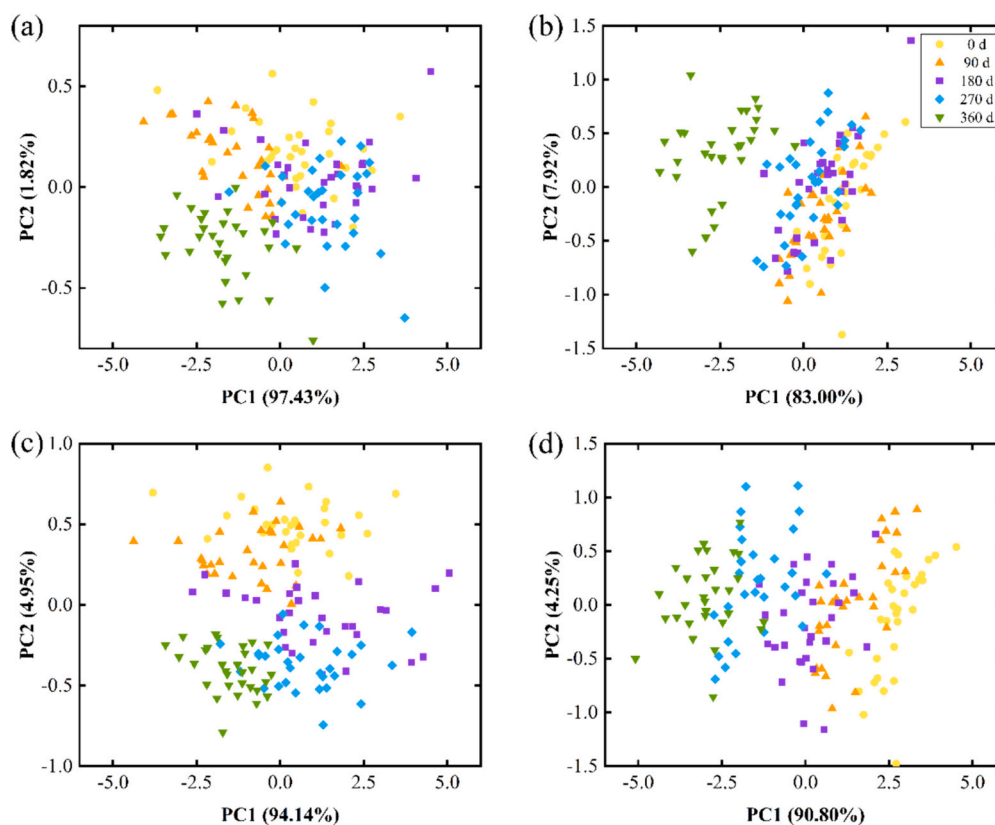


Fig. 3. PCA score plots of tea samples stored under 4 °C and 25 °C with raw spectra (a,c) and SNV preprocessed spectra (b,d), respectively. Note: PC: principal component; d: day.

period under 25 °C conditions. Nevertheless, because PCA is an unsupervised algorithm, we were unable to add labels to facilitate more precise qualitative differentiation in this study.

3.3.2. Establishment of qualitative identification model

The PLS-DA, SVM and RF models were employed to predict the storage period of tea samples from hyperspectral data in the full spectral range of 400–1000 nm. The statistical parameters of PLS-DA, SVM and RF in both calibration and prediction sets for different storage treatments were shown in Table 1. Compared with other qualitative discriminant algorithms, RF achieves better results under calibration conditions, and its discriminant rate in both sample sets 1 and 2 reaches 100%. However, when the predictive conditions were taken into account, the maximum discriminant rate was obtained from the PLS-DA

Table 1

Comparison of different classification models of green tea storage period for sample set 1 and set 2.

