Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 7, No. 3, 2025

Origin Identification of Jujubes Based on Hyperspectral Imaging Technology

Lei Fei, Haili Yang*, Jiahao Zeng, Xilong Liao, Hao Xia, Fuhao Han, Yuqi Zhou

School of Mechanical Engineering, Sichuan University of Science & Engineering, Yibin 644000, Sichuan Province, China

* **Corresponding author**: Haili Yang

Abstract: To achieve non-destructive, rapid, batch detection and accurate identification of jujube origins, a recognition model for jujube origin was constructed in this study based on hyperspectral technology combined with machine learning methods. Jujubes from six different provinces—Liaoning, Xinjiang, Yunnan, Hebei, Hubei, and Guangxi—were selected as research subjects, and their hyperspectral data were collected. The original spectral data were preprocessed using multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky-Golay smoothing (SG), first derivative (FD), and second derivative (SD) methods. Subsequently, support vector machine (SVM), random forest (RF), and partial least squares discriminant analysis (PLS-DA) were employed to establish jujube origin recognition models, which were evaluated using confusion matrix visualization and F1 score. The results showed that the SNV-SVM combination model achieved the highest F1 score of 99.44% and a classification accuracy of 99.43%, which were 0.15% and 0.31% higher in F1 score, and 0.16% and 0.32% higher in accuracy, respectively, compared to the SD-PLS-DA and SD-RF combination models. In summary, the combination of hyperspectral technology and the SNV-SVM method enables non-destructive, rapid, and batch identification of jujubes from different origins, providing strong technical support for jujube research and production.

Keywords: Hyperspectral Technology; Machine Learning; Jujube; Origin Identification.

1. Introduction

As a traditional medicinal and edible plant resource, jujubes are widely favored [1]. Taxonomically belonging to the genus Ziziphus in the family Rhamnaceae, jujubes are also commonly named "red dates" or "Guanzao", with their plants characterized as small trees or shrubs. The fruits contain various amino acids, carbohydrates, and vitamins, and these nutritional components and bioactive substances have been proven to enhance human immune function [2]. Practices have shown that due to the spatial heterogeneity of ecological factors such as light intensity, temperature range, and precipitation distribution, the quality traits of jujubes produced in different geographical regions exhibit significant differences [3]. In market transactions, origin labeling has become a key indicator for evaluating the quality grade of jujubes. However, false labeling of origin information frequently occurs in the current circulation field, especially on e-commerce platforms [4]. Such behavior not only disrupts the normal market operation mechanism but also infringes on the legitimate rights and interests of consumers. In the existing technical system, the identification of jujube origins mainly relies on high-performance liquid chromatography (HPLC) [5] and near-infrared spectroscopy (NIRS) [6]. Although these chemical detection methods have high accuracy, they have obvious limitations: prolonged detection cycles, excessive dependence on precision instruments, and low work efficiency. Thus, traditional methods are difficult to fully meet the market demand for rapid and non-destructive identification technologies. Hyperspectral imaging (HSI) technology, which has important application value in the field of digital imaging, is essentially an advanced technical system integrating spectral analysis and spatial information acquisition. This technology enables simultaneous collection of continuous spectral information of samples and spatial distribution data of measured objects [7][8]. The adoption of non-contact detection methods significantly improves the efficiency of batch detection while maintaining sample integrity. Practices have shown that the application value of HSI technology is increasingly prominent in the fields of agricultural and food quality monitoring, as well as the origin identification of traditional Chinese medicinal materials. He Wenwen et al. [9] combined improved convolutional neural networks (CNN) with hyperspectral imaging in the purity identification of corn seeds, and the optimized CNN model showed superiority in multiple indicators such as precision and recall. Peng Jianheng et al. [10] achieved breakthrough progress in the authenticity identification of lotus root starch. The collaborative application of hyperspectral imaging technology and deep learning frameworks resulted in a test accuracy of 99.69%. The isolation forest algorithm was used for outlier elimination, and feature wavelength extraction was realized through methods such as CARS, BOSS, and CAMM. A CAMM-CNN model for visual identification of starch adulteration was constructed. Sun Xue et al. [11] achieved remarkable results in the cross-regional identification of Poria cocos samples. After spectral collection and preprocessing of samples from 7 provinces, two modeling methods (PLS-DA and LinearSVC) were adopted. The FD+LinearSVC combination performed excellently in classification tasks at the provincial, county, and batch levels, with all indicators exceeding 97%. Thus, the model system has extremely prominent applicability in non-destructive detection scenarios.

At the current stage, the practical application of hyperspectral imaging (HSI) technology in the research direction of jujube origin identification remains insufficient. Against this background, this study attempts to integrate hyperspectral image acquisition with machine learning algorithms to conduct distinguishability verification for jujube samples from different sources. Practices have shown that the exploration of this method is expected to provide technical support for establishing an accurate jujube origin

traceability system. Thus, the innovative combination of these technologies can not only effectively improve the efficiency of quality supervision for jujube products but also significantly enhance their market competitiveness and economic benefits, ultimately promoting the sustainable development of the entire jujube industrial chain.

