ACTA SCIENTIARUM POLONORUM
Acta Sci. Pol. Technol. Aliment. 24(1) 2025, 27–45
www.food.actapol.net pISSN 1644-0730 eISSN 1898-9594 http://doi.org/10.17306/J.AFS.001277

ORIGINAL PAPER

Received: 13.08.2024
Accepted: 24.10.2024

RAPID DETECTION OF COMMON ADDITIVES IN TEA FOR QUALITY ASSURANCE VIA MID-INFRARED SPECTROSCOPY AND MACHINE LEARNING

Weiyu Liu[1,2], Yuduan Lin[1,3], Cuihua Liu[1], Honghao Cai[1]✉, Hui Ni[4,5]

[1]Department of Physics, School of Science, Jimei University
Xiamen, Fujian Province, **China**
[2]School of Materials Science and Engineering, Guilin University of Technology
Guangxi Province, **China**
[3]School of Electronic Science and Engineering, Xiamen University
Xiamen, Fujian Province, **China**
[4]College of Food and Biology Engineering, Jimei University
Xiamen, Fujian Province, **China**
[5]Fujian Provincial Key Laboratory of Food Microbiology and Enzyme Engineering
Xiamen, Fujian, **China**

ABSTRACT

Background. Tea, a globally popular beverage, is frequently adulterated with various additives for profit. However, research on these additives – particularly their common combinations in fraudulent practices – remains limited.

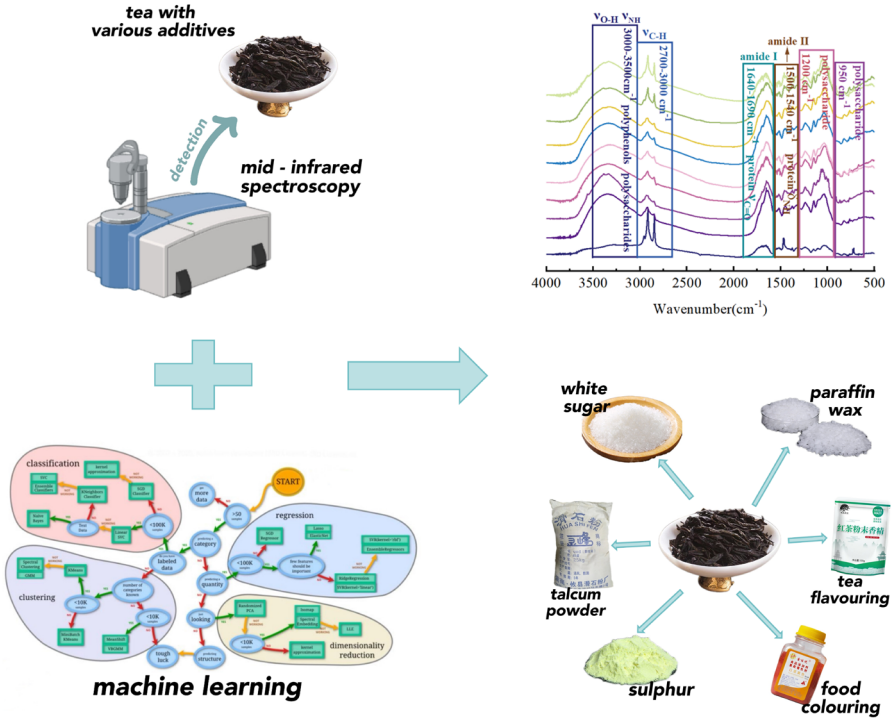
Materials and methods. This study uses mid-infrared (MIR) spectroscopy and machine learning techniques to identify such additives in tea. We trained several models – AdaBoost, random forest, K-nearest neighbours (KNN), support vector classification, Gaussian naive Bayes, and decision tree – to detect adulterants such as white sugar, talcum powder, sulphur, paraffin wax, food colouring, and flavourings, as well as their combinations. To improve accuracy and efficiency, we employed the Successive Projections Algorithm (SPA) and Competitive Adaptive Reweighted Sampling (CARS) for feature selection, prioritising features that exhibited the strongest correlations with the additives.

Results. Compared to models relying solely on raw spectra, those incorporating SPA and CARS consistently achieved high accuracy and reduced detection time by at least 90%. The selected wavenumbers also serve as biomarkers for specific additives. The SPA-KNN model performed exceptionally well, with an accuracy of 0.956, macro-precision of 0.964, macro-recall of 0.956, and a macro-F1 score of 0.956, all with a detection time of just 0.9 seconds.

Conclusion. These results highlight the effectiveness of combining MIR spectroscopy with advanced classifiers and feature selection algorithms, offering a rapid and precise method for detecting tea additives. This approach benefits producers, distributors, and consumers alike.

Keywords: adulteration, food quality, K-nearest neighbours, classification algorithm, feature selection

✉hhcai@jmu.edu.cn, https://orcid.org/0000-0002-1870-8061



Graphical Abstract

INTRODUCTION

Tea is one of the world's most popular beverages, cherished for its flavour and taste. However, tea merchants may illegally add certain additives during production to conceal quality defects, enhance the colour of tea leaves, reduce production costs, and increase the price. Common additives include white sugar, talcum powder, sulphur, paraffin wax, food colouring, and flavourings. Talcum powder, a widely used industrial mineral in body and face powders, has been linked to an increased risk of cancer (Ng et al., 2012; Nnorom, 2011). Research has shown that excessive use of talcum powder may elevate cancer risks (Li et al., 2016). Although sulphur is an essential trace element in the human body, excessive consumption can have harmful effects on health. Substances like flavouring, paraffin, food colouring and white sugar, are permitted food additives within regulated limits, but exceeding these limits can negatively impact health. Tea with added white sugar is prone to moisture absorption, making it challenging to assess its quality over time. If individuals with diabetes consume deteriorated tea, it may trigger their condition. Additives in production may also introduce impurities from raw materials, leading to contamination with toxic compounds that can cause acute and chronic poisoning. For instance, the antioxidant butylated hydroxyanisole, a common preservative, can accumulate in the body and damage the thyroid system, disrupt metabolism and growth, and pose neurotoxic and carcinogenic risks (Zhang et al., 2023). Furthermore, the addition of illegal additives in tea production is prohibited under China's food safety standard GB/T14456.17.

