Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rapid Authenticity Identification and Adulteration Level Detection of Perilla Oil Adulterated with Soybean Oil Based on Hyperspectral Imaging Technology

Zhen Dong
Jilin Business and Technology College
Fujian Xu
Jilin Business and Technology College
Haijuan Tian
52841263@qq.com

Jilin Business and Technology College
Taogang Hu
Jilin Business and Technology College
Jingwei Zhang
Jilin Business and Technology College
Weiqi Hu
Jilin Business and Technology College
Changcheng Liu
Hangzhou Hyperspectral Imaging Technology Co.,Ltd

Research Article

Keywords: Hyperspectral imaging technology, Perilla oil, Soybean oil, Authenticity identification, Adulteration level detection

Posted Date: April 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9119900/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Rapid Authenticity Identification and Adulteration Level Detection of Perilla Oil Adulterated with Soybean Oil Based on Hyperspectral Imaging Technology

Zhen Dong[1,2] · Fujian Xu[1,2] · Haijuan Tian[1,2]* · Taogang Hu[1,2] · Jingwei Zhang[1,2] ·Weiqi Hu[1,2] · Changcheng Liu [3]

(1. School of Grain Engineering and Nutritional Science, Jilin Business and Technology College, Changchun 130507, China；2. Jilin Province Key Laboratory of Grain and Oil Processing，Jilin Business and Technology College，Changchun 130507, China；3. Hangzhou Hyperspectral Imaging Technology Co.,Ltd，Hangzhou 310024，China)

* Corresponding author: 52841263@qq.com (Tian Haijuan)

Abstract

Perilla seed oil has high nutritional value and is expensive, so it is frequently adulterated with other low-priced vegetable oils to cut costs, which infringes upon the rights and interests of consumers. To achieve rapid authenticity identification and adulteration level detection of perilla oil adulterated with soybean oil, visible-near infrared hyperspectral imaging technology was used to collect spectral data from 168 perilla oil samples with different adulteration ratios. Ten spectral preprocessing methods were employed, combined with two feature extraction algorithms, Successive Projections Algorithm (SPA) and Competitive Adaptive Reweighted Sampling (CARS). Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) models were constructed for qualitative authentication of perilla oil. Light Gradient Boosting Machine (LightGBM) and Stacking models were established for quantitative detection of adulteration levels. The results showed that SVM was the optimal qualitative identification model. SPA feature extraction enhanced its performance, achieving 100% identification accuracy under optimal preprocessing. RF and XGBoost models exhibited excellent fitting capabilities, but feature extraction reduced their generalization ability. Derivative-based and Standard Normal Variate (SNV)-related preprocessing methods were generally the optimal choices for qualitative modeling. For quantitative detection, the single LightGBM model outperformed the Stacking model. The 2-Der-CARS-LightGBM combination was identified as the optimal quantitative model, with a Coefficient of Determination of Cross-validation (R^2_{CV}) of 0.9628 and a Root Mean Square Error of Prediction (RMSEP) of 0.0273. The residuals were randomly and uniformly distributed without systematic error, enabling accurate prediction of perilla oil adulteration levels.

Keywords Hyperspectral imaging technology · Perilla oil · Soybean oil · Authenticity identification · Adulteration level detection

Introduction

Edible oils are essential in the daily human diet, and their quality and safety are directly related to public health (Deng et al. 2022; Niu et al. 2021). However, driven by economic interests, the adulteration of high-value edible oils persists globally, not only infringing on consumer rights but also potentially causing serious public health incidents (Jamwal et al. 2021; Su et al. 2021). The frequent occurrence of edible oil adulteration in recent years underscores the urgency of developing rapid, accurate, and efficient

techniques for authenticity identification and quantitative adulteration detection (Aykas et al. 2022; Duraipandian et al. 2019; Huang et al. 2021). Perilla oil, rich in α-linolenic acid and phenolic bioactive compounds, offers various health benefits (Kangwan et al. 2021; Xia et al. 2025) and holds high market value, making it a frequent target for adulteration. Soybean oil, due to its low price and similar properties to perilla oil, is commonly used as an adulterant. The adulteration of perilla oil with soybean oil significantly degrades its nutritional quality and harms consumer interests. Traditional methods for detecting edible oil adulteration, such as chromatography and chemical assays, although accurate, have limitations including complex pretreatment, being time-consuming, high cost, and requiring professional operation, making them unsuitable for large-scale rapid screening needs (Rogeboz et al. 2024; Wei et al. 2023). Therefore, developing rapid, non-destructive alternative technologies has become a research hotspot.

Hyperspectral imaging (HSI) technology can simultaneously acquire spatial and spectral information from samples (Li et al. 2022; Tang et al. 2023). By analyzing the response of substances to light across different wavelengths, it reveals differences in internal composition. With advantages such as being rapid, non-destructive, and information-rich, HSI shows great potential in identifying adulteration in agricultural products and food (Castro et al. 2023; Cui et al. 2022; Hu et al. 2023; Pu et al. 2023). Combining HSI with chemometrics and machine learning has become a powerful tool for edible oil adulteration detection. HSI (400–1000 nm) combined with Savitzky–Golay smoothing and various wavelength selection methods was used to detect adulteration in four edible oils, including rose oil, achieving a validation accuracy of 93% with a Histogram-based Gradient Boosting model (Aqeel et al. 2026).The same spectral range was employed to identify adulteration in four edible oils, including almond oil, achieving 100% validation accuracy with Linear Discriminant Analysis (LDA) (Aqeel et al. 2024).HSI (900-1650 nm) was used to quantitatively detect palm kernel oil adulteration concentration in virgin coconut oil, achieving a prediction coefficient of determination (R^2) of 0.991 with Partial Least Squares Regression (PLSR) (Jermwongruttanachai et al. 2024).HSI (900-1700 nm) combined with machine learning was used to detect adulteration of extra virgin olive oil with hazelnut oil, refined olive oil, and olive pomace oil, achieving perfect classification with Artificial Neural Network and Partial Least Squares Discriminant Analysis(Malavi et al. 2024).HSI (870-1720 nm) was applied to analyze camellia seed oil adulterated with soybean oil, achieving 100% identification accuracy with both Support Vector Machine (SVM) and LDA, and a prediction correlation coefficient of 0.9789 with Synergy Interval Partial Least Squares(Shang et al. 2024).A comparison was made between fluorescence HSI and infrared spectroscopy for detecting adulteration in extra virgin olive oil, with infrared spectroscopy combined with SVM achieving 99.1% accuracy, while fluorescence HSI combined with Convolutional Neural Network (CNN) achieved 94.2% accuracy(Jiao et al. 2025).These studies demonstrate the good applicability and high accuracy of HSI technology in authenticating different edible oils.

Currently, systematic research on detecting perilla oil adulteration using hyperspectral technology is scarce, especially in-depth discussions on the suitability of spectral preprocessing, feature extraction, and various models. This study combines HSI technology with chemometric methods to detect and analyze the authenticity and different adulteration levels of perilla oil. Through a systematic comparative investigation of different methodologies, this study seeks to determine the optimal combination of spectral preprocessing, feature extraction, and modeling for the detection of perilla oil adulteration. It thereby provides a rapid, accurate, and non-destructive technical scheme and theoretical foundation for the quality supervision of high-value edible oils.

Materials and Methods

Materials

The perilla oil used in the experiment was purchased from Jiangxi Kangshanyuan Oil Co., Ltd., and the soybean oil was purchased from Luhua Fragrant Peanut Oil Co., Ltd. Perilla oil was used as the base oil, and soybean oil as the adulterant. The adulterant oil was mixed into the base oil at concentration levels of 3%, 5%, 8%, 10%, 12%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, and 90%. Eight parallel samples were prepared for each concentration level. The total number of samples was 168, including 136 binary adulterated samples, 24 pure perilla oil samples, and 8 pure soybean oil samples. Each sample was prepared with a volume of 10.00±0.05 mL, thoroughly mixed by shaking, and then stored at 25°C in the laboratory for 24 hours before hyperspectral data acquisition.

Hyperspectral data acquisition

Hyperspectral images were acquired using a visible-near infrared (Vis-NIR) hyperspectral imaging system (Hangzhou Hyperspectral Imaging Technology Co., Ltd). The system consisted of a tungsten lamp light source, lens, camera, computer, and monitor. The spectral acquisition range was 374-1010 nm, with a spectral resolution better than 2.8 nm, a sampling interval of 2.2 nm, and 300 spectral bands. A schematic diagram of the acquisition system is shown in Fig. 1. For hyperspectral data acquisition, each sample was poured into a 35 mL ceramic evaporation dish and placed on the stage. Each image scanned four samples, and each group of samples was scanned three times.