2. Materials and Methods

2.1. Samples

To ensure the geographical representativeness and consistency of the research samples, this study uniformly collected jujube samples from six different provinces (Liaoning, Xinjiang, Yunnan, Hebei, Hubei, and Guangxi) in the same year. Detailed information on the sample origins is shown in Table 1. During sample selection, all jujubes were directly purchased from local individual households in the producing areas. By directly collaborating with farmers or operators, the product sources were strictly traced to ensure the authenticity and reliability of the origin information for each jujube. As shown in Figure 1, the jujube samples from different origins included 210 samples per origin, grouped into 35 sets of 6 samples each. A total of 1260 jujube samples from the six origins were scanned to construct the sample library. The samples were divided into a training set (882 samples) and a test set (378 samples) at a ratio of 7:3.

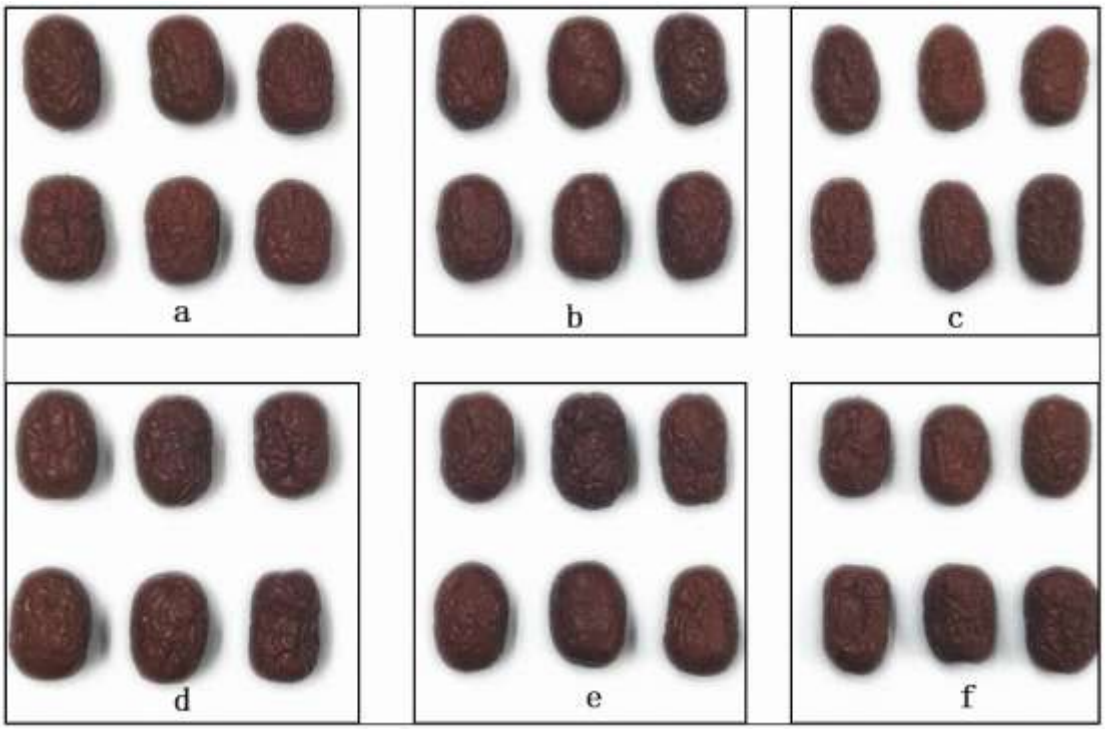


Figure 1 Jujube samples from different producing areas (1. Hezhou, Guangxi; 2. Hotan, Xinjiang; 3. Cangzhou, Hebei; 4. Shiyan, Hubei; 5. Tengchong; Yunnan 6. Chaoyang, Liaoning)

Table 1 Information on the producing areas of jujube samples

No	Region	Origin
S1	Northeast China	Chaoyang (City, Liaoning)
S2	Northwest China	HOtan (City, Xinjang)
S3	Southwest China	Tengchong (City, Yunnan)
S4	North China	Cangzhou (City, Hebei)
S5	Central China	Shiyan (City, Hubei)
S6	South China	Hezhou (City, Guangxi)

2.2. Instruments and Equipment

One GaiaField - N17E - HR hyperspectral camera (spectral range of 900 - 1700 nm, spectral resolution of 5 nm, and 512 bands), provided by Jiangsu Dualix Spectral Imaging Technology Co., Ltd.; four FSL halogen light sources (power of 55 W), provided by Foshan Electrical and Lighting Co., Ltd.; one ASUS computer (equipped with the hyperspectral data acquisition software SpecVIEW), provided by ASUSTek Computer Inc., Taiwan, China; one BCE55I - 10CN electronic balance (measurement accuracy of 0.01 mg), provided by Beijing Sartorius Instrument System Co., Ltd.