Sample treatment	Models	Parameters	Discriminant rate	
			Calibration set	Prediction set
Set 1	SVM	$c = 2896.3100, g = 0.0078$	100%	88%
	PLS-DA	PCs = 20	98%	98%
	RF	N = 400	100%	70%
Set 2	SVM	$c = 512, g = 0.0039$	95%	88%
	PLS-DA	PCs = 13	95%	96%
	RF	N = 100	100%	78%

Note: Set 1: samples stored at 4 °C for 360 days. Set 2: samples stored at 25 °C for 360 days. c: penalty coefficient; g: kernel function parameters; PCs: principal components; N: n-estimators.

model. For set 1, PLS-DA achieved a predictive accuracy of 98% with PCs = 20. For set 2, PLS-DA achieved a predictive accuracy of 96% with PCs = 13. In contrast, the RF discrimination rate did not exceed 80% in either sample sets. Overfitting of data sets in random forests has been observed, which may be caused by the noise in the training data of classification problem (Lang, Tiancai, Shan, & Xiangyan, 2021). Mario et al.'s study on dimensionality reduction methods for hyperspectral data classification demonstrated that PLS-DA outperforms SVM (Mario Fordellone & Mencarelli, 2018), a finding consistent with our results. Additionally, PLS-DA maximizes the total variance of the data using a minimum of components (the latent X scores), rendering it a superior option for our analyses.

3.4. Quantitative prediction of catechins and caffeine

The qualitative results clearly show that PLS-DA exhibits superior performance compared to other algorithms. Therefore, PLSR was adopted to predict the major chemical components in green tea samples. For Ester catechins, $R_p = 0.9265$ for set 1 and $R_p = 0.9172$ for set 2; for simple catechins, $R_p = 0.9073$ for set 1 and $R_p = 0.9227$ for set 2; for total catechins, $R_p = 0.9408$ for set 1 and $R_p = 0.9218$ for set 2, indicated that models capable of achieving accurate quantification were built. For EGCs, simple catechins, and total catechins, RPDs were all >1.4 , indicating that the model performance is excellent. In sets 1 and 2 at both 4 °C and 25 °C, similar quantitative results of EGCs were observed, a finding consistent with the analysis results in Section 3.1. However, the PLSR model showed poor predictive ability for C and caffeine content ($R_p < 0.65$). The low C content in green tea may be the reason for the model's inability to predict C content. The stability of caffeine results in a small range of content variation, and as we know a wide range of data is the basis for stable modeling. Although these two components cannot be accurately quantified, our chemical results indicate that neither C nor

caffeine changes significantly during storage, so in practice there is no need to establish relevant monitoring models.

3.5. Visualized distribution of catechins

In the chemical analysis, it was observed that caffeine was relatively stable during storage and therefore no further visualization was

required. Accordingly, only changes in catechins during storage were visualized: simple catechins (Fig. 4A), ester catechins (Fig. 4B), and total catechins (Fig. 4C). The samples were divided into groups according to storage time and temperature. The predicted catechin contents of all pixels were mapped with a linear color scale using different colors from red to blue (color bar in Fig. 4) representing different catechin contents arranged in descending order to facilitate subsequent analysis. As