Infrared (IR) spectroscopy is widely recognised as a powerful tool for analysing chemical components based on specific absorption frequencies corresponding to functional groups. Compared with other detection methods, such as nuclear magnetic resonance (NMR), mass spectrometry and Raman spectroscopy, IR spectroscopy offers advantages in speed, cost-effectiveness and ease of use, making it widely applied in food research. Machine learning, a key branch of artificial intelligence, uses large datasets and statistical

methods to enable computers to learn from data and generate models autonomously (Gao et al., 2024a; 2024b; Qiu et al., 2024). This empowers machines to make predictions, decisions, and identifications based on past experiences. Combining IR spectroscopy with machine learning enhances prediction accuracy and efficiency, facilitating automated analysis of large-scale data to uncover hidden patterns and relationships. Researchers have investigated various applications, such as detecting microbial spoilage (Ellis et al., 2002; 2004), identifying food varieties and geographical origins (Chen et al., 2023; Özdemir et al., 2024; Xiao et al., 2024), and recognising food adulteration (Da Costa Filho et al., 2022; Valand et al., 2020).

In tea research, IR spectroscopy has been used to differentiate the origin of tea (He et al., 2012; Zhuang et al., 2017), assess its quality (Ding et al., 2022; Xia et al., 2024), and identify different varieties (Zhang et al., 2024a; 2024b). While researchers (Amsaraj et al., 2023; Li et al., 2016) have conducted quantitative analyses of additives such as sunset yellow and talcum powder in tea using regression models, many other commonly used additives remain unexplored. Moreover, no comprehensive model has yet been developed that can simultaneously identify multiple additives in tea. Current research mainly focuses on individual additives, but adding combinations of multiple additives is a common practice in illicit tea production.

IR instruments often generate spectral data exhibiting high collinearity, and mathematical models based on the full spectrum, which includes many irrelevant spectral variables, can reduce prediction accuracy. Traditional manual selection of spectral ranges may either lose important information or retain excessive redundant information. Furthermore, from a model interpretation standpoint, it remains challenging for analytical chemists and chemometricians to identify the specific wavelengths or combinations responsible for the properties of interest. Both experimentally and theoretically, the calibration model's performance can be improved by using selected informative wavelengths rather than the full spectrum (Li et al., 2009). To address this, effective methods, such as the successive projections algorithm (SPA) (Tang et al., 2018) and competitive adaptive reweighted sampling (CARS) (Li et al., 2023) have been employed to eliminate redundant information from wavelengths in this study.

The objective of this research is to explore the application of mid-infrared (MIR) spectroscopy in differentiating various additives in tea. The specific goals are:

1. Establishing models using MIR spectroscopy to differentiate between various additives, including talcum powder, food colouring (sunset yellow), white sugar, tea flavouring, sulphur, and paraffin wax, employing six classifiers: AdaBoost, random forest (RF), K-nearest neighbour (KNN), support vector classification (SVC), Gaussian naive Bayes (GNB), and decision tree (DT).
2. Investigating the differentiation of tea samples with multiple additive combinations.
3. Determining the optimal spectral preprocessing combination.
4. Utilising SPA and CARS to select the most relevant wavenumbers and improve model performance.

This study takes into account the complex and diverse adulteration practices encountered in real-world scenarios, making it a valuable addition to the existing research on tea additive identification.

MATERIALS AND METHODS

Sample preparation

The products used in the experiment were as follows: Wuyi rock tea (First Class Teas Inc., Wuyishan City, Fujian, China), white sugar (Sanlvyuan Food Factory, Tong'an District, China), talcum powder (Guilin Guiguang Talc Development Co., Ltd., China), food colouring (Sunset Yellow) (Dongguan Jinjiahe Food Co., Ltd., China), paraffin wax (Penglei Chemical Division, China), tea flavouring (Dongguan Jinjiahe Food Co., Ltd., China), and sulphur (Dongguan Jinjiahe Food Co., Ltd., China). All the products were sealed in plastic bags with desiccants and stored at 25°C in the laboratory.