2.3. Data Processing

2.3.1. Acquisition and Correction of Hyperspectral Images

The hyperspectral imaging system is shown in Figure 2. The samples were placed on a black paper sheet (10 cm × 10 cm) and flattened with a glass plate to ensure a smooth and compact surface. They were then positioned at the center of the acquisition platform, 35.5 cm away from the lens of the hyperspectral imager. Hyperspectral image data with a resolution of 640×666 pixels × 512 bands were collected. The hyperspectral acquisition parameters were set as follows: R/G/B wavelengths of 954.15 nm/927.55 nm/904.28 nm, gain of 2, and exposure time of 2 ms.

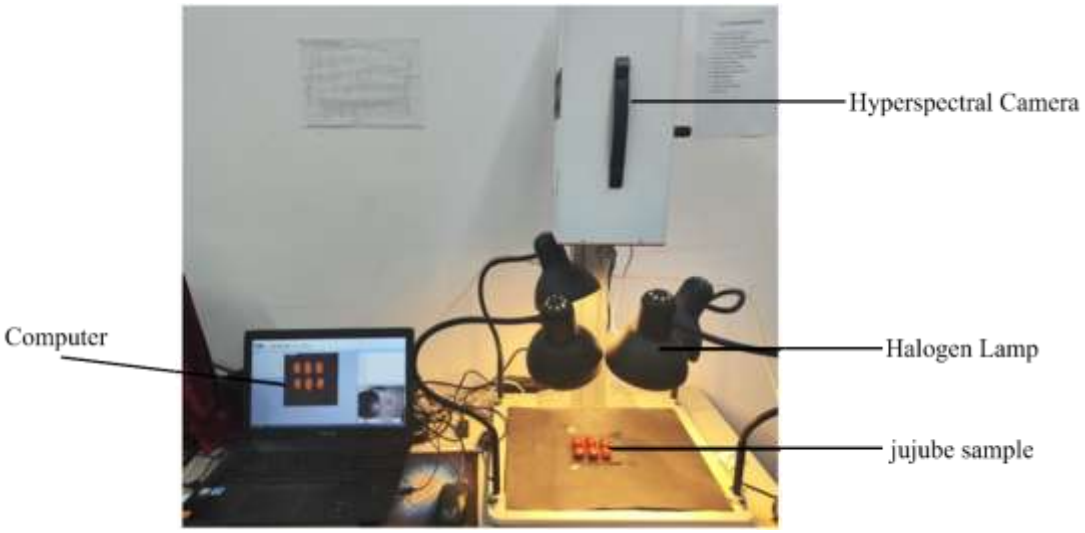


Figure 2 Hyperspectral Imaging System

Due to the dark current in the camera and the uneven intensity distribution of the system light source, the collected hyperspectral images are affected by significant noise and cannot reflect the true characteristics of the samples. Therefore, whiteboard and blackboard images need to be collected for black-white correction to reduce the impact of camera noise and uneven light source intensity on the data, thereby obtaining relatively stable spectral reflectance [12]. The original images were corrected using the black-white correction method described in Equation (1).

$$L_C = \frac{L_0 - L_d}{L_W - L_d} \quad (1)$$

Wherein:

L_C ——R is the corrected hyperspectral image;

L_0 ——R raw is the hyperspectral image before correction;

L_W ——R white is the white reference image;

L_d ——R black is the black reference image;

2.3.2. Hyperspectral Regions of Interest (ROI)

To accurately extract the spectral data of jujubes, this study applied Otsu's thresholding algorithm to the corrected hyperspectral images for segmenting samples and backgrounds. By combining image binarization with morphological operations, background interference was eliminated to achieve precise segmentation of hyperspectral images and extract Regions of Interest (ROI). Each jujube was regarded as an independent ROI, and 6 ROIs were extracted from each sample. For each ROI, the spectral reflectance of all internal pixels was extracted, and the average spectral reflectance within the ROI was calculated, which was used as the one-dimensional spectral data of the jujube.

2.3.3. Preprocessing of Spectral Data

During the acquisition of spectral data, it is often affected by factors such as noise, stray light, and baseline drift, which thereby impairs the modeling effect. Therefore, it is necessary to preprocess the raw spectral data before modeling to improve the classification effect and stability of the model [13]. In this study, five spectral preprocessing methods were adopted, namely Multiplicative Scatter Correction (MSC), Savitzky - Golay (SG) convolution smoothing, Standard Normal Variable (SNV), First Derivative (FD), and Second Derivative (SD). These methods were separately applied to process the raw hyperspectral data, and the classification accuracy rates after different preprocessing steps were compared.