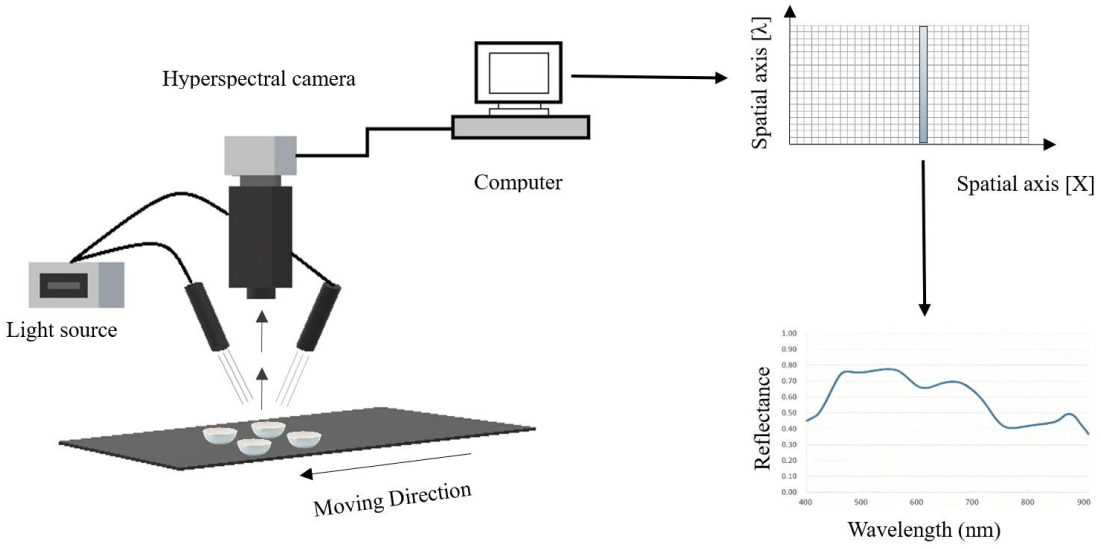


Fig. 1 Hyperspectral image acquisition system and process

Region of interest (ROI) extraction and spectral correction

ENVI® v5.3 software (Exelis Visual Information Solutions, Inc., USA) was used to extract ROIs from the acquired images. The area where the sample was located in each image was designated as the ROI. The spectral reflectance of all pixels within the ROI was extracted, and their average spectral reflectance was calculated as the spectral data for the sample. To reduce the influence and noise caused by detector

sensitivity, camera physical differences, and external interference, the obtained raw hyperspectral images required blackboard and whiteboard correction to convert them to reflectance spectral mode. The black-and-white correction formula for the image is:

$$R = \frac{R_0 - R_B}{R_W - R_B}$$

Where: R is the corrected image information; R_0 is the original image information; R_B is the blackboard image information, obtained by covering the lens with the lens cap; R_W is the whiteboard image information, obtained using a polytetrafluoroethylene plate.

Spectral preprocessing

To reduce the impact of instrument noise, baseline drift, light scattering effects, and sample matrix interference on the quality of spectral data during hyperspectral imaging, and to improve the accuracy and stability of subsequent modeling, this study employed 10 spectral data preprocessing methods on the raw spectra. These included: 1st Derivative (1-Der), 2nd Derivative (2-Der), Savitzky-Golay Smoothing (SG), Normalization, Baseline Correction (BC), Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Median Filtering (MF), 1-Der-SNV, and 2-Der-SNV.

Feature extraction

The spectral data of perilla oil obtained by HSI technology suffer from issues like high dimensionality and redundancy. Directly using them for modeling can easily lead to the curse of dimensionality, reducing model generalization ability and computational efficiency. To screen for characteristic wavelengths highly correlated with perilla oil authenticity identification and adulteration level detection, this study employed two feature extraction algorithms: Successive Projections Algorithm (SPA) and Competitive Adaptive Reweighted Sampling (CARS) to perform feature selection on the preprocessed spectral data(Cao et al. 2022).

Sample partitioning

To establish and validate the perilla oil detection models, the collected hyperspectral data were divided into a calibration set and a validation set using the Kennard-Stone (KS) algorithm. The principle of the KS algorithm is to first select the two samples with the farthest Euclidean distance into the calibration set, and then iteratively add the sample from the validation set with the smallest distance to the calibration set (Farber and Kurouski 2022). Finally, the samples were divided into a calibration set (101 samples) and a validation set (67 samples) at a ratio of 3:2. The validation set served as an independent prediction set to evaluate the external predictive performance of the models.

Qualitative identification of perilla oil authenticity

Three algorithms, SVM, Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were used to establish qualitative identification models for perilla oil authenticity. SVM, based on statistical learning theory, classifies by finding the optimal separating hyperplane, handles nonlinear high-dimensional data using kernel functions, and is suitable for small-sample, high-dimensional hyperspectral data with strong

generalization ability (Prakash et al. 2024). RF is a Bagging ensemble algorithm based on multiple decision trees and a voting mechanism for classification. It uses bootstrap resampling, is resistant to overfitting, robust, and requires simple preprocessing (N. K et al. 2024). XGBoost, improved from gradient boosting decision trees (GBDT), iteratively optimizes the error, incorporates L1 and L2 regularization and learning rate to control complexity, effectively reducing overfitting while balancing accuracy and stability (He et al. 2022). All three algorithms were implemented using Python 3.11.9.

Quantitative detection of perilla oil adulteration level

Light Gradient Boosting Machine (LightGBM) models were established and evaluated for perilla oil adulteration levels. Stacking ensemble models were established and evaluated based on the nine combinations derived from the optimal three preprocessing methods for each of the full spectra, SPA, and CARS feature extraction approaches used with the LightGBM model.LightGBM, as a lightweight gradient boosting algorithm, uses a single-layer structure, directly predicts the adulteration ratio from hyperspectral features, offering high computational efficiency, outstanding nonlinear fitting capability, and strong resistance to noise and overfitting, making it suitable for high-dimensional redundant spectral data (Cai et al. 2021). Stacking is a two-layer ensemble framework. The first layer uses partial least squares (PLS) and Ridge as base models to mitigate collinearity interference and generate primary predictions through K-fold cross-validation. The second layer uses LightGBM as the meta-model, fusing the primary predictions with original features for secondary nonlinear fitting, combining linear and nonlinear advantages to mine data correlations more accurately (Zhao and Cheng 2022).

Both types of models were evaluated using four core metrics: Coefficient of Determination of Cross-validation (R^2_{CV}), Root Mean Square Error of Cross-Validation (RMSECV), Coefficient of Determination of Prediction set (R^2_{Pred}), and Root Mean Square Error of Prediction (RMSEP). R^2_{CV} and R^2_{Pred} values closer to 1 indicate stronger explanatory power and better fitting and prediction performance of the model for the variables. Smaller RMSECV and RMSEP values indicate smaller prediction errors and stronger stability and generalization ability. Both algorithms were implemented using Python 3.11.9.

Results and discussion

Hyperspectral characteristics of samples

The spectral reflectance curves of perilla oil samples with different adulteration levels are shown in Fig. 2. Within the spectral range of 374-1010 nm, the hyperspectral trends of all samples were generally consistent. The spectral characteristics of perilla oil with different adulteration levels and soybean oil were highly similar, making it difficult to directly distinguish differences between samples based on absorbance alone. Therefore, it is necessary to combine chemometric methods with hyperspectral data to achieve authenticity identification and adulteration level detection of perilla oil.