The sample preparation scheme is presented in Figure 1. Since the use of additives in tea is prohibited in China, specific dosage information from published sources is unavailable. Therefore, based on information from illicit vendors, it was determined that during the tea frying process, the ratio of additives to fresh tea leaves is typically 1:50. At this ratio, additives are not easily detectable by consumers, and the small quantity of additives contributes to the practicality of this

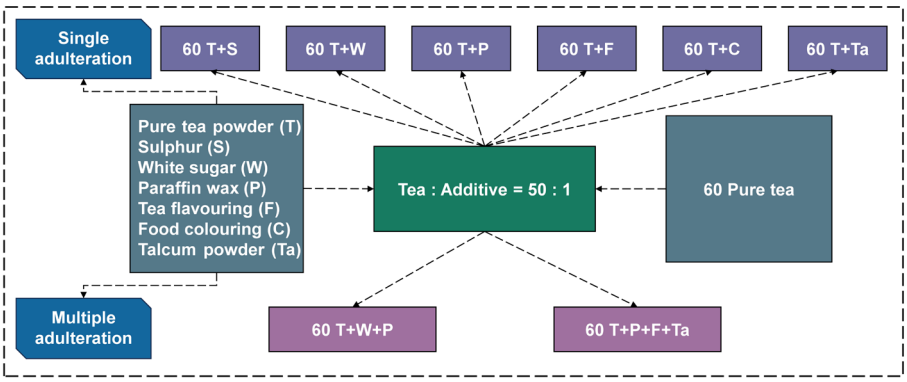


Fig. 1. Overview of sample production

study. Tea production with additives was outsourced to First Class Teas Inc. In total, eight types of additives were used: (1) white sugar, (2) talcum powder, (3) food colouring, (4) paraffin wax, (5) tea flavouring, (6) sulphur, (7) white sugar + paraffin wax, and (8) talcum powder + paraffin wax + tea flavouring.

After drying the samples in a drying chamber (Shangcheng Co., Ltd., China) at 43°C for 2 hours, they were weighed using a balance with 0.0001 g accuracy (Lichen Co., Ltd., China). The tea was ground into fine particles and passed through a 100-mesh sieve to ensure uniform particle size. The tea was then blended with the additives in the prescribed proportions. Subsequently, 3 mg of either pure or adulterated tea leaves were weighed and combined with potassium bromide in a 1:10 ratio to prepare samples using the potassium bromide pellet technique. The pure and adulterated tea tablets (3 mg) were then produced following this method (Donald et al., 1954). A total of 60 tablets were prepared for pure tea and each type of adulterated tea, resulting in 540 tablets in total.

Acquisition of MIR Spectra

IR spectral acquisition was performed at a constant temperature of 25°C using a WQF-500 Fourier transform infrared (FTIR) spectrometer (Beifen-Ruili Analytical Instrument Co., Ltd., China). The instrument was operated with Main FTOS Suite software (Beifen-Ruili Analytical Instrument Co., Ltd., China). Spectra were recorded in absorbance mode, covering the range of 4500 to 400 cm^{-1}, with 256 scans and a total of 4253 variables. The measurement background, represented by potassium bromide, was subtracted from the sample spectra. Three measurements were taken for each sample and then averaged.

MIR Spectra Preprocessing

First, all spectra were pretreated using min-max normalization (MM-norm) (Borkin et al., 2019) and automatic baseline correction (BC) (Shao et al., 2007) methods via the main FTOS Suite software. To further optimise the spectra for efficient information extraction, the standard normal variate (SNV) (Barnes et al., 1989), Savitzky-Golay (SG) (Zimmermann et al., 2013), and first and second derivatives (FD/SD) (Rinnan et al., 2009; Whitbeck, 1981) were applied. These preprocessing methods were implemented in Python using Spyder compiler software (Anaconda, Inc., USA). MM-norm scales the data range into a specific interval, reducing the parameter update step size during the gradient descent process and thereby improving the algorithm's convergence speed. BC aids in identifying true peaks by removing or adjusting the baseline offset in the signal or spectrum, improving the accuracy and precision of the analysis. It also facilitates the identification and interpretation of specific features. SNV is primarily used to eliminate the effects of solid particle size and surface scattering, reduce interference in spectral data, and enhance data comparability and interpretability. SG helps reduce signal vibration, minimise noise, enhance signal characteristics, and suppress unwanted spectral features caused by the instrument and sample. FD removes the influence of baseline drift and smooths

background interference, improving resolution. SD enhances spectral resolution, eliminates background noise, and increases the accuracy and reliability of IR spectral analysis. The performance of these combinations was evaluated using spectra pretreated with the MM-norm and BC.

Machine Learning Workflow

The workflow of the machine learning model is depicted in Figure 2. The dataset was randomly split into a training set (n = 540) and a test set (n = 90) at a 6:1 ratio, as shown in the upper dotted box in Figure 2. The data underwent preprocessing with various combinations of methods, and the optimal combination was determined through 10-fold cross-validation and a grid search, implemented in Python using the scikit-learn toolbox, as depicted in the left section of the middle-dotted box in Figure 2. Feature wavenumbers were then selected using SPA and CARS, as shown in the right section of the middle-dotted box in Figure 2. In this study, nine classification models were built using six machine learning classifiers: AdaBoost (Ying et al., 2013), RF (Huang et al., 2023), KNN (Trigueiro et al., 2019), SVC (Brereton et al., 2010), GNB (Wang et al., 2016), and DT (Barbosa et al., 2014), as shown in the lower dotted box of Figure 2. The code and parameter settings for the algorithms are available in the supplementary materials.