2.3.4. Data Analysis

All experimental data were processed and analyzed using Python 3.8 and its related scientific computing libraries (e.g., NumPy, Pandas, scikit-learn, matplotlib, etc.). First, the collected hyperspectral data were preprocessed, and then the spectral data under different preprocessing methods were input into classification models such as Support Vector Machine (SVM), Random Forest (RF), and Partial Least Squares Discriminant Analysis (PLS-DA) for training and testing. Model performance was evaluated using metrics including accuracy, precision, recall, and F1 score, and a confusion matrix was used for visual analysis of classification results. In addition, to further verify the discrimination ability of the models, the t-SNE method was adopted for dimensionality reduction and visualization of high-dimensional features. All experiments were repeated multiple times under the same hardware environment to ensure the reliability and reproducibility of the results.

2.4. Construct a classification model for jujube origin identification

2.4.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a method widely applied in the field of supervised learning [14]. Its core idea lies in maximizing the class margin by constructing an optimal hyperplane in the feature space to achieve high efficiency in classification and regression tasks. Through the use of kernel functions, input data are mapped to a high-dimensional space, enabling effective separation of jujube samples from different origins by hyperplanes defined by support vectors. After optimizing the positions of support vectors, the classification margin is improved [15]. In this study, the grid search method was adopted to optimize model parameters, including the selection of kernel function type (Kernel), the determination of penalty parameter (C), and the setting of radial basis kernel parameter (gamma). Five-fold cross-validation was implemented to evaluate parameter combinations, thereby reducing the randomness caused by random data segmentation, and the final optimal model parameters were determined based on this.

2.4.2. Random Forest (RF)

Random Forest (RF), as an ensemble learning method, forms a "forest" by combining multiple decision trees [16]. The prediction results of all decision trees are integrated through majority voting to complete the classification task. If the number of decision trees (n_estimators) is set too small, the stability and generalization ability of the model will be affected, and overfitting or underfitting phenomena are likely to occur. Examples show that when the number of decision trees is too large, the computational load increases significantly, the operation efficiency decreases, and the risk of system crash rises. Thus, the reasonable selection of the number of decision trees plays a key role in improving the accuracy and stability of the model. In this study, the grid search method was used for parameter optimization. The parameters adjusted separately on the training set included the number of decision trees (n_estimators), the maximum depth of a single tree (max_depth), the minimum number of samples required to split an internal node (min_samples_split), and the minimum number of samples required for a leaf node (min_samples_leaf). Five-fold cross-validation was applied to evaluate parameter combinations, reducing the impact of randomness caused by random data segmentation, and the determination of optimal model parameters was finally completed based on this.

2.4.3. Partial Least Squares Discriminant Analysis (PLS-DA)

Partial Least Squares Discriminant Analysis (PLS-DA) is a supervised discriminant learning method that combines the advantages of Partial Least Squares Regression (PLS) and Discriminant Analysis (DA). It can effectively handle problems such as high-dimensional small samples and multicollinearity, and is widely applied in the field of hyperspectral classification [17]. PLS-DA introduces DA on the basis of PLS, integrating sample information into the process of latent variable extraction. Its core is to transform the classification task into a virtual regression framework: first, a binary dummy response matrix (Y) is constructed according to the samples, and then PLS regression is performed on the original high-dimensional predictor variable matrix (X) and the dummy response matrix Y. By maximizing the covariance between Y and X, a series of orthogonal latent variables are extracted to achieve supervised dimensionality reduction.

2.5. Model Evaluation

After constructing the jujube origin identification model system, the conventional practice often involves examining the accuracy metric based on multiple combinations of training and test sets. It should be noted that when the data distribution is imbalanced, the reliability of the evaluation results will be questioned if only the accuracy is relied upon as the evaluation dimension [18]. Examples show that this study adopted the visual analysis method of confusion matrix to achieve a comprehensive evaluation of the model's classification performance. As a widely used classification evaluation tool in the field of machine learning, the confusion matrix can intuitively present the mapping relationship between predicted categories and actual categories [19]. It is worth noting that in the structure of this matrix, the row vectors correspond to the actual category affiliations, while the column vectors represent the distribution of predicted categories. In addition to the aforementioned metrics, evaluation parameters such as Recall, Precision, and F1 score are also commonly used. These parameters reveal the performance characteristics of the classification model from different perspectives. Their calculation formulas are shown below.

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

$$F1 = \frac{precision \times recall \times 2}{precision + recall} \tag{4}$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \tag{5}$$

Where,

TP (True Positive) — refers to the number of true positive samples, i.e., the number of samples predicted as the target category by the model and actually belonging to that category;

FN (False Negative) — refers to the number of false negative samples, i.e., the number of samples actually belonging to the target category but predicted as other categories by the model;

FP (False Positive) — refers to the number of false positive samples, i.e., the number of samples not actually belonging to the target category but incorrectly predicted as the target category by the model.