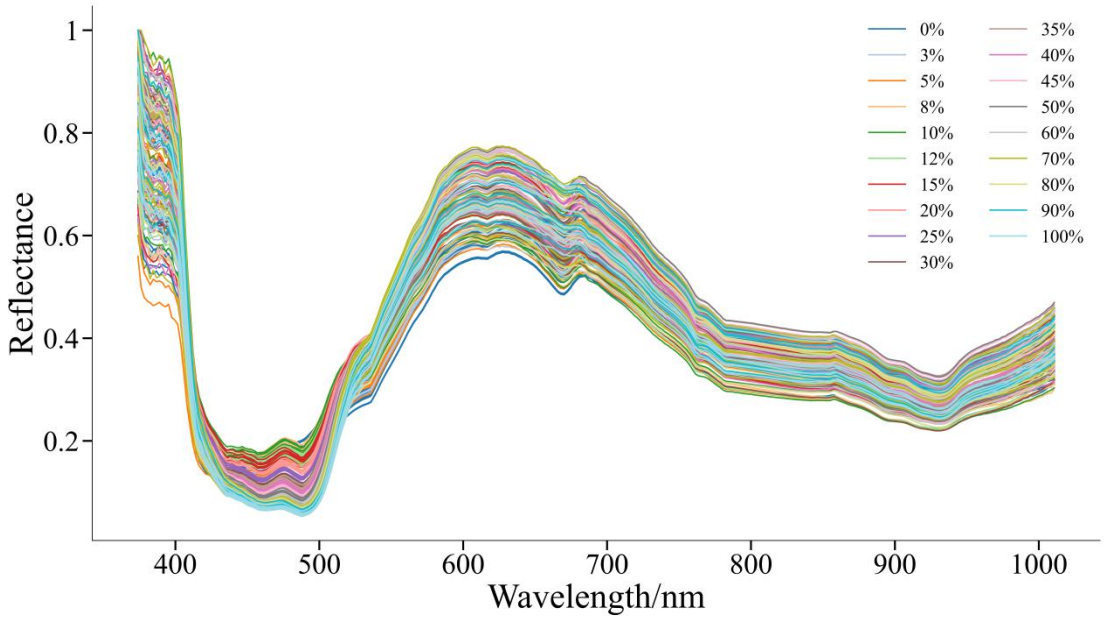
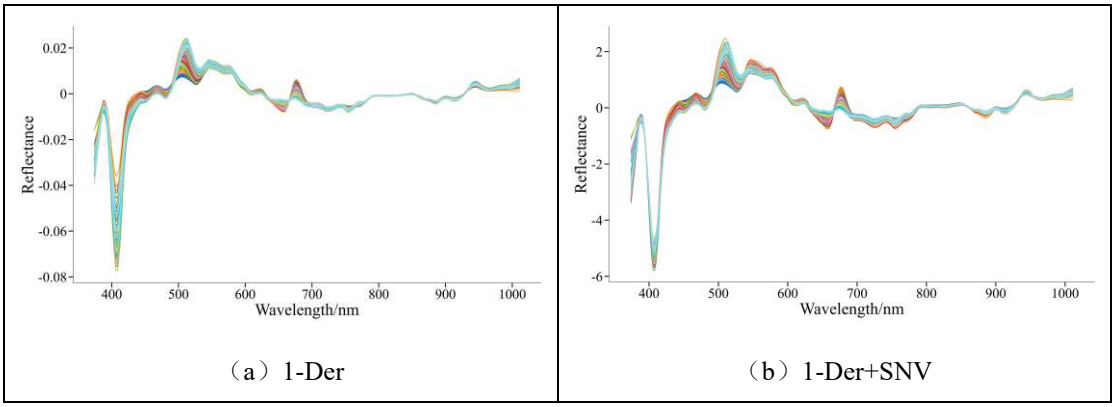


Fig. 2 Spectral reflectance curves of all samples

Spectral preprocessing and feature band extraction

The effect of applying the 10 preprocessing methods to the hyperspectral data is depicted in Fig. 3. 1-Der and 2-Der can highlight the rate of change and curvature characteristics of absorption peaks, respectively, enhancing peak identification ability; when combined with SNV, they can also eliminate scattering and baseline drift. BC can remove baseline drift and restore true absorption peaks. MF and SG can suppress noise and smooth curves, with SG being more suitable for spectra with gentle peak shapes. MSC can eliminate scattering interference among samples, while SNV reduces scattering errors and improves data comparability. Normalization unifies signal dimensions and eliminates intensity scale differences. Various methods optimize spectral data from perspectives such as denoising, baseline correction, scattering elimination, and scale unification, providing high-quality data for subsequent modeling.



（a）1-Der　　（b）1-Der+SNV

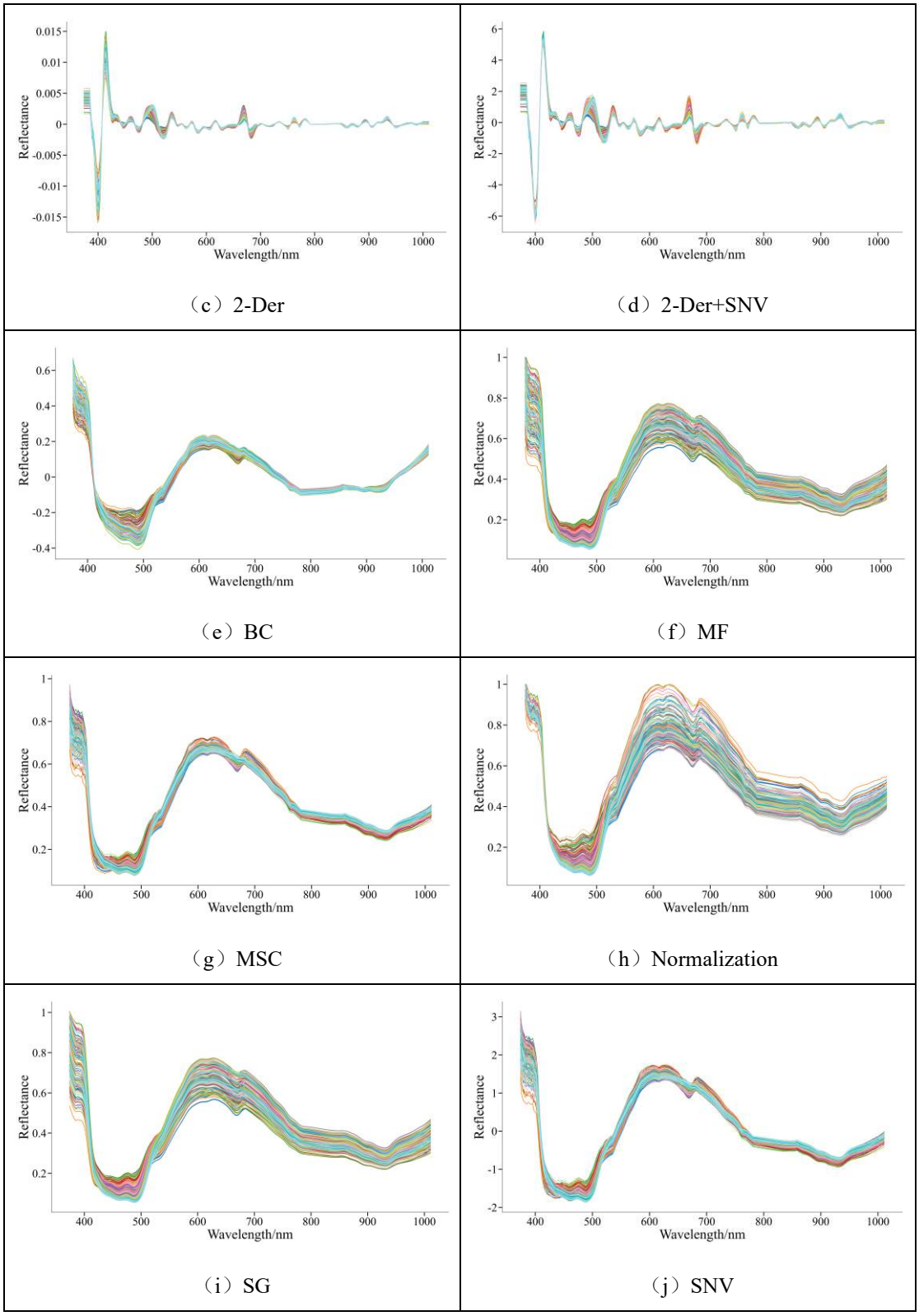


Fig. 3 Ten preprocessed spectral images

Feature wavelengths extracted by SPA yielded optimal wavelength numbers in descending order: BC-SPA > 1-Der-SPA = Origin-SPA > MF-SPA > MSC-SPA = SG-SPA > 2-Der-SNV-SPA = SNV-SPA > 2-Der-SPA > Normalization-SPA > 1-Der-SNV-SPA, corresponding to optimal wavelength counts of 29, 19, 19, 18, 16, 16, 14, 14, 11, 10, and 8, accounting for 9.67%, 6.33%, 6.33%, 6.00%, 5.33%, 5.33%,

4.67%, 4.67%, 3.67%, 3.33%, and 2.67% of the total wavelengths, respectively.