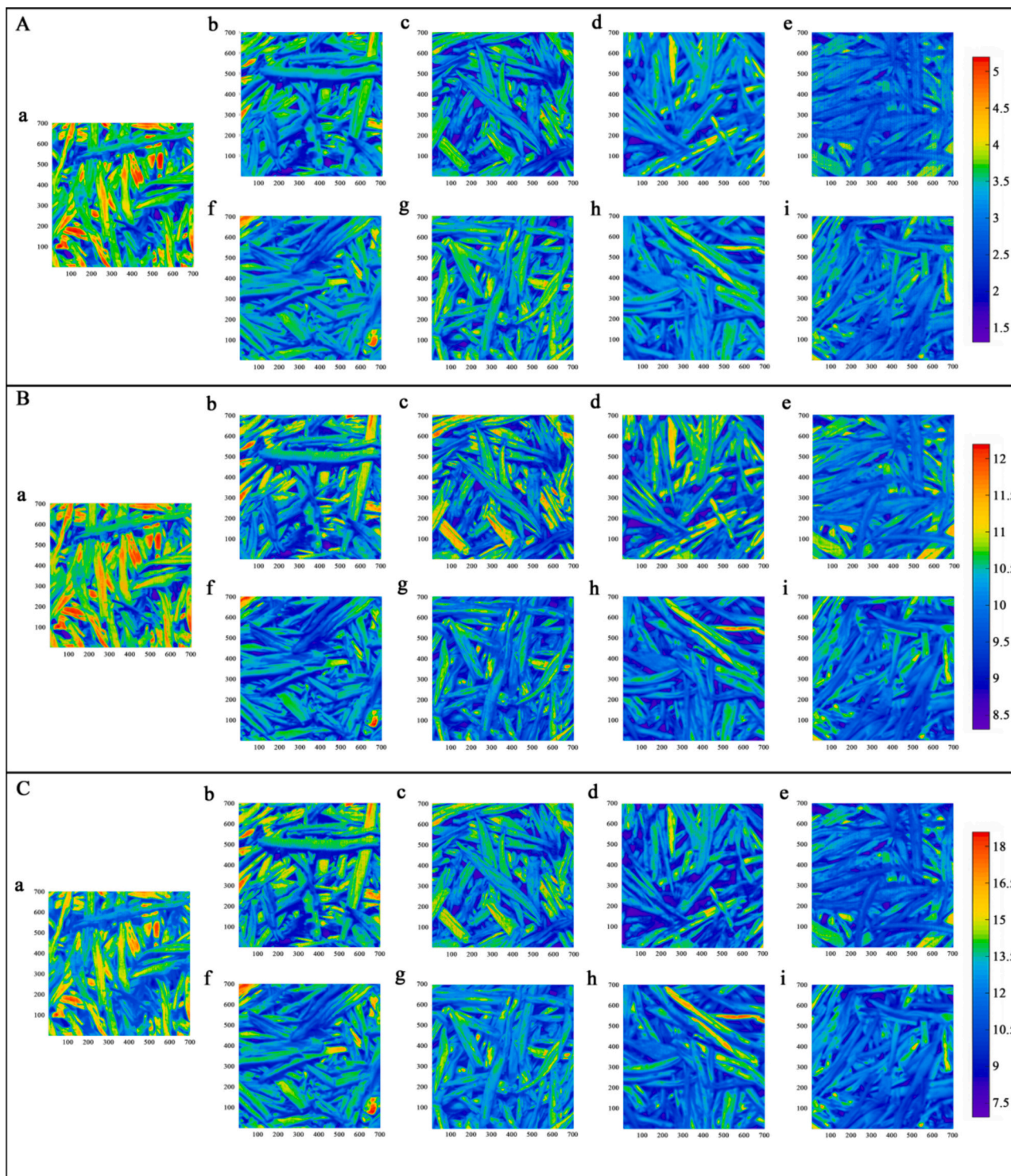


Fig. 4. Visualization maps of catechin content in green tea samples.

Note: a represents samples stored for 0 days, b, c, d, e represent samples stored at 4 °C for 90, 180, 270, and 360 days, respectively, and f, g, h, i represent samples stored at 25 °C for 90, 180, 270, and 360 days, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

illustrated in Fig. 4, a significant decrease occurred after 90 days of storage for all three types of catechins, whereas no further significant change was observed after 180, 270, and 360 days. This observed stability was consistent with the analysis results from Section 3.1. Meanwhile, the catechin content of the green tea samples stored at 25 °C decreased more than those stored at 4 °C. Overall, visualization provides an intuitive way to monitor changes in catechin content during green tea storage.

3.6. Discussion

Green tea loses flavor and salubriousness during storage due to degradation and oxidation of chemical components such as catechins. There are various methods that have been used to evaluate the storage period of tea (seen in Table S1).

Conventional detection methods for non-volatile and volatile substances have been used to study the optimal storage time for different tea types. For example, Xie et al. conducted a metabolomics study utilizing ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UHPLC-QTOF/MS) on the non-volatile components of white tea with different storage times (Xie et al., 2019). Additionally, the same study examined volatile components metabolically to predict the optimal storage time for white tea. Xie et al.'s research primarily focused on substances such as 8-CN-ethyl-2-pyrrolidone, flavan-3-ols, pyroglutamic acid, and catechins. Li et al. conducted a comprehensive analysis and comparison of the volatile components in organic green teas with consecutive storage periods (1–16 years) using simultaneous distillation extraction, gas chromatography–mass spectrometry (SDE-GC–MS), and chemometrics (Li, Han, Mei, Wang, & Han, 2024). Although Li et al.'s study provides a methodology to study the changes of volatile compounds in green teas during the storage process. However, the samples were produced from 2006 to 2021, and the differences in the samples themselves may affect the analysis of the effect of storage time on the samples.