3. Results and Analysis

Figure 3 shows the average spectral curves of jujubes from different origins. As can be seen from the figure, the overall variation trend of the spectral curves of jujubes from different origins (Chaoyang, Liaoning; Hetian, Xinjiang; Tengchong, Yunnan; Shiyan, Hubei; Hezhou, Guangxi) is similar. In the wavelength range of 800~1800 nm: the reflectance reaches its peak around 900~1200 nm; fluctuates and decreases between 1200~1400 nm; drops rapidly after 1400 nm; and tends to stabilize after 1600 nm. However, there are subtle differences in the curves among different origins in local wavelength bands: obvious differences in reflectance are observed among jujubes from the six origins in 800~1000 nm; significant reflectance differences exist between Shiyan, Hubei and Tengchong, Yunnan in 1200~1400 nm; and distinct reflectance gaps are found between Shiyan, Hubei and other origins after 1400 nm. These differences to a certain extent reflect the unique spectral reflection characteristics of jujubes from different origins, which are influenced by factors such as the growing environment. Such differences can serve as potential basis for distinguishing jujubes from different origins and facilitate subsequent research on origin tracing based on spectral technology.

Through analyzing the attribution of different peaks in the original spectra, several absorption peaks are clearly observed in Figure 2. The absorption peaks in the range of 980 nm~1050 nm may be related to the stretching vibrations of C-OH bonds in molecules, including proteins and oligosaccharides in jujubes [20]. Meanwhile, the absorption peaks around 1220 nm~1250 nm, 1450 nm~1500 nm, and 1620 nm~1650 nm are presumably attributed to the bending and stretching vibrations of N-H and C-N bonds in the amide groups of organic compounds (proteins and amino acids) [21][22].

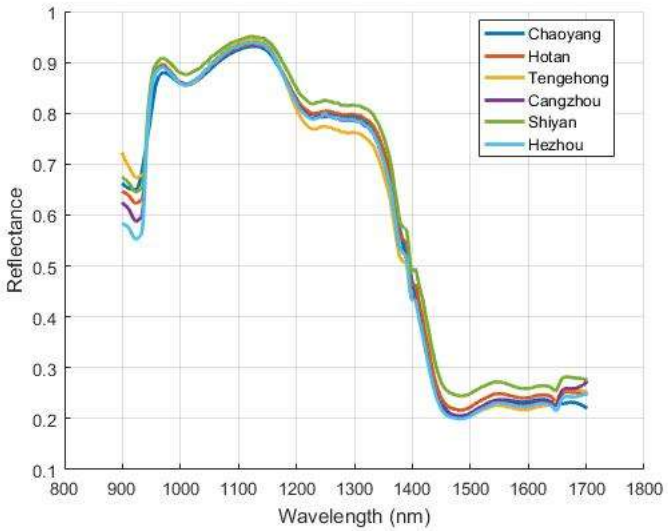


Figure 3 Average Spectral Curve

The raw hyperspectral data in the near-infrared band of jujube samples from different origins were preprocessed using five methods: Standard Normal Variable (SNV), Savitzky-Golay (SG) convolution smoothing, Multiplicative Scatter Correction (MSC), Second Derivative (SD), and First Derivative (FD). The spectral curves after preprocessing are shown in Figure 4. As observed from the figure, the raw spectral curve (a) exhibits complex fluctuations, indicating difficulties in feature recognition. SNV preprocessing (b) alters the overall trend of the spectrum and adjusts the reflectance range of some bands. SG preprocessing (c) smooths the curve to a certain extent, reducing noise while preserving spectral features. MSC preprocessing (d) corrects deviations caused by scattering in the spectrum, making the curve closer to the ideal shape. SD (e) and FD (f) preprocessing methods can highlight differences and changes in specific spectral bands: SD shows discrete fluctuations, while FD presents obvious peaks and valleys in relevant bands. Different preprocessing methods process the raw spectrum from various perspectives, thereby optimizing spectral data, improving the effectiveness of subsequent analyses, and facilitating clearer extraction of spectral features of jujube samples.