Feature wavelengths extracted by CARS yielded optimal band numbers in descending order: BC-CARS > 1-Der-SNV-CARS > MF-CARS > 2-Der-CARS > 1-Der-CARS = Origin-CARS = SNV-CARS > SG-CARS > MSC-CARS > 2-Der-SNV-CARS = Normalization-CARS, corresponding to optimal band counts of 25, 24, 13, 11, 10, 10, 10, 7, 6, 5, and 5, accounting for 8.33%, 8.00%, 4.33%, 3.67%, 3.33%, 3.33%, 3.33%, 2.33%, 2.00%, 1.67%, and 1.67% of the total wavelengths, respectively.

Qualitative identification of perilla oil authenticity

SVM qualitative modeling analysis

Full-band spectral data contain all wavelength information acquired by hyperspectral imaging. The authenticity identification accuracies of full-band SVM models under different spectral preprocessing methods are shown in Table 1. With the original spectrum (Origin) without any preprocessing, the model accuracies for the calibration set, validation set, and prediction set were 97.03%, 95.04%, and 89.55%, respectively. SNV and 1-Der-SNV methods exhibited the best modeling performance under full bands, with calibration, validation, and prediction set accuracies all reaching 100%, achieving completely accurate identification of adulterated perilla oil samples and pure oils. 2-Der-SNV performed next best, with calibration and prediction set accuracies of 100% and validation set accuracy of 99.02%.

Table 1 Authenticity identification accuracy of full-band SVM models under different spectral preprocessing methods

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	97.03	95.04	89.55
1-Der	92.08	92.1	76.12
1-Der+SNV	100	100	100
2-Der	90.1	90.11	79.1
2-Der+SNV	100	99.02	100
BC	92.08	92.1	76.12
MF	97.03	95.04	89.55
MSC	94.06	91.09	91.04
Normalization	98.02	93.02	95.52
SG	97.03	95.04	89.55
SNV	100	100	100

The authenticity identification accuracies of SVM models using different preprocessing methods combined with SPA feature band extraction are shown in Table 2. The calibration set accuracy of the models was comprehensively improved, and the overall identification stability was optimized. After SPA feature band extraction, except for 1-Der-SNV, the calibration set accuracy of models for all other preprocessing methods, including the unpreprocessed original spectrum, reached 100%. This indicates that SPA can accurately screen feature bands highly correlated with perilla oil authenticity identification, effectively eliminating the interference of redundant information from full-band spectra during model training, allowing the model to perfectly fit the differences in authenticity characteristics of samples

during the calibration stage. This is a core advantage of feature band extraction compared to full-band modeling.

At the validation and prediction set levels, the overall performance of the models was also significantly improved after SPA feature band extraction. All prediction set accuracies were above 90%. Among them, MF, MSC, SG, and SNV methods performed best, with calibration, validation, and prediction set accuracies all reaching 100%, making them the optimal preprocessing combinations under SPA feature band extraction. Even methods like 1-Der and BC, which performed poorly under full bands, saw their prediction set accuracies increase to 98.51% and 100%, respectively, after SPA feature band extraction, with validation set accuracies all above 95%. This demonstrates that SPA can effectively enhance the model's ability to identify unknown samples.

Table 2 Authenticity identification accuracy of SVM models under different spectral preprocessing methods combined with SPA feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	100	97.98	100
1-Der	100	95.99	98.51
1-Der+SNV	95.05	94.06	92.54
2-Der	100	97.06	100
2-Der+SNV	100	97.06	100
BC	100	97.03	100
MF	100	100	100
MSC	100	100	100
Normalization	100	99.02	100
SG	100	100	100
SNV	100	100	100

The authenticity identification accuracies of SVM models using different preprocessing methods combined with CARS feature band extraction are shown in Table 3. The identification performance of the models was generally intermediate between full-band and SPA-based models, and highly sensitive to the suitability of the preprocessing method. The fluctuation range of model accuracy under different preprocessing methods was significantly larger than for bands extracted by SPA. Origin and MF performed best under CARS feature band extraction, with model calibration, validation, and prediction set accuracies reaching 100%, 98.99%, and 100%, respectively. This indicates that the characteristic information of the raw spectrum without preprocessing and the spectrum after simple MF filtering aligns highly with CARS's adaptive feature selection logic. CARS can accurately extract the core features for authenticity identification, enabling completely accurate identification by the model.

In contrast, the performance of models using MSC and Normalization dropped sharply after CARS feature band extraction. The prediction set accuracy for MSC was only 77.61%, and for Normalization was 80.6%, with neither achieving 100% calibration set accuracy. This suggests a severe mismatch between the spectral features processed by these methods and the feature bands selected by CARS. CARS struggled to screen effective authenticity identification information from the preprocessed spectra, leading to a significant reduction in the model's ability to identify unknown samples. 1-Der and SG methods showed similar performance, characterized by good calibration and validation set performance

but a sharp drop in prediction set performance. Both had calibration set accuracies around 98% and validation set accuracies around 98%, but prediction set accuracies were only 88.06%, indicating significant overfitting and poor adaptability to unknown samples.

Table 3 Authenticity identification accuracy of SVM models under different spectral preprocessing methods combined with CARS feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	100	98.99	100
1-Der	98.02	98.04	88.06
1-Der+SNV	100	96.97	100
2-Der	100	97.03	95.52
2-Der+SNV	100	98.04	100
BC	99.01	97.06	95.52
MF	100	98.99	100
MSC	91.09	91.09	77.61
Normalization	89.11	89.13	80.6
SG	98.02	98.04	88.06
SNV	99.01	97	91.04

In summary, feature band extraction can significantly enhance the modeling potential of raw hyperspectral data for perilla oil adulteration identification. Both SPA and CARS can effectively mine key spectral information and greatly improve SVM model accuracy. Among the two feature extraction methods, SPA achieved 100% recognition rates in calibration, validation, and prediction sets under various preprocessing conditions, demonstrating superior model accuracy and stability compared to CARS.

RF qualitative modeling analysis

The authenticity identification accuracies of full-band RF models under different preprocessing methods are shown in Table 4. The calibration set identification accuracy of RF models reached 100% under all preprocessing methods (including the raw spectrum without preprocessing). This result reflects the extremely strong fitting and learning ability of the RF model for the spectral characteristics of the perilla oil adulteration system. Even with noise, baseline drift, scattering effects, and other irrelevant information present in the raw spectrum, the model can still completely identify and fit the core characteristic differences of sample authenticity during the training stage. This is a significant advantage of the RF model as an ensemble learning algorithm in classification modeling.

After 1-Der and 2-Der preprocessing, both the validation and prediction set accuracies of the models reached 100%, achieving completely accurate identification of pure perilla oil and samples adulterated with soybean oil. 1-Der-SNV and 2-Der-SNV performed next best, with validation set accuracies of 98.04% and 99.02%, and prediction set accuracies both at 98.51%. Although not reaching 100%, they still maintained high identification reliability. It is speculated that the reason for this phenomenon is that while combined preprocessing eliminates spectral interference, it might also slightly over-screen some low-dimensional effective features, leading to a slight decrease in the model's ability to recognize

unknown samples.

Table 4 Authenticity identification accuracy of full-band RF models under different spectral preprocessing methods

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	100	90.05	95.52
1-Der	100	100	100
1-Der+SNV	100	98.04	98.51
2-Der	100	100	100
2-Der+SNV	100	99.02	98.51
BC	100	92.04	95.52
MF	100	91.03	97.01
MSC	100	95.1	97.01
Normalization	100	92.1	89.55
SG	100	91.03	97.01
SNV	100	93.11	97.01

The authenticity identification accuracies of RF models using different preprocessing methods combined with SPA feature band extraction are shown in Table 5. After SPA feature band extraction, except for 1-Der-SNV which had a calibration set accuracy of 99.01%, all other preprocessing methods maintained 100% calibration set accuracy. This indicates that the feature bands selected by SPA still retain the core spectral information for perilla oil authenticity identification, and the strong feature fitting ability of the RF model was not significantly affected. At the validation set level, the 1-Der preprocessing performed best, with an accuracy of 98.99%. At the prediction set level, the prediction set accuracies for 1-Der, 2-Der, and Normalization were all 98.51%, representing the optimal level under SPA feature band extraction. However, compared to the 100% achieved under full bands, there was a clear decline. Overall, the compatibility between SPA feature band extraction and the RF model is poor. Only derivative-based preprocessing methods like 1-Der and 2-Der maintained relatively high model identification performance under this dimension, and no overall optimization or improvement in model performance was achieved. This reflects a poor fit between the feature selection logic of the SPA algorithm and the feature learning mechanism of the RF model.