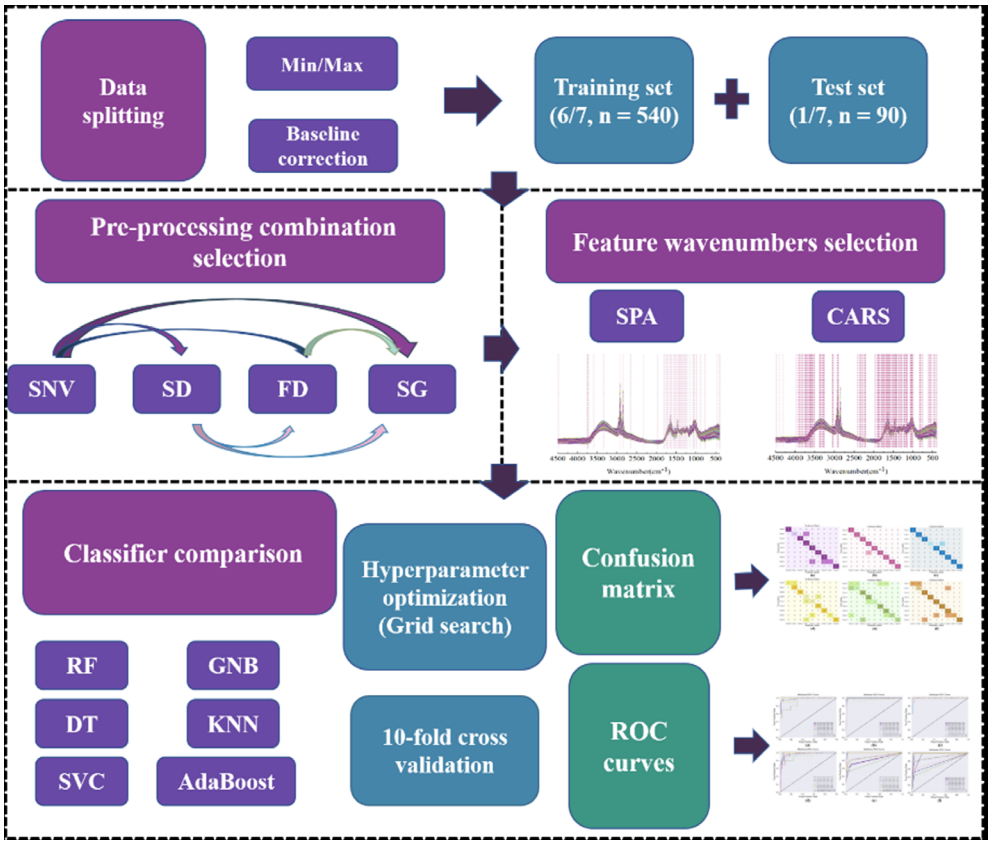


Fig. 2. Diagram of the machine learning workflow. (Min/Max = min-max normalization, SNV = standard normal variate, SD = second derivative, FD = first derivative, SG = Savitzky-Golay, SPA = successive projections algorithm, CARS = competitive adaptive reweighted sampling, RF = random forest, DT = decision tree, KNN = K-nearest neighbour, SVC = support vector classification, GNB = Gaussian naive Bayes)

Evaluation Metrics

To assess the fit and performance of the models and identify the one with optimal performance, this study calculates several evaluation metrics, including accuracy, precision, recall, F1 score, macro-precision, macro-recall, and macro-F1 score. The specific definitions and calculation methods for these indicators are provided in Table S1 in the supplementary materials. Additionally, the confusion matrix was visualised, and receiver operating characteristic (ROC) curves were plotted to provide a clearer evaluation of the model, as shown in the lower dotted box in Figure. 2.

RESULTS AND DISCUSSION

MIR Spectra Analysis

The MIR spectra of the pure tea samples and tea samples with the additives are shown in Figure 3. To improve the clarity of the spectra and facilitate interpretation, automatic baseline correction and min/max normalization were applied. The pure tea samples and most tea samples with additives exhibited similar absorption bands, although with distinct differences in shape and intensity. The broad absorption band at approximately 3500 to 3000 cm^{-1} is primarily attributed to the stretching vibrations of the hydroxyl group (-OH) and nitrogen–hydrogen bond (N-H) in polyphenols, polysaccharides, and hydrogen-containing compounds (Wu et al., 2019). The hydroxyl group (-OH) band results from the moisture content in the tea leaves. The band at approximately 3000 to 2700 cm^{-1} is mainly associated with the stretching vibrations of the hydrocarbon bond (C-H) generated by the protein and amino acid components in the tea leaves. Additionally, the band near 1690 to 1640 cm^{-1} corresponds to the amide I band of protein molecules, primarily attributed to the stretching vibrations of carbonyl bonds (C=O) generated by polyphenolic compounds like catechins and theaflavins in tea leaves. The band near 1540 to 1500 cm^{-1} represents the amide II band of proteins, mainly due to the in-plane bending vibrations of the peptide bond (-CONH-), which involves the nitrogen–hydrogen bond (N-H) generated by polyphenolic compounds (Xiang, 2011). Furthermore, the spectral region between 1200 and 950 cm^{-1} is characteristic of polysaccharides (Ma et al., 2018).

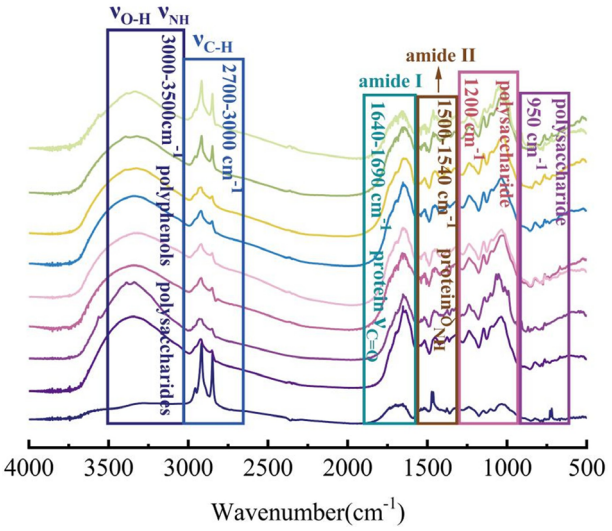


Fig. 3. Mid-infrared spectra of tea with wavenumber identifiers: ▬ = tea + paraffin wax, ▬ = pure tea, ▬ = tea + white sugar, ▬ = tea + talcum powder, ▬ = tea + sulphur, ▬ = tea + food colouring, ▬ = tea + tea flavouring, ▬ = tea + talcum powder + paraffin wax + tea flavouring, ▬ = tea + white sugar + paraffin wax