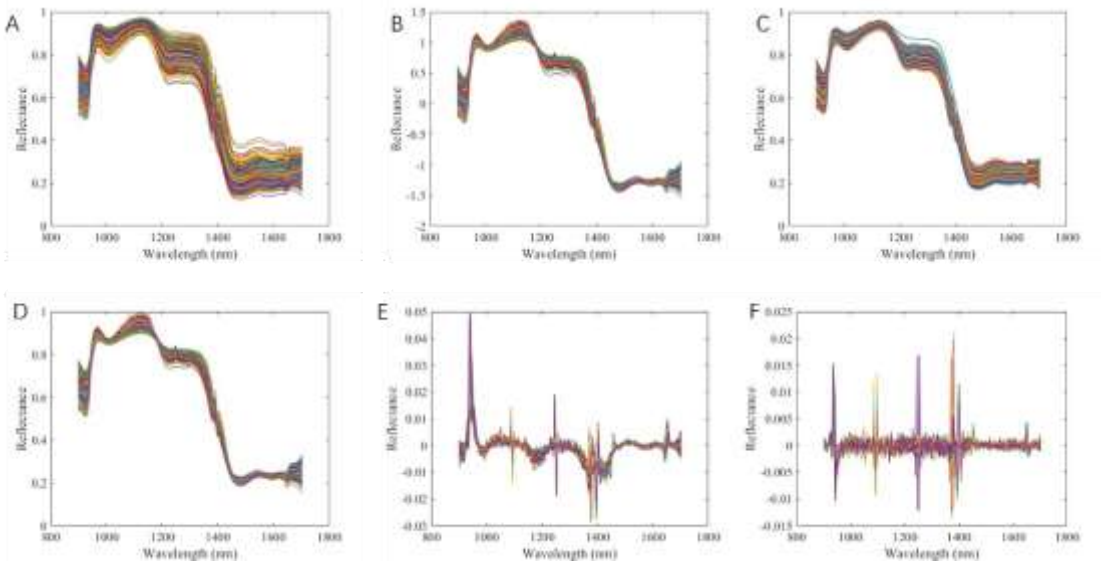


Figure 4 Original spectral curves of jujube samples from different producing areas (a) and spectral curves after pretreatment with SNV (b), SG (c), MSC (d), SD (e), and FD (f) methods

4. Discussion

4.1. Performance Evaluation of Classification Models

Table 2 Evaluation results of paired combinations of pretreatment methods and classification models for jujube samples from different producing areas

model	Preprocessing	Training Set				Test Set			
		Accuracy	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score
RF	none	93.53%	93.58%	93.49%	93.53%	85.61%	85.71%	85.52%	85.62%
	SG	93.29%	93.28%	93.25%	93.26%	84.00%	84.23%	83.90%	84.06%
	MSC	93.80%	93.87%	93.75%	93.81%	88.92%	89.05%	88.84%	88.94%
	SNV	93.57%	93.67%	93.52%	93.60%	89.33%	89.48%	89.25%	89.36%
	FD	98.97%	98.98%	98.96%	98.97%	98.79%	98.82%	98.78%	98.80%
	SD	99.49%	99.50%	99.49%	99.50%	99.11%	99.15%	99.11%	99.13%
PLS.DA	none	90.16%	90.24%	90.07%	90.16%	89.49%	89.68%	89.40%	89.54%
	SG	89.83%	89.94%	89.74%	89.84%	89.17%	89.33%	89.07%	89.20%
	MSC	93.21%	93.10%	93.15%	93.12%	92.88%	92.84%	92.82%	92.83%
	SNV	93.47%	93.40%	93.41%	93.41%	93.53%	93.61%	93.48%	93.54%
	FD	97.92%	97.93%	97.90%	97.91%	97.90%	97.93%	97.88%	97.91%
	SD	99.49%	99.50%	99.49%	99.50%	99.27%	99.30%	99.27%	99.29%
SVM	none	92.12%	92.14%	92.06%	92.10%	91.84%	91.94%	91.78%	91.86%
	SG	91.98%	92.00%	91.91%	91.96%	91.67%	91.78%	91.61%	91.70%
	MSC	92.85%	92.83%	92.79%	92.81%	92.24%	92.29%	92.18%	92.24%
	SNV	99.68%	99.68%	99.67%	99.68%	99.43%	99.45%	99.43%	99.44%
	FD	98.95%	98.95%	98.94%	98.94%	98.38%	98.40%	98.37%	98.38%
	SD	99.11%	99.11%	99.10%	99.11%	98.54%	98.57%	98.53%	98.55%

The data presented in Table 2 indicates significant differences in the performance of three classification models—Random Forest (RF), Partial Least Squares Discriminant Analysis (PLS-DA), and Support Vector Machine (SVM)—in identifying jujube origins across different spectral preprocessing methods. Under the Second Derivative (SD) preprocessing, the RF model achieved the optimal performance, with test set accuracy, precision, recall, and F1 score reaching 99.11%, 99.15%, 99.11%, and 99.13%, respectively. For instance, when using Savitzky-Golay (SG) preprocessing, the F1 score of the RF model dropped to only 84.06%, representing a substantial gap of 15.07%. Similarly, the PLS-DA model under SD preprocessing exhibited excellent performance, with all four test set metrics exceeding 99%, including an F1 score of 99.29%. In contrast, without any preprocessing, the F1 score of the PLS-DA model decreased to 89.54%, reflecting a performance degradation of 9.75%. The most outstanding results were observed in the SVM model combined with Standard Normal Variable (SNV) preprocessing, where the test set accuracy (99.43%), precision (99.45%), recall (99.43%), and F1 score (99.44%) all set the highest records in this study. For example, with SG preprocessing, the F1 score of the SVM model was 91.70%, a difference of 7.74 percentage points from the optimal result. Further analysis revealed that compared to the SD-PLS-DA combination, the SNV-SVM combination improved the F1 score and accuracy by 0.15% and 0.16%, respectively; compared to the SD-RF combination, the improvements were 0.31% and 0.32%, respectively. Derivative transformation preprocessing (i.e., FD and SD) generally enhanced model performance, as evidenced by the nearly 99% excellent metrics of the RF and PLS-DA models under these two preprocessing methods. A comprehensive evaluation of model stability showed that the SVM model had the strongest adaptability. Notably, the SNV-SVM combination not only achieved the best classification performance but also demonstrated the highest generalization ability. Overall, the confusion matrix of the SNV-SVM (C) combination yielded the highest discrimination accuracy, providing data support for practical applications such as jujube origin tracing.