Table 5 Authenticity identification accuracy of RF models under different spectral preprocessing methods combined with SPA feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	100	91.06	91.04
1-Der	100	98.99	98.51
1-Der+SNV	99.01	93.02	92.54
2-Der	100	96.02	98.51
2-Der+SNV	100	98.04	97.01
BC	100	96.05	88.06
MF	100	92.07	95.52

MSC	100	94.09	95.52
Normalization	100	90.11	98.51
SG	100	92.1	89.55
SNV	100	92.1	97.01

The authenticity identification accuracies of RF models using different preprocessing methods combined with CARS feature band extraction are shown in Table 6. The model identification performance was characterized by a slight decrease in calibration set accuracy and a significant overall reduction in validation and prediction set accuracies. At the calibration set level, the accuracies for MF, Normalization, and SG were all below 100%, indicating that CARS's adaptive feature selection might remove a small amount of core information crucial for model fitting for some preprocessed spectra, causing a slight decline in the feature fitting ability of the RF model. At the validation set level, 2-Der-SNV performed best, achieving 100% accuracy, making it the only method to achieve perfect validation identification under this dimension. At the prediction set level, accuracies under all preprocessing methods showed a significant downward trend, generally falling below 97%, which is the core manifestation of the model's declining generalization ability.

Table 6 Authenticity identification accuracy of RF models under different spectral preprocessing methods combined with CARS feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Origin	100	91.12	80.6
1-Der	100	96.08	83.58
1-Der+SNV	100	94.06	92.54
2-Der	100	94.09	94.03
2-Der+SNV	100	100	97.01
BC	100	93.11	95.52
MF	99.01	92.1	85.07
MSC	100	91.09	85.07
Normalization	93.07	88.12	85.07
SG	97.03	94.09	85.07
SNV	100	97	89.55

Comprehensive analysis shows that the RF model exhibits extremely strong feature fitting ability. Under all three modeling approaches, calibration set accuracy was basically 100%, with only a slight decrease for a few preprocessed spectra under CARS feature extraction. The raw full-band spectrum without preprocessing kept prediction set accuracy above 95%. However, after SPA and CARS feature band extraction, prediction set accuracy dropped to 91.04% and 80.6%, respectively, showing a clear downward trend overall. Overall, the complete spectral information of the full band is more suitable for the RF model. SPA and CARS band selection not only failed to improve performance but generally reduced the model's generalization ability.

XGBoost qualitative modeling analysis

The authenticity identification accuracies of full-band XGBoost models under different preprocessing methods are shown in Table 7. After preprocessing with the four derivative-related methods (1-Der, 1-Der-SNV, 2-Der, 2-Der-SNV), the calibration set accuracy of the models all reached 100%, and validation set accuracy was 99.02%. Specifically, 2-Der preprocessing achieved 100% prediction set accuracy, while 1-Der, 1-Der-SNV, and 2-Der-SNV preprocessing achieved 98.51% prediction set accuracy. This indicates that derivative transformation can effectively enhance feature information in the spectra related to differences in sample composition, eliminate background interference. This feature enhancement logic aligns well with the gradient boosting learning mechanism of the XGBoost model, improving the model's fitting and identification capabilities.

Table 7 Authenticity identification accuracy of full-band XGBoost models under different spectral preprocessing methods

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Orign	99.01	92.1	97.01
1-Der	100	99.02	98.51
1-Der+SNV	100	99.02	98.51
2-Der	100	99.02	100
2-Der+SNV	100	99.02	98.51
BC	92.08	92.1	76.12
MF	99.01	92.1	97.01
MSC	100	95.1	98.51
Normalization	100	93.11	89.55
SG	99.01	93.08	97.01
SNV	99.01	96.08	97.01

The authenticity identification accuracies of XGBoost models using different preprocessing methods combined with SPA feature band extraction are shown in Table 8. Compared to the full band, the overall model performance was characterized by relatively high calibration set accuracy, but generally decreased and more volatile prediction set accuracy, with no optimization or improvement achieved. Notably, at the prediction set level, the performance decline was evident for most models. Only 2-Der and Normalization achieved prediction set accuracies of 98.51%.

Table 8 Authenticity identification accuracy of XGBoost models under different spectral preprocessing methods combined with SPA feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Orign	99.01	95.07	94.03
1-Der	100	99.02	94.03
1-Der+SNV	100	91.09	89.55
2-Der	99.01	97.98	98.51
2-Der+SNV	100	97.03	97.01
BC	100	95.04	97.01
MF	99.01	95.04	95.52

MSC	100	92.07	97.01
Normalization	100	94.03	98.51
SG	100	94.09	92.54
SNV	100	96.05	92.54

The authenticity identification accuracies of XGBoost models using different preprocessing methods combined with CARS feature band extraction are shown in Table 9. The model identification performance showed a significant decrease in calibration set accuracy and a substantial overall decline in validation and prediction set accuracies. At the calibration set level, only 2-Der, 2-Der-SNV, and BC achieved 100% accuracy. Compared to full-band and SPA-based models, the feature fitting ability of the models was significantly reduced. At the validation set level, 2-Der-SNV and 1-Der had relatively higher accuracies, both at 98.04%, while other methods were below 96%, indicating poor overall validation performance. At the prediction set level, accuracies for all preprocessing methods showed a substantial downward trend, which is the core manifestation of performance degradation. Only the 2-Der-SNV preprocessing maintained a relatively high prediction accuracy of 98.51%, making it the optimal method under this condition.

Table 9 Authenticity identification accuracy of XGBoost models under different spectral preprocessing methods combined with CARS feature band extraction

Preprocessing method	Calibration accuracy (%)	Validation accuracy (%)	Prediction accuracy (%)
Orign	93.07	93.08	74.63
1-Der	99.01	98.04	82.09
1-Der+SNV	93.07	93.08	74.63
2-Der	100	92.13	89.55
2-Der+SNV	100	98.04	98.51
BC	100	90.14	95.52
MF	92.08	92.1	76.12
MSC	91.09	91.09	77.61
Normalization	89.11	89.13	80.6
SG	91.09	91.09	77.61
SNV	99.01	95.99	89.55

The XGBoost model is more suitable for full-band modeling. SPA and CARS feature extraction can lead to the loss of effective features, resulting in a significant decline in model performance. This model highly relies on complete spectral information. The effectiveness of preprocessing is strongly dimension-dependent, with derivative-based preprocessing being generally optimal. Under full-band conditions, 2-Der performed best, achieving 100% identification accuracy.

Quantitative detection of perilla oil adulteration level

LightGBM model for perilla oil adulteration level

The accuracy of LightGBM models for perilla oil adulteration level under full-band spectra with different

preprocessing methods is shown in Table 10. Under full bands, derivative-based preprocessing performed excellently. The model with 2-Der preprocessing was optimal, with R^2_{CV}, RMSECV, R^2_{Pred}, and RMSEP values of 0.9752, 0.0468, 0.9879, and 0.0311, respectively. This indicates that derivative transformation effectively enhances feature information in the spectra related to the adulteration level, aligning well with the learning mechanism of the LightGBM model.

Table 10 Prediction accuracy of perilla oil adulteration level using full-band hyperspectral LightGBM models under different spectral preprocessing methods

Preprocessing Method	R^2_{CV}	RMSECV	R^2_{Pred}	RMSEP
Origin	0.9188	0.0889	0.9581	0.0605
1-Der	0.9672	0.0538	0.9845	0.0387
1-Der+SNV	0.9553	0.0615	0.9882	0.0339
2-Der	0.9752	0.0468	0.9879	0.0311
2-Der+SNV	0.9729	0.0516	0.9852	0.0347
BC	0.9152	0.0835	0.9609	0.0613
MF	0.9250	0.0844	0.9504	0.0660
MSC	0.9254	0.0817	0.9732	0.0507
Normalization	0.8834	0.1050	0.9076	0.0919
SG	0.9166	0.0906	0.9500	0.0661
SNV	0.9234	0.0830	0.9760	0.0478

The accuracy of LightGBM models for perilla oil adulteration level using SPA feature bands with different preprocessing methods is shown in Table 11. Based on hyperspectral data after SPA feature band extraction, the raw spectrum model still maintained good detection performance. The impact trend of different preprocessing methods on model performance was generally consistent with that under full bands. 1-Der preprocessing was the optimal method, with model R^2_{CV}, RMSECV, R^2_{Pred}, and RMSEP values of 0.9637, 0.0591, 0.9826, and 0.0379, respectively. Overall, after dimensionality reduction using SPA feature bands, model performance under various preprocessing methods was close to that of the full bands, with no significant performance loss observed.