The addition of various additives alters the composition of tea, resulting in noticeable differences in the spectra related to the constituent bonds. Therefore, additives can be detected in tea through molecular bond information from the fingerprint region. Specifically, the addition of paraffin wax, white sugar + paraffin wax, and talcum powder + paraffin wax + tea flavouring resulted in significant alterations in the infrared spectra of the tea samples. These changes can be attributed to several factors:

1. **Unique absorption features**: As shown in Figure 4, paraffin wax exhibits strong bands between 3000 and 2750 cm^{-1}. Paraffin may exhibit its own characteristic absorption bands in the infrared spectra. For example, alkyl groups in paraffin wax may display absorption bands corresponding to the stretching vibrations of the hydrocarbon bond (C-H) within the wavenumber range from 3000 to 2750 cm^{-1}, as shown in Figure 3.
2. **Band intensity and shape alterations**: The incorporation of paraffin wax into tea affects its original nutrient content, volatile compound content, and overall composition. This leads to changes in the

absorption characteristics and vibrational modes of the infrared spectrum. Depending on the light absorption properties of the wax components and their interactions with the constituents of tea, these alterations can either amplify or attenuate the intensity and shape of the absorption bands.

3. **Band shifts**: The presence of wax can cause shifts in the positions of certain absorption bands between 1750 and 1500 cm^{-1} and 1200 and 1000 cm^{-1}, as shown in Figure 3. This occurs because the chemical components in paraffin wax may interact with the tea constituents, affecting the vibrational frequencies of the molecules.

Experienced tea experts can initially detect additives like paraffin wax by comparing the taste, aroma, and colour of the tea with known tea samples. In contrast, IR spectroscopy allows for a molecular-level analysis of the additives. Combining both approaches enables a more comprehensive evaluation of tea quality.

Notably, as shown in Figure 4, the IR spectrum of sulphur exhibits unusual band shapes. This is due to the use of natural sulphur, which exists as sulphur ore and contains high levels of impurities such as sand and rock. This results in discrepancies between its spectrum and the standard sulphur spectrum (Mohammadi et al., 2022). The distinctive paraffin bands observed between 3000 and 2750 cm^{-1} led to changes in peak intensity following its addition to tea. In contrast, the nearly complete absence of characteristic peaks for talcum powder and food flavouring compounds between 3500 and 2000 cm^{-1} led to a reduction in band intensity upon their incorporation into the tea. These spectral variations provide clear evidence of the presence of these additives, demonstrating the potential of infrared spectroscopy to detect and identify adulterants in tea samples.

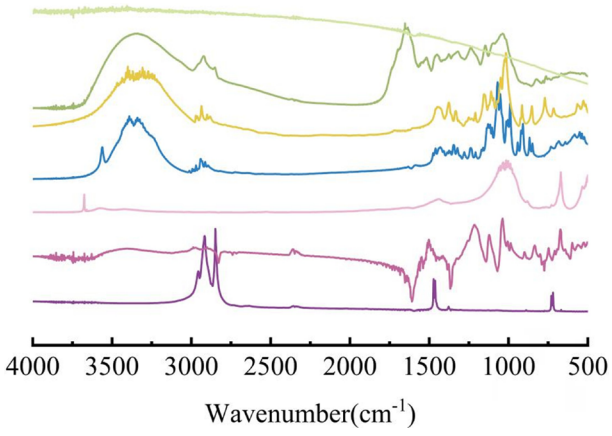


Fig. 4. Mid-infrared spectra of additives with wavenumber identifiers: ▬ = paraffin wax, ▬ = food colouring, ▬ = talcum powder, ▬ = white sugar, ▬ = tea flavouring, ▬ = pure tea, and ▬ = sulphur

In 2016, Li et al. conducted an experiment in which tea was adulterated with talcum powder. The addition of talcum powder did not cause any noticeable changes in the IR spectra, consistent with previous results. In 2023, Rani Amsaraj et al. studied tea adulteration with sunset yellow using machine learning. By comparing the IR spectra from these two experiments with those obtained from this study, similar bands and intensities were observed, indicating the validity of the findings. Both studies demonstrate the feasibility of quantitatively identifying additives using IR and machine learning. This study, therefore, aimed to use the AdaBoost, GNB, SVC, KNN, RF, and DT algorithms to develop classification models for different additives.

Optimisation of the spectral preprocessing combination

To optimise the combination of spectral preprocessing, the performances of various preprocessing methods were calculated and compared using the AdaBoost classifier, as shown in Table S2 in the supplementary materials. The results indicated that the AdaBoost classifier effectively distinguishes tea additives, with a 10-fold cross-validation score ranging from 0.836 to 0.960. The optimal combination of spectral preprocessing was SNV&SD&SG, which achieved a discrimination rate of 0.960. The second-best combination of spectral preprocessing methods was FD&SG, yielding a discrimination rate of 0.947. Notably, applying FD/SD processing to only the raw* data resulted in lower scores, and combining SNV with derivatives also resulted in a decrease in scores. The results obtained by applying FD to the raw* data were the lowest, with a score of only 0.844. These results highlight the crucial role of SG as a preprocessing method. The reduced accuracy of the infrared spectra without smoothing treatment can be attributed to the presence of noise and stray

signals. Interference from noise and stray signals can cause peak broadening or false peaks, leading to a decrease in the accuracy of the IR spectra. In general, the cross-validation scores for various combinations of preprocessing methods were higher than those for the raw* data.