When it comes to non-destructive testing, the classification of total polyphenol content and storage period of Iron Buddha tea was investigated by Xiong et al. based on multi-spectral imaging (C. Xiong et al., 2015). But their study examined tea from different batches, which also lead to differences in the substrate. Yang et al. discriminated the storage time of Pu-erh tea using E-tongue (ET) and E-eye (EE) based on deep learning and achieved very accurate results (Yang et al., 2021). However, it is important to note that the tea samples were stored for 8 years, 6 years, 4 years, 2 years and 0 years. The storage time span and the wide variation in the samples make it easier to discriminate the storage period of the tea. Zhang et al. used indicator displacement assay (IDA) to qualitatively and quantitatively monitor the freshness during the preservation process of green tea (Zhang, Yuan, Ren, Ning, & Wang, 2023) using a low-cost pH dye and metal ions as indicator and receptor, respectively, which is destructive to the tea itself and has low accuracy compared to other tea storage period evaluation methods. Xiao et al. developed a PCA-SVM model based on surface-enhanced raman spectroscopy (SERS) to rapidly predict the storage time of green tea. However, the study was limited to establishing discriminant model and lacked further quantitative analysis (Xiao et al., 2023).

Our methodology has several advantages over those of the aforementioned studies. First, our tea samples were all from the same batch. Second, the samples were subjected to 4 °C and 25 °C temperature storage conditions, simulating cold and room temperature tea storage. Third, the storage period of tea was set at 90-day intervals beginning at day 0, which is a more reasonable time interval and a better guide for the actual production storage and transportation of green tea. Fourth, our study employed hyperspectral image technology to obtain sample information without chemically altering or damaging the tea. We compared three models before deciding on PLS-DA, which had an accuracy of 98% in predicting the chemical contents of set 1 and 96% in predicting set 2. We then elaborated on the hyperspectral data to

quantitatively analyze the main chemical components of green tea. Fifth, we performed quantitative analysis on the primary freshness indicators catechin and caffeine, enabling us to determine tea freshness precisely. Sixth, we employed a qualitative visual analysis of catechin content to test the accuracy and reliability of the quantitative model. In summary, our study provides a rapid and accurate method of evaluating the freshness of stored green tea that does not damage or chemically alter the stored material.

4. Conclusion

Our study presents a novel, efficient, and nondestructive method for evaluating the freshness of green tea using HSI. The results showed that the PLS-DA model constructed using HSI data predicted the freshness of tea samples with 98% accuracy for set 1 and 96% accuracy for set 2. The PLSR models adequately predicted the quantities of EGC, EC, EGCG, easter catechins, simple catechins, and total catechins, with $R_p > 0.85$. Moreover, the model provided satisfactory predictions of all catechin concentrations, with $RPD > 1.0$. However, the PLSR model did not achieve accurate prediction of C and caffeine due to the low levels of C and the stabilization of caffeine. Finally, the distribution of catechins in green tea was visualized. In conclusion, HSI technology is a fusion of spectral technology and image processing technology, with both spectral analysis capability and image resolution capability. It acquires a large and redundant amount of data, which therefore needs to be downsampled before modeling. HSI can accurately, non-destructively, and rapidly determine the freshness of green tea. In future studies, color and aroma changes during green tea storage will be further monitored. Meanwhile, storage methods to extend the shelf life of green tea will be explored, especially there are many types of green teas, and although the quality trends of green teas during storage are roughly similar, it is worthwhile to conduct an in-depth study to see whether the model we developed can predict green teas other than Lu'an Guapian.

CRedit authorship contribution statement

Feilan Li: Writing – review & editing, Writing – original draft, Conceptualization. **Jingfei Shen:** Visualization, Software, Data curation, Conceptualization. **Qianfeng Yang:** Formal analysis, Data curation, Conceptualization. **Yongning Wei:** Data curation, Conceptualization. **Yifan Zuo:** Data curation, Conceptualization. **Yujie Wang:** Formal analysis, Data curation, Conceptualization. **Jingming Ning:** Formal analysis, Conceptualization. **Luqing Li:** Supervision, Resources, 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 study was financially supported by National Key Research and Development Program (2023YFD1601300), National Natural Science Foundation of China (32202543), National Key Research and Development Program (2021YFD1601102), Key Research and Development Program of Anhui Province (202104h04020023), the earmarked fund for CARS (CARS-19).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fochx.2024.101538>.