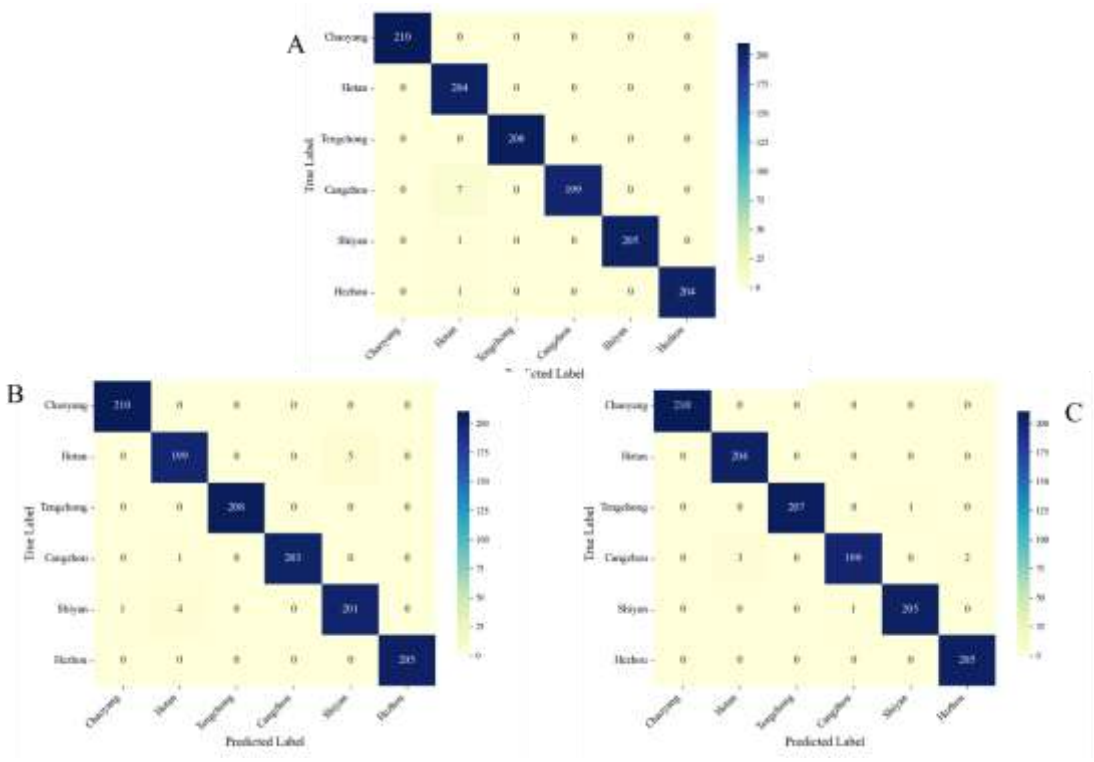


Figure 5 Discrimination results of confusion matrices for SD-PLS-DA (A), SD-RF (B), and SNV-SVM (C)

5. Conclusions

In this experiment, jujube samples purchased from six different provinces (Chaoyang, Liaoning; Hetian, Xinjiang; Tengchong, Yunnan; Cangzhou, Hebei; Shiyan, Hubei; Hezhou, Guangxi) were selected as the objects for spectral data collection. Three classification models—Random Forest (RF), Support Vector Machine (SVM), and Partial Least Squares Discriminant Analysis (PLS-DA)—were combined with five spectral preprocessing methods, namely Multiplicative Scatter Correction (MSC), Savitzky-Golay (SG) convolution smoothing, Standard Normal Variable (SNV), First Derivative (FD), and Second Derivative (SD), to establish origin identification models using hyperspectral data of jujube samples from different provinces.

Through the evaluation of paired combinations of preprocessing methods and classification models, three combinations exhibited optimal performance: SNV-SVM (precision: 99.45%, recall: 99.43%, F1 score: 99.44%);SD-PLS-DA (precision: 99.30%, recall: 99.27%, F1 score: 99.29%);RF-SD (precision: 99.15%, recall: 99.11%, F1 score: 99.13%).Further visualization evaluation using confusion matrices revealed that the SNV-SVM combination achieved the highest classification accuracy of 99.43%, followed by SD-PLS-DA (99.27%) and SD-RF (99.11%). All three models demonstrated excellent performance. Comprehensive assessment indicated that the hyperspectral technology combined with the SNV-SVM model exhibited the best overall performance, enabling rapid and accurate identification of jujube origins. This method holds promising application prospects in origin authentication.