Feature Selection via SPA and CARS

IR spectroscopy typically covers a wide range of wavenumbers, some of which are irrelevant to the model. Therefore, it is necessary to eliminate these interfering wavenumbers and select the most relevant ones to improve the model's accuracy and reduce modelling time. In 2019, Sun et al. proposed using SPA and CARS coupled with stepwise regression to select characteristic wavelengths, thereby enhancing the correlation coefficient of the prediction set in the established multiple linear regression model. In this study, SPA and CARS were employed to select the most relevant frequencies for the classification model. As a result, 41 effective feature wavenumbers were obtained via SPA, and 297 effective feature wavenumbers were obtained via CARS from the spectra of 540 samples. These features are highlighted as dashed lines in Figure 5.

Both SPA and CARS are feature selection methods used in spectral analysis. SPA progressively selects feature variables with maximum correlation to an additive target (Esteki et al., 2016), whereas CARS adaptively adjusts feature weights by considering competition between features and selects those with the strongest classification ability. Both algorithms may employ attentional focusing strategies that prioritise relevant features, improving classification accuracy and enhancing differentiation between additives. The choice of feature selection strategy ultimately depends on the specific experimental design, data characteristics, and algorithmic considerations. SPA extracts features by calculating the area under the peak region in the spectra, while CARS extracts features through fitting or peak analysis of the spectra. SPA focuses only on the peak region, whereas CARS typically selects more features, including peak position, peak intensity, and peak width. Thus, SPA and CARS select different wavenumbers because of their underlying principles and evaluation criteria. Although the wavenumbers selected by SPA and CARS

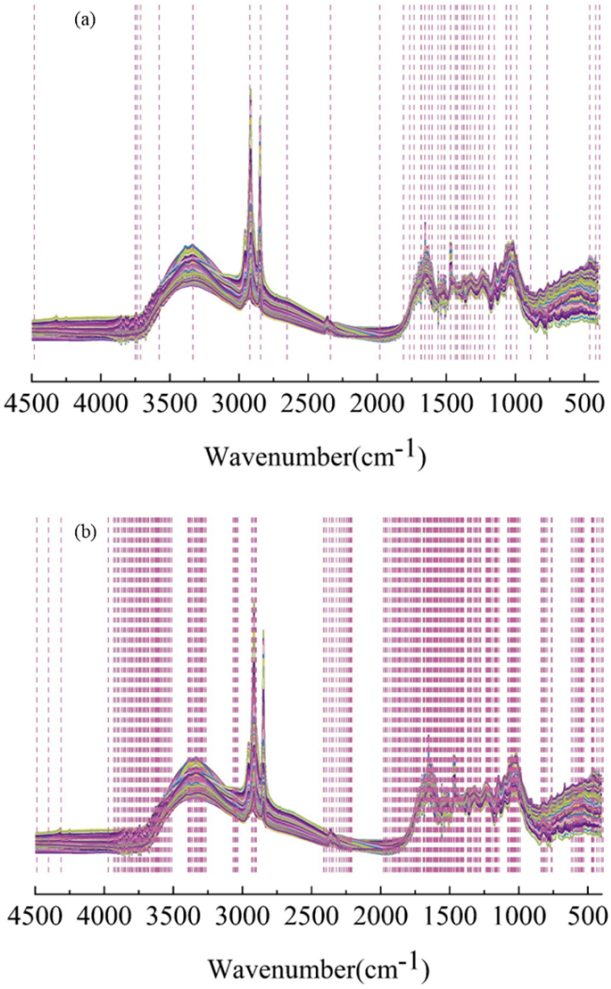


Fig. 5. The wavenumbers selected by (a) SPA and (b) CARS

were not identical, significant overlaps between them indicated that certain features in the IR spectra were considered important across various analysis methods. This overlap potentially offers more reliable and consistent feature information, facilitating accurate analysis and identification of additives.

Moreover, the selected wavenumbers can be used to identify additives. As shown in Figure 3, significant differences were observed in the spectral comparison between tea adulterated with paraffin wax and pure tea in the ranges of 3000 to 2750 cm^{-1}, 1750 to 1500 cm^{-1}, and 1200 to 1000 cm^{-1}. Notably, similar band variations were also observed in tea adulterated with talcum powder + paraffin wax + tea flavouring. These variations were primarily due to paraffin wax adulteration,