References

- Baek, N., Kim, Y., Duncan, S., Leitch, K., & O'Keefe, S. (2021). (-)-epigallocatechin Gallate stability in ready-to-drink (RTD) green tea infusions in TiO₂ and oleic-acid-modified TiO₂ Poly(lactic acid) film packaging stored under fluorescent light during refrigerated storage at 4 °C. *Foods*, 10(4), 723. <https://doi.org/10.3390/foods10040723>
- China Tea Distribution Association. (2023). China Tea Production and Sales Situation Report 2022. <https://www.ctma.com.cn/>. Retrieved from Accessed March, 2023.
- Dai, W., Lou, N., Xie, D., Hu, Z., Song, H., Lu, M., & Lin, Z. (2020). N-ethyl-2-pyrrolidinone-substituted Flavan-3-Ols with anti-inflammatory activity in lipopolysaccharide-stimulated macrophages are storage-related marker compounds for green tea. *Journal of Agricultural and Food Chemistry*, 68(43), 12164–12172. <https://doi.org/10.1021/acs.jafc.0c03952>
- Donlao, N., & Ogawa, Y. (2019). The influence of processing conditions on catechin, caffeine and chlorophyll contents of green tea (*Camellia sinensis*) leaves and infusions. *LWT*, 116, Article 108567. <https://doi.org/10.1016/j.lwt.2019.108567>
- Faudone, G., Arifi, S., & Merk, D. (2021). The medicinal chemistry of caffeine. *Journal of Medicinal Chemistry*, 64(11), 7156–7178. <https://doi.org/10.1021/acs.jmedchem.1c00261>
- Guo, Y., Yang, X., Wang, H., Yang, X., Su, X., Kong, J., & Qian, H. (2022). Study on the establishment of quality discrimination model of Longjing 43 green tea (*Camellia sinensis*(L.) Kuntze). *Journal of Applied Research on Medicinal and Aromatic Plants*, 31, Article 100389. <https://doi.org/10.1016/j.jarmap.2022.100389>
- ISO, b. (2005). *ISO 14502-2: 2005*.
- Kang, Z., Geng, J., Fan, R., Hu, Y., Sun, J., Wu, Y., & Liu, C. (2022). Nondestructive testing model of mango dry matter based on fluorescence hyperspectral imaging technology. *Agriculture*, 12(9), 1337. <https://doi.org/10.3390/agriculture12091337>
- Kochman, J., Jakubczyk, K., Antoniewicz, J., Mruk, H., & Janda, K. (2021). Health benefits and chemical composition of Matcha green tea: A review. *Molecules*, 26(1), 85. <https://doi.org/10.3390/molecules26010085>
- Lang, L., Tiancai, L., Shan, A., & Xiangyan, T. (2021). An improved random forest algorithm and its application to wind pressure prediction. *International Journal of Intelligent Systems*, 36(8), 4016–4032. <https://doi.org/10.1002/int.22448>
- Lee, G., & Lee, K. (2021). Feature selection using distributions of orthogonal PLS regression vectors in spectral data. *Biodata Mining*, 14(1), 7. <https://doi.org/10.1186/s13040-021-00240-3>
- Li, J., Han, S., Mei, X., Wang, M., & Han, B. (2024). Changes in profiles of volatile compounds and prediction of the storage year of organic green tea during the long-term storage. *Food Chemistry*, 437, Article 137831. <https://doi.org/10.1016/j.foodchem.2023.137831>
- Li, L., Li, M., Liu, Y., Cui, Q., Bi, K., Jin, S., & Zhang, Z. (2021). High-sensitivity hyperspectral coupled self-assembled nanoporphyrin sensor for monitoring black tea fermentation. *Sensors and Actuators B: Chemical*, 346, Article 130541. <https://doi.org/10.1016/j.snb.2021.130541>
- Lu, X., Sun, J., Mao, H., Wu, X., & Gao, H. (2017). Quantitative determination of rice starch based on hyperspectral imaging technology. *International Journal of Food Properties*, 20(sup1), S1037–S1044. <https://doi.org/10.1080/10942912.2017.1326058>