Table 11 Prediction accuracy of perilla oil adulteration level using SPA-band hyperspectral LightGBM models under different spectral preprocessing methods

Preprocessing Method	R^2_{CV}	RMSECV	R^2_{Pred}	RMSEP
Origin	0.9186	0.0871	0.9590	0.0627
1-Der	0.9637	0.0591	0.9826	0.0379

1-Der+SNV	0.9559	0.0681	0.9702	0.0467
2-Der	0.9606	0.0618	0.9702	0.0478
2-Der+SNV	0.9405	0.0805	0.9702	0.0425
BC	0.9261	0.0838	0.9644	0.0534
MF	0.9094	0.0900	0.9640	0.0578
MSC	0.9189	0.0878	0.9758	0.0471
Normalization	0.9010	0.0942	0.9183	0.0889
SG	0.9115	0.0908	0.9641	0.0581
SNV	0.9359	0.0791	0.9630	0.0564

The accuracy of LightGBM models for perilla oil adulteration level using CARS feature bands with different preprocessing methods is shown in Table 12. The cross-validation performance of the raw spectrum model improved compared to full-band and SPA-band models. 2-Der preprocessing remained optimal, with all model evaluation metrics being the best among all combinations: R^2_{CV} = 0.9628, RMSECV = 0.0608, R^2_{Pred} = 0.9863, RMSEP = 0.0273.

Table 12 Prediction accuracy of perilla oil adulteration level using CARS-band hyperspectral LightGBM models under different spectral preprocessing methods

Preprocessing Method	R^2_{CV}	RMSECV	R^2_{Pred}	RMSEP
Origin	0.9359	0.0771	0.9585	0.0564
1-Der	0.9378	0.0722	0.9816	0.0395
1-Der+SNV	0.9581	0.0623	0.9843	0.0342
2-Der	0.9628	0.0608	0.9863	0.0273
2-Der+SNV	0.9558	0.0664	0.9832	0.0362
BC	0.8910	0.0990	0.9460	0.0703
MF	0.9382	0.0777	0.9718	0.0471
MSC	0.8109	0.1262	0.9325	0.0821
Normalization	0.6719	0.1819	0.8451	0.0992
SG	0.4961	0.2245	0.6438	0.1494
SNV	0.8639	0.1126	0.9506	0.0645

Stacking model for perilla oil adulteration level

The accuracy of Stacking models for perilla oil adulteration level is shown in Table 13. Compared to the original LightGBM models, the Stacking ensemble models did not achieve performance improvement in any of the corresponding preprocessing-band extraction combinations. Instead, they showed varying degrees of accuracy decrease and error increase. Even the optimal combination for the Stacking model had lower prediction accuracy than the optimal result of the original LightGBM model under the same band condition. This indicates that in the perilla oil adulteration detection system of this study, the Stacking ensemble model, constructed based on the optimal preprocessing for the LightGBM model, did not leverage the advantages of ensemble modeling and performed worse than the single LightGBM model. The reason for the decreased model capability might be that in the Stacking framework using PLS and Ridge as base models, the base models, being primarily linear fitting models, provided limited incremental information and failed to complement the meta-model effectively. Additionally, the two-layer modeling structure increased model complexity and the risk of noise propagation, leading to a decline in overall generalization performance.

Table 13 Accuracy of Stacking models for perilla oil adulteration level

Preprocessing method	Feature extraction method	R^2_{CV}	RMSECV	R^2_{Pred}	RMSEP
1-Der+SNV		0.9364	0.0744	0.9692	0.0530
2-Der	Full-band	0.9532	0.0670	0.9822	0.0423
2-Der+SNV		0.9427	0.0733	0.9690	0.0559
1-Der		0.9376	0.0752	0.9710	0.0490
1-Der+SNV	SPA	0.9598	0.0650	0.9758	0.0420
2-Der+SNV		0.9636	0.0638	0.9694	0.0413
1-Der+SNV		0.9550	0.0638	0.9792	0.0453
2-Der	CARS	0.9659	0.0579	0.9800	0.0449
2-Der+SNV		0.9454	0.0723	0.9758	0.0490

In summary, 2-der-CARS-LightGBM is identified as the best model for detecting perilla oil adulteration levels. The fitting and residual distribution characteristics of this model for the cross-validation set and prediction set are shown in Figures 4, 5, 6, and 7. In the cross-validation set scatter plot, the correlation coefficient R^2_{CV} between predicted and actual adulteration levels reached 0.9628, with RMSECV of 0.0608. All data points are tightly distributed around the fitting line without obvious deviation, indicating good fitting ability and stability of the model during the cross-validation stage. In the prediction set scatter plot, the correlation coefficient R^2_{Pred} further improved to 0.9863, with RMSEP as low as 0.0273. The data points align even more closely with the fitting line, demonstrating the model's excellent quantitative prediction accuracy for perilla oil adulteration levels. In both the cross-validation set residual plot and the prediction set residual plot, the residual values are randomly and uniformly distributed around zero, with no obvious systematic bias. Moreover, the fluctuation range of prediction set residuals is significantly smaller than that of the cross-validation set, and the absolute residual values are generally at a low level. This indicates that the model has no systematic error, the prediction results

for unknown samples are reliable, and it can accurately reflect the actual adulteration level of perilla oil. This further validates the applicability and superiority of the 2-Der-CARS-LightGBM model in the quantitative detection of perilla oil adulteration levels.

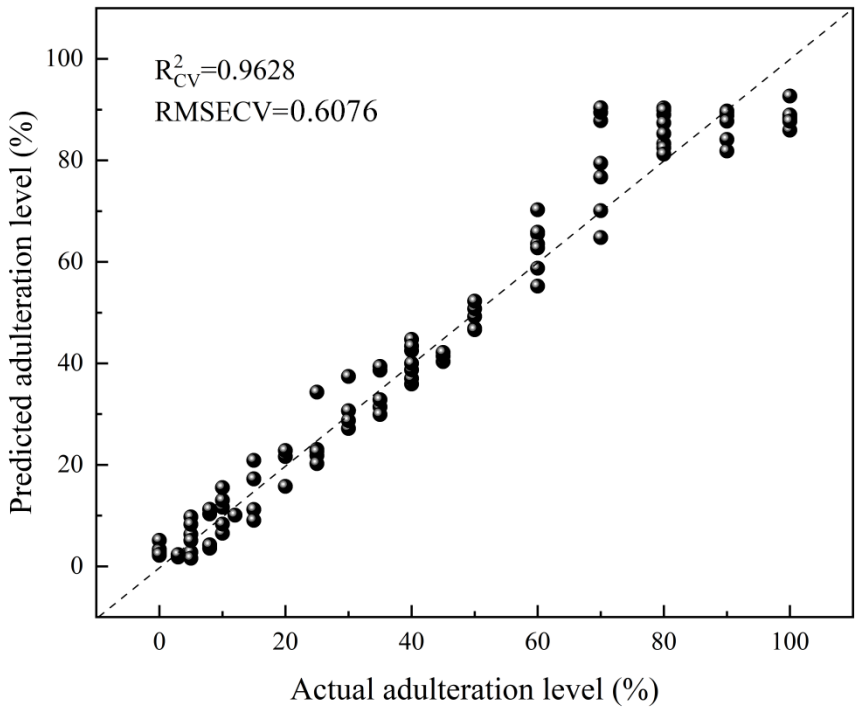


Fig. 4 Cross-validation scatter plot of actual and predicted adulteration levels of perilla oil based on the 2-Der-CARS-LightGBM model