Acknowledgment

Fund Project: Graduate Innovation Fund Project of Sichuan University of Science & Engineering (Y2024023).

References

[1] Zhao Ailing, Li Dengke, Wang Yongkang, et al. Evaluation of Nutritional Characteristics and Germplasm Screening of Jujube Cultivar Resources[J]. Journal of Plant Genetic Resources, 2010, 11(06): 811-816.

[2] Liu Lihua. High-yield cultivation and storage-freshness preservation techniques of jujube in western Liaoning[J]. Forest By-Products and Speciality in China, 2021, (05): 37-39.

[3] Dai Qi, Bai Yuanding, Ye Qiaobo, et al. Research progress on chemical components and pharmacological effects of jujube from different producing areas[J]. Chinese Journal of Drug Evaluation, 2023, 40(06): 506-511.

[4] Shao Wuxia. Study on the strategies for enhancing the competitiveness of the regional public brand of Ningyang jujube[D]. Tai'an: Shandong Agricultural University, 2023.

[5] Li Shang, Fang Cuilan, Xiao Lun, et al. Research progress on separation and detection methods of main active components in jujube[J/OL]. Journal of Hygiene Research, 1-9[2025-07-02].

[6] Li Shixin, Fan Xiaoping. Research and application of image processing and recognition in monitoring fruit and vegetable ripeness[J]. Packaging Engineering, 2024, 45(03): 153-164.

[7] HUANG HP, HU XJ, TIAN JP, et al. Rapid and nondestructive prediction of amylose and amylopectin contents in sorghum based on hyperspectral imaging [J]. Food Chem, 2021, 359: 129954.

[8] Gu Jiasheng, Liu Zijian, Zhou Cong, et al. Study on origin identification of green prickly ash based on hyperspectral imaging technology[J]. Journal of Food Safety & Quality, 2023, 14(21): 210-218.

[9] He Wenwen, Tian Chenghua. A method for identifying maize variety purity based on hyperspectral features using deep learning[J]. Crop Research, 2024, 38(04): 272-277+313.

[10] Peng Jianheng, Hu Xinjun, Zhang Jiahong, et al. Identification and adulteration detection of lotus root starch using hyperspectral imaging technology combined with deep

learning[J]. Spectroscopy and Spectral Analysis, 2025, 45(06): 1759-1767.

[11] Sun Xue, Zhang Dengting, Wang Hui, et al. Origin identification of Poria cocos based on hyperspectral imaging technology[J]. China Journal of Chinese Materia Medica, 2023, 48(16): 4337-4346.

[12] Shao Y, Liu Y, Xuan G, et al. Detection and analysis of sweet potato defects based on hyperspectral imaging technology[J]. Infrared Physics & Technology, 2022, 127: 104403.

[13] Diwupengyao, Bian Xihui, Wang Zifang, et al. Research on the selection of spectral preprocessing methods[J]. Spectroscopy and Spectral Analysis, 2019, 39(09): 2800-2806.

[14] Jia Dan, Ming Mina, Lei Lei, et al. Identification of vegetable oil types based on PCA-ACO-SVM algorithm and FTIR technology[J]. Chinese Journal of Light Scattering, 2024, 36(04): 454-460.

[15] LV YP, LV WB, HAN KX, et al. Determination of wheat kernels damaged by fusarium head blight using monochromatic images of effective wavelengths from hyperspectral imaging coupled with an architecture self-search deep network [J]. Food Control, 2022, 135: 108819.

[16] Yang Chunxia. Application of three-dimensional fluorescence spectroscopy combined with multivariate statistical analysis in food quality and safety control[J]. China Food Additives, 2025, 36(06): 96-102.

[17] Lei Rong, Yang Jianling, Zhou Yanan, et al. Quality evaluation of Shenmai Granules based on multi-index component quantification combined with partial least squares-discriminant analysis and entropy weight-TOPSIS method[J]. Chinese Traditional and Herbal Drugs, 2025, 56(02): 467-475.

[18] FAWCETT T. An introduction to roc analysis [J]. Pattern Recogn Lett,2006,27(8):861.

[19] Qian Min, Wang Zhe, Cheng Lianhua. Study on the causes of coal mine transportation accidents based on configuration analysis and machine learning[J]. Journal of Xi'an University of Science and Technology, 2025, 45(02): 286-295.

[20] Cen H, He Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality[J]. Trends in Food Science & Technology, 2007, 18(2): 72-83.

[21] Hua R, Sun S Q, Zhou Q, et al. Discrimination of fritillary according to geographical origin with fourier transform infrared spectroscopy and two-dimensional correlation ir spectroscopy[J]. Journal of Pharmaceutical and Biomedical Analysis, 2003, 33(2): 199-209.

[22] Türker-Kaya S, Huck C W. A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis[J]. Molecules, 2017, 22(1): 168.