- Lv, H., Feng, X., Song, H., Ma, S., Hao, Z., Hu, H., & Chen, P. (2023). Tea storage: A not thoroughly recognized and precisely designed process. *Trends in Food Science & Technology*, 140, Article 104172. <https://doi.org/10.1016/j.tifs.2023.104172>
- Mangels, D. R., & Mohler, E. R., 3rd. (2017). Catechins as potential mediators of cardiovascular health. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 37(5), 757–763. <https://doi.org/10.1161/ATVBAHA.117.309048>
- Mario Fordellone, A. B., & Mencarelli, F. (2018). Partial least squares discriminant analysis: A dimensionality reduction method to classify hyperspectral data. *Italian Journal of Applied Statistics*, 31(2). <https://doi.org/10.26398/IJAS.0031-010>
- Munnaf, M. A., Guerrero, A., Nawar, S., Haesaert, G., Van Meirvenne, M., & Mouazen, A. M. (2021). A combined data mining approach for on-line prediction of key soil quality indicators by Vis-NIR spectroscopy. *Soil and Tillage Research*, 205, Article 104808. <https://doi.org/10.1016/j.still.2020.104808>
- Musial, C., Kuban-Jankowska, A., & Gorska-Ponikowska, M. (2020). Beneficial properties of green tea Catechins. *International Journal of Molecular Sciences*, 21(5). <https://doi.org/10.3390/ijms21051744>
- Nie, F., Zhu, W., & Li, X. (2020). Decision tree SVM: An extension of linear SVM for non-linear classification. *Neurocomputing*, 401, 153–159. <https://doi.org/10.1016/j.neucom.2019.10.051>
- Nobari Moghaddam, H., Tamiji, Z., Akbari Lakeh, M., Khoshayand, M. R., & Haji Mahmoodi, M. (2022). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, 107, Article 104343. <https://doi.org/10.1016/j.jfca.2021.104343>
- Ren, G., Li, T., Wei, Y., Ning, J., & Zhang, Z. (2021). Estimation of Congou black tea quality by an electronic tongue technology combined with multivariate analysis. *Microchemical Journal*, 163, Article 105899. <https://doi.org/10.1016/j.microc.2020.105899>
- Sanaeifar, A., Huang, X., Chen, M., Zhao, Z., Ji, Y., Li, X., & Yu, X. (2020). Nondestructive monitoring of polyphenols and caffeine during green tea processing using Vis-NIR spectroscopy. *Food Science & Nutrition*, 8(11), 5860–5874. <https://doi.org/10.1002/fsn3.1861>
- Sun, J., Tang, K., Wu, X., Dai, C., Chen, Y., & Shen, J. (2018). Nondestructive identification of green tea varieties based on hyperspectral imaging technology. *Journal of Food Process Engineering*, 41(5), Article e12800. <https://doi.org/10.1111/jfpe.12800>
- Tang, Y., Wang, F., Zhao, X., Yang, G., Xu, B., Zhang, Y., & Li, L. (2023). A nondestructive method for determination of green tea quality by hyperspectral imaging. *Journal of Food Composition and Analysis*, 123, Article 105621. <https://doi.org/10.1016/j.jfca.2023.105621>
- Wang, J., Li, M., Wang, H., Huang, W., Li, F., Wang, L., & Wan, X. (2022a). Decoding the specific Roasty aroma Wuyi rock tea (*Camellia sinensis*: Dahongpao) by the sensomics approach. *Journal of Agricultural and Food Chemistry*, 70(34), 10571–10583. <https://doi.org/10.1021/acs.jafc.2c02249>
- Wang, J., Li, M., Wang, H., Huang, W., Li, F., Wang, L., & Wan, X. (2022b). Decoding the specific Roasty aroma Wuyi rock tea (*Camellia sinensis*: Dahongpao) by the sensomics approach. *Journal of Agricultural and Food Chemistry*, 70(34), 10571–10583. <https://doi.org/10.1021/acs.jafc.2c02249>