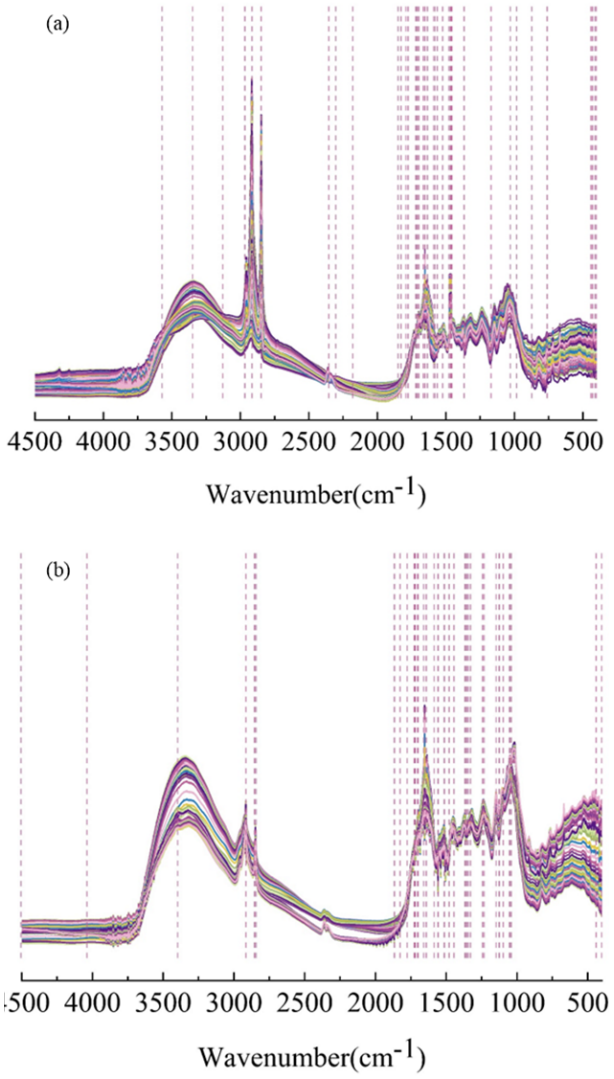


Fig. 6. The wavenumbers selected by SPA: (a) tea + paraffin wax, (b) tea + talcum powder + paraffin wax + tea flavouring

which altered the chemical composition of the tea by introducing artificial additives, leading to an increase in fat content and changes in sugar content. Figure 6a presents the characteristic wavenumbers of tea and paraffin wax, selected through SPA analysis, and compares them to the infrared spectrum of paraffin wax. These wavenumbers align with the distinctive 'fingerprint' of paraffin wax. Additionally, Figure 6 illustrates that the characteristic wavenumbers identified by SPA closely correspond to the major bands of the respective additives. The wavenumbers for tea with other additives, as identified by SPA, are provided in Fig. S1 to S6 in the supplementary materials. These results indicate that SPA is a reliable method for identifying the type of additive.

Model comparison

The six classification algorithms were tested using the optimal spectral preprocessing combination and selected features to generate the optimal model with the highest prediction accuracy, efficiency and generalisation ability. Table S3 in the supplementary materials presents the classification results and performance of the algorithms on both full raw* spectra data and the data processed with SPA and CARS. The results demonstrated that SPA while improving accuracy by 0.017 to 0.106, significantly reduced processing time by at least 97%. Similarly, CARS improved accuracy by 0.017 to 0.102, while decreasing processing time by at least 91%. The SPA-RF achieved the highest accuracy of 0.962, while the SPA-DT model had the shortest processing time of 0.33 seconds. In conclusion, SPA outperformed CARS in reducing processing time while maintaining high accuracy, making it more suitable for large-scale sample detection.

To validate the model, 90 external tea samples were tested. The test dataset consisted of 10 spectra of pure tea and 80 spectra of adulterated tea (tea adulterated with white sugar, talcum powder, food colouring, paraffin wax, tea flavouring, sulphur, white sugar + paraffin wax and talcum powder + paraffin wax + tea flavouring). Table S4 in the supplementary materials summarises the model performance on the test set. The confusion matrices, presented in Figure S7 of the supplementary materials, provide specific classification results and detailed validation information. The ROC curves for the six classifiers are also presented in Figure S8 in the supplementary materials.

The results from the test set differ from those of the training set, as shown in Table S4. When ranked by accuracy on the test set, the order is SPA-KNN > SPA-SVC > SPA-AdaBoost > SPA-GNB > SPA-DT > SPA-RF. Most algorithms saw a decrease in accuracy, which could be attributed to potential differences in data quality and distribution between the training and test sets. Among these, SPA-KNN achieved the best overall result, with the highest accuracy of 0.956.

In the confusion matrix, SPA-KNN misclassified just one sample from the tea adulterated with sugar, two samples from the tea adulterated with food colouring, and one sample from the tea adulterated with talcum powder + paraffin wax + tea flavouring. Notably, these misclassifications occurred specifically in the positions where the respective single adulterants are presented in Fig. S7 in the supplementary materials. Additionally, KNN demonstrated the largest area under the ROC curve compared to the other six algorithms, as shown in Fig. S8 in the supplementary materials.

In 2021, Yang et al. (Jun et al., 2021) integrated diffuse reflectance MIR spectroscopy with chemometrics to rapidly discriminate adulterants in Radix Astragali, achieving a correct prediction rate of 100.0% using the KNN classification method. Both previous studies and the present one demonstrate that KNN, a nonparametric method based on the principle of locality, is particularly well-suited for classifying IR spectroscopy data with complex, unknown distributions and similar IR spectral features.

Common analytical methods include high-performance liquid chromatography (HPLC) (Sun et al., 2019), NMR spectroscopy (Maraschin et al., 2016), surface-enhanced Raman spectroscopy (Lin et al., 2021), and IR spectroscopy. Compared to these methods, our work developed a more complex classification model capable of distinguishing multiple types of additives and their combinations. We employed feature selection algorithms to enhance detection speed while maintaining satisfactory accuracy. The application of SPA and CARS algorithms enabled us to extract valid information from spectra more rapidly than the combination of HPLC-diode array detection with second-order calibration based on alternating trilinear decomposition algorithm (2.07~8.10 min). In contrast to NMR spectroscopy combined with partial least squares-discriminant analysis and RF, which achieved high accuracy (>75%, reaching 90% in the best case), we used SPA with a KNN model, achieving an accuracy of 0.956. Our method significantly outperforms those reported in the literature, both in terms of speed and accuracy.