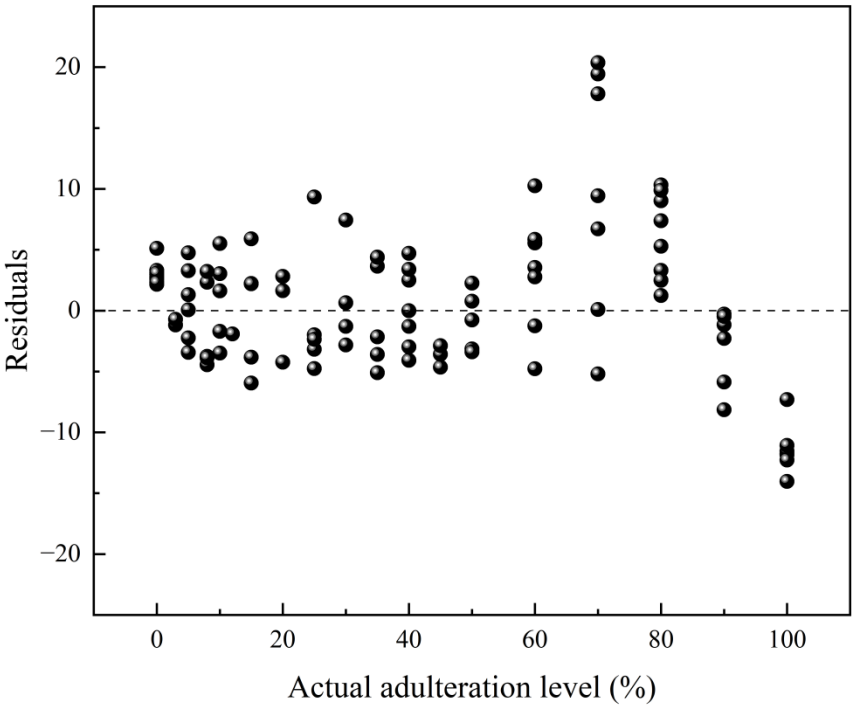


Fig. 5 Cross-validation residual plot of adulteration level prediction for perilla oil based on the 2-Der-CARS-LightGBM model

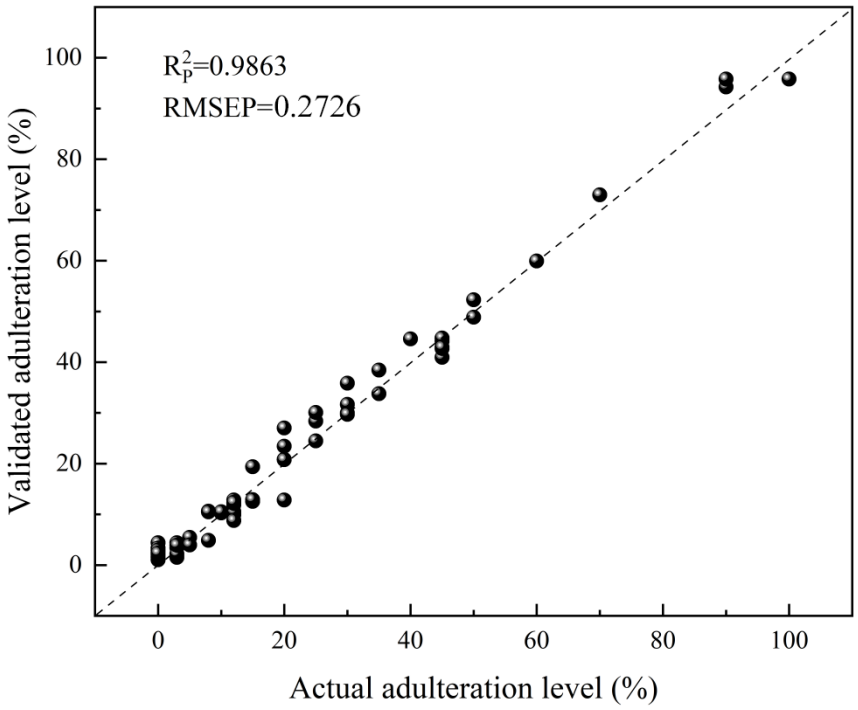


Fig. 6 Prediction set scatter plot of actual and predicted adulteration levels of perilla oil based on the 2-Der-CARS-LightGBM model

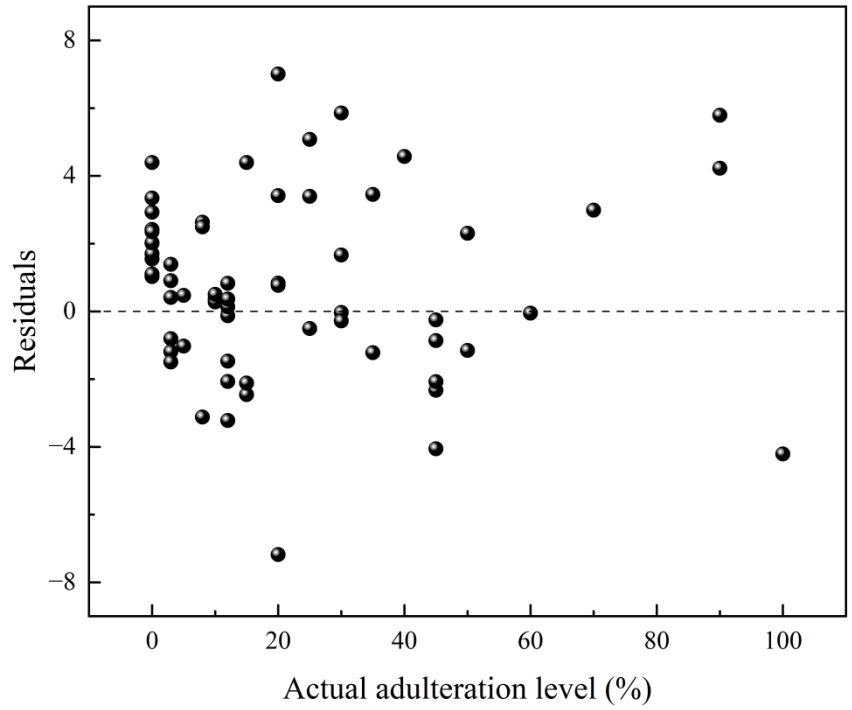


Fig. 7 Prediction set residual plot of adulteration level prediction for perilla oil based on the 2-Der-CARS-LightGBM model

Conclusion

This study successfully achieved authenticity identification and quantitative detection of adulteration

levels in perilla oil adulterated with soybean oil by combining Vis-NIR hyperspectral imaging technology with chemometric methods. It clarified the suitability patterns among different spectral preprocessing, feature extraction methods, and machine learning models. For qualitative identification, SVM and RF models established after preprocessing achieved 100% accuracy in distinguishing perilla oil authenticity. The SVM model demonstrated the best overall performance, with SPA feature band extraction significantly optimizing its identification performance and stability. In contrast, RF and XGBoost models were highly dependent on complete full-band spectral information, and feature extraction reduced their generalization ability. Derivative-based and SNV-related preprocessing methods were generally the optimal choices for qualitative modeling. For quantitative detection, the LightGBM model outperformed the Stacking ensemble model. The 2-Der-CARS-LightGBM combination was identified as the optimal detection model, with R^2_{CV} reaching 0.9628 and RMSEP as low as 0.0273. This model enables accurate quantitative prediction of perilla oil adulteration levels, with a residual distribution free of systematic error and excellent predictive reliability for unknown samples. The study only focused on the binary adulteration of perilla oil with soybean oil, and the detection effect on multi-component adulteration needs to be further verified. Future research can expand the adulteration types of perilla oil, and combine HSI with deep learning models (e.g., CNN, Transformer) to further improve the detection accuracy and speed. The results of this study provide an efficient technical method and theoretical support for the rapid quality detection of perilla oil, and also offer a reference approach for adulteration identification research in fatty foods.

Author contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Zhen Dong, Fujian Xu, Haijuan Tian, Taogang Hu, Jingwei Zhang, Weiqi Hu, and Changcheng Liu. The first draft of the manuscript was written by Zhen Dong and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by Science and Technology Development Plan Project of Jilin Province, China (No. 20250202017NC).

Declarations

Conflict of interest The authors declare no competing interests.