- Wang, Y., Ren, Y., Kang, S., Yin, C., Shi, Y., & Men, H. (2024). Identification of tea quality at different picking periods: A hyperspectral system coupled with a multibranch kernel attention network. *Food Chemistry*, 433, Article 137307. <https://doi.org/10.1016/j.foodchem.2023.137307>
- Wei, Y., Li, T., Xu, S., Ni, T., Deng, W.-W., & Ning, J. (2021). The profile of dynamic changes in yellow tea quality and chemical composition during yellowing process. *LWT*, 139, Article 110792. <https://doi.org/10.1016/j.lwt.2020.110792>
- Wei, Y., Wu, F., Xu, J., Sha, J., Zhao, Z., He, Y., & Li, X. (2019). Visual detection of the moisture content of tea leaves with hyperspectral imaging technology. *Journal of Food Engineering*, 248, 89–96. <https://doi.org/10.1016/j.jfoodeng.2019.01.004>
- Xiao, M., Chen, Y., Zheng, F., An, Q., Xiao, M., Wang, H., & Dai, Q. (2023). Predicting the storage time of green tea by myricetin based on surface-enhanced Raman spectroscopy. *npj Science of Food*, 7(1), 28. <https://doi.org/10.1038/s41538-023-00206-1>
- Xie, D., Dai, W., Lu, M., Tan, J., Zhang, Y., Chen, M., & Lin, Z. (2019). Nontargeted metabolomics predicts the storage duration of white teas with 8-C-N-ethyl-2-pyrrolidinone-substituted flavan-3-ols as marker compounds. *Food Research International*, 125, Article 108635. <https://doi.org/10.1016/j.foodres.2019.108635>
- Xiong, C., Liu, C., Pan, W., Ma, F., Xiong, C., Qi, L., & Zheng, L. (2015). Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system. *Food Chemistry*, 176, 130–136. <https://doi.org/10.1016/j.foodchem.2014.12.057>
- Xiong, J., Lin, R., Bu, R., Liu, Z., Yang, Z., & Yu, L. (2018). A Micro-damage detection method of Litchi fruit using hyperspectral imaging technology. *Sensors*, 18(3), 700. <https://doi.org/10.3390/s18030700>
- Yang, Z., Gao, J., Wang, S., Wang, Z., Li, C., Lan, Y., & Li, S. (2021). Synergistic application of E-tongue and E-eye based on deep learning to discrimination of Pu-erh tea storage time. *Computers and Electronics in Agriculture*, 187, Article 106297. <https://doi.org/10.1016/j.compag.2021.106297>
- Yuan, H., Chen, X., Shao, Y., Cheng, Y., Yang, Y., Zhang, M., & Wu, Z. (2019). Quality evaluation of green and dark tea grade using electronic nose and multivariate statistical analysis. *Journal of Food Science*, 84(12), 3411–3417. <https://doi.org/10.1111/1750-3841.14917>
- Zhang, Y., Yuan, W., Ren, Z., Ning, J., & Wang, Y. (2023). Indicator displacement assay for freshness monitoring of green tea during storage. *Food Research International*, 167, Article 112668. <https://doi.org/10.1016/j.foodres.2023.112668>
- Zhou, J., Huang, S., Zhou, T., Armaghani, D. J., & Qiu, Y. (2022). Employing a genetic algorithm and grey wolf optimizer for optimizing RF models to evaluate soil liquefaction potential. *Artificial Intelligence Review*, 55(7), 5673–5705. <https://doi.org/10.1007/s10462-022-10140-5>
- Zhu, S., Chao, M., Zhang, J., Xu, X., Song, P., Zhang, J., & Huang, Z. (2019). Identification of soybean seed varieties based on hyperspectral imaging technology. *Sensors*, 19(23), 5225. <https://doi.org/10.3390/s19235225>
- Zhu, Y., Wang, J., Wu, Y., Shang, Z., Ding, Y., & Hu, A. (2021). A fluorescent sensor array-based electronic tongue for Chinese tea discrimination. *Journal of Materials Chemistry C*, 9(17), 5676–5681. <https://doi.org/10.1039/D1TC000491C>