References

Aqeel M, Munawar H, Sohaib A, Khan KB, Deng Y (2026) Spectral band selection for nondestructive detection of edible oil adulteration using hyperspectral imaging and chemometric analysis. Journal of Food Measurement and Characterization 20(2):1482-1503. https://doi.org/10.1007/s11694-025-03805-6

Aqeel M, Sohaib A, Iqbal M, Rehman HU, Rustam F (2024) Hyperspectral identification of oil adulteration using machine learning techniques. Current Research in Food Science 8:100773. https://doi.org/https://doi.org/10.1016/j.crfs.2024.100773

Aykas D-P, Menevseoglu A, Gunes N (2022) Information theory and machine learning based authentication of flaxseed oil using portable and handheld vibrational spectroscopy sensors. Chinese Journal of Analytical Chemistry 50(4):100064. https://doi.org/https://doi.org/10.1016/j.cjac.2022.100064

Cai J, Li X, Tan Z, Peng S (2021) An assembly-level neutronic calculation method based on LightGBM

algorithm. Annals of Nuclear Energy 150:107871. https://doi.org/https://doi.org/10.1016/j.anucene.2020.107871

Cao Y, Jia N, Li X, Zhu J, Li B (2022) Quantitative analysis of typical fluoroquinolone antibiotics in livestock feed using terahertz spectrum. Journal of Food Process Engineering 45(9):e14107. https://doi.org/https://doi.org/10.1111/jfpe.14107

Castro W, Mejía J, De-La-Torre M, Acevedo-Juárez B, Tech ARB, Avila-George H (2023) Radial grid reflectance correction for hyperspectral images of fruits with rounded surfaces. Computers and Electronics in Agriculture 213:108179. https://doi.org/https://doi.org/10.1016/j.compag.2023.108179

Cui J, Li K, Hao J, Dong F, Wang S, Rodas-González A et al. (2022) Identification of Near Geographical Origin of Wolfberries by a Combination of Hyperspectral Imaging and Multi-Task Residual Fully Convolutional Network. Foods 11(13):1936. https://doi.org/ https://doi.org/10.3390/foods11131936

Deng J, Zhang X, Li M, Jiang H, Chen Q (2022) Feasibility study on Raman spectra-based deep learning models for monitoring the contamination degree and level of aflatoxin B1 in edible oil. Microchemical Journal 180:107613. https://doi.org/https://doi.org/10.1016/j.microc.2022.107613

Duraipandian S, Petersen JC, Lassen M (2019) Authenticity and Concentration Analysis of Extra Virgin Olive Oil Using Spontaneous Raman Spectroscopy and Multivariate Data Analysis. Applied Sciences 9(12):2433. https://doi.org/https://doi.org/10.3390/app9122433

Farber C, Kurouski D (2022) Raman Spectroscopy and Machine Learning for Agricultural Applications: Chemometric Assessment of Spectroscopic Signatures of Plants as the Essential Step Toward Digital Farming. Frontiers in Plant Science 13 - 2022. https://doi.org/10.3389/fpls.2022.887511

He W, He H, Wang F, Wang S, Li R, Chang J et al. (2022) Rapid and Uninvasive Characterization of Bananas by Hyperspectral Imaging with Extreme Gradient Boosting (XGBoost). Analytical Letters 55(4):620-633. https://doi.org/10.1080/00032719.2021.1952214

Hu Y, Sun J, Zhan C, Huang P, Kang Z (2023) Identification and quantification of adulterated Tieguanyin based on the fluorescence hyperspectral image technique. Journal of Food Composition and Analysis 120:105343. https://doi.org/https://doi.org/10.1016/j.jfca.2023.105343

Huang Z-M, Xin J-X, Sun S-S, Li Y, Wei D-X, Zhu J et al. (2021) Rapid Identification of Adulteration in Edible Vegetable Oils Based on Low-Field Nuclear Magnetic Resonance Relaxation Fingerprints. Foods 10(12):3068. https://doi.org/https://doi.org/10.3390/foods10123068

Jamwal R, Amit, Kumari S, Sharma S, Kelly S, Cannavan A et al. (2021) Recent trends in the use of FTIR spectroscopy integrated with chemometrics for the detection of edible oil adulteration. Vibrational Spectroscopy 113:103222. https://doi.org/https://doi.org/10.1016/j.vibspec.2021.103222

Jermwongruttanachai P, Pathaveerat S, Noypitak S (2024) Quantification of the adulteration concentration of palm kernel oil in virgin coconut oil using near-infrared hyperspectral imaging. Journal of Integrative Agriculture 23(1):298-309. https://doi.org/https://doi.org/10.1016/j.jia.2023.08.002

Jiao Z, Song L, Zhang Y, Dai J, Liu Y, Zhang Q et al. (2025) A comparative study of fluorescence hyperspectral imaging and FTIR spectroscopy combined with chemometrics for the detection of extra virgin olive oil adulteration. Journal of Food Measurement and Characterization 19(3):1761-1776. https://doi.org/10.1007/s11694-024-03069-6

Kangwan N, Pratchayasakul W, Kongkaew A, Pintha K, Chattipakorn N, Chattipakorn SC (2021) Perilla Seed Oil Alleviates Gut Dysbiosis, Intestinal Inflammation and Metabolic Disturbance in Obese-Insulin-Resistant Rats. Nutrients 13(9):3141. https://doi.org/https://doi.org/10.3390/nu13093141

Li B, Han Z, Wang Q, Sun Z, Liu Y (2022) Study on Qualitative Impact Damage of Loquats Using Hyperspectral Technology Coupled with Texture Features. Foods 11(16):2444. https://doi.org/https://doi.org/10.3390/foods11162444

Malavi D, Raes K, Van Haute S (2024) Integrating near-infrared hyperspectral imaging with machine learning and feature selection: Detecting adulteration of extra-virgin olive oil with lower-grade olive oils and hazelnut oil. Current Research in Food Science 9:100913. https://doi.org/https://doi.org/10.1016/j.crfs.2024.100913

N. K S, Patil A, Reddy DR, K RU, Basavaraj P (2024) Predictive Analytics for Milk Quality Using Random Forest (RF) Algorithm. 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS):1-5. https://doi.org/10.1109/CSITSS64042.2024.10816917

Niu B, Zhang H, Zhou G, Zhang S, Yang Y, Deng X et al. (2021) Safety risk assessment and early warning of chemical contamination in vegetable oil. Food Control 125:107970. https://doi.org/https://doi.org/10.1016/j.foodcont.2021.107970

Prakash R, Shantha Kumar SK, Umar A, Chandan H (2024) Predicting food adulterants in milk using Support Vector Machine (SVM). International Journal of Science and Research Archive 13(1):2257-2268. https://doi.org/10.30574/ijsra.2024.13.1.1848

Pu H, Yu J, Sun D-W, Wei Q, Wang Z (2023) Feature construction methods for processing and analysing spectral images and their applications in food quality inspection. Trends in Food Science & Technology 138:726-737. https://doi.org/https://doi.org/10.1016/j.tifs.2023.06.036

Rogeboz P, Latado H, Sharma A, Chaubey N, Kadian S, Chavez E et al. (2024) Oil Adulteration Evaluation Using High Performance Thin Layer Chromatography. Food Analytical Methods 17(9):1336-1347. https://doi.org/10.1007/s12161-024-02659-w

Shang Y, Bao L, Bi H, Guan S, Xu J, Gu Y et al. (2024) Authenticity Discrimination and Adulteration Level Detection of Camellia Seed Oil via Hyperspectral Imaging Technology. Food Analytical Methods 17(3):450-463. https://doi.org/10.1007/s12161-024-02577-x

Su N, Weng S, Wang L, Xu T (2021) Reflectance Spectroscopy with Multivariate Methods for Non-Destructive Discrimination of Edible Oil Adulteration. Biosensors 11(12):492. https://doi.org/https://doi.org/10.3390/bios11120492

Tang Y, Wang F, Zhao X, Yang G, Xu B, Zhang Y et al. (2023) A nondestructive method for determination of green tea quality by hyperspectral imaging. Journal of Food Composition and Analysis 123:105621. https://doi.org/https://doi.org/10.1016/j.jfca.2023.105621

Wei H, Yang D, Mao J, Zhang Q, Cheng L, Yang X et al. (2023) Accurate quantification of TAGs to identify adulteration of edible oils by ultra-high performance liquid chromatography-quadrupole-time of flight-tandem mass spectrometry. Food Research International 165:112544. https://doi.org/https://doi.org/10.1016/j.foodres.2023.112544

Xia J, Wang Y, Li X, Liu L, Zhang P, Dai W et al. (2025) The mechanism of perilla oil in regulating lipid metabolism. Food Chemistry 476:143318. https://doi.org/https://doi.org/10.1016/j.foodchem.2025.143318

Zhao AB, Cheng T (2022) Stock return prediction: Stacking a variety of models. Journal of Empirical

Finance 67:288-317. https://doi.org/https://doi.org/10.1016/j.jempfin.2022.04.001