This research provides a cost-effective and rapid method for food regulatory agencies to monitor tea adulteration. However, practical applications of IR spectroscopy may be affected by environmental factors such as temperature and humidity, which can cause fluctuations in detection results. Additionally, the performance of the machine learning model is heavily dependent on the quality and quantity of the training data. Insufficient or unrepresentative data may result in recognition errors. It is also important to note that fraudulent practices in real-world scenarios can be even more complex, with techniques such as fumigation used to mask the true nature of the tea The effectiveness of our proposed method has not yet been validated in such situations.

Our method offers a sensitivity that can reach an adulterating concentration of 2%, is suitable for both single and multiple adulterating scenarios, and requires only 0.9 seconds for model operation. Therefore, it has significant potential for practical application, including detecting the adulteration of meat with hydrocolloids (Qiu et al., 2024), identifying the practice of adulterating food with lower-quality ingredients (Gao et al., 2024c), determining adulterated food species (Liu et al., 2023) and tracing adulteration based on geographical origins (Xiao et al., 2024).

CONCLUSIONS

This study explored the combination of MIR spectroscopy, spectral preprocessing methods, and feature selection techniques for detecting additives in tea. The optimal spectral preprocessing combination, SNV&SD&SG, achieved a 0.960 discrimination rate on the training set. SPA, CARS, and the optimal spectral preprocessing combination effectively improved algorithm accuracy and processing time. Among the classification models tested on the test set, the SPA-KNN model delivered the most comprehensive performance, with a prediction accuracy of 0.956, macro-precision of 0.964, macro-recall of 0.956, a macro-F1 score of 0.956, and a detection time of 0.9 seconds. Additionally, the selected features can assist in identifying the types of additives. These results suggest that MIR spectroscopy, when paired with machine learning, has the potential for differentiating tea adulterated with various additives.

Although the Chinese government has expressly prohibited the use of additives in tea, unscrupulous traders continue to engage in illegal practices for

financial gain. These additives pose serious risks to consumer health and significantly damage the reputation of tea products. The timely detection and prevention of adulteration through reliable testing methods are critical to promoting standardisation and transparency across the tea industry, ensuring the authenticity and purity of tea. Future research on food adulteration should prioritise the integration of multispectral or hyperspectral imaging technologies, which offer richer spectral data and can enhance detection accuracy and sensitivity. Moreover, leveraging deep learning algorithms, such as 1D convolutional neural networks, may outperform traditional machine learning techniques in classifying and identifying complex samples. To enhance the generalisation ability of detection models, future efforts should focus on developing a standardised dataset encompassing various tea samples and their potential adulterants for robust model training and validation.

FUNDING

This work was supported by the Natural Science Foundation of Fujian Province of China [grant number 2022J01821 and 2022J05163]; the Natural Science Foundation of China [grant number 32172339 and 11705068].

SUPPLEMENTARY MATERIALS

1. AdaBoost

1.1 The code used to build the model

```
from sklearn.ensemble import AdaBoostClassifier
kfold =
KFold(n_splits=10,shuffle=True,random_state=0)
clf = DecisionTreeClassifier(max_depth=4, min_samples_leaf=5, random_state=42)
ada_clf=AdaBoostClassifier(estimator=clf,learni
ng_rate=0.7000000000000001,n_estimators=10,
random_state=42)
ada_clf.fit(X, y)
score = cross_val_score(ada_clf, X, y, cv=kfold)
y_pred = ada_clf.predict(X_test)
score = ada_clf.score(X_test, y_test)
```

1.2 The main hyperparameters

The main hyperparameters were set: "learning_rate" as 0.7000000000000001, the "base_estimator" as DecisionTreeClassifier() and "n_estimators" as 10.

2. Support Vector Classification

2.1 The code used to build the model

```
from sklearn.svm import SVC
kfold =
KFold(n_splits=10,shuffle=True,random_state=0)
param_grid = {'C':10}
clf = SVC(**param_grid,probability=True)
clf.fit(X, y)
score = cross_val_score(clf, X, y, cv=kfold)
y_pred = clf.predict(X_test)
score = clf.score(X_test, y_test)
```

2.2 The main hyperparameters

The main hyperparameters were set: kernel as 'rbf', "gamma" as "scale", "C" as 10, "class_weight" as 'None', and "decision_function_shape" as 'ovr'.

3. Gaussian Naive Bayes

3.1 The code used to build the model

```
from sklearn.naive_bayes import GaussianNB
kfold =
KFold(n_splits=10,shuffle=True,random_state=0)
clf=GaussianNB(var_smooth-
ing=0.0012328467394420659)
clf.fit(X, y)
score = cross_val_score(clf, X, y, cv=kfold)
y_pred = clf.predict(X_test)
score = clf.score(X_test, y_test)
```

3.2 The main hyperparameters

The main hyperparameters were set: "var_smoothing" as 0.0012328467394420659.

4. K Neighbors Classifier

4.1 The code used to build the model

```
from sklearn.neighbors import KNeighborsClassifier
kfold =
KFold(n_splits=10,shuffle=True,random_state=0)
clf = KNeighborsClassifier(n_neighbors=3)
clf.fit(X, y)
score = cross_val_score(clf, X, y, cv=kfold)
y_pred = clf.predict(X_test)
score = clf.score(X_test, y_test)
```

4.2 The main hyperparameters
The main hyperparameters were set: "n_neighbors" as 3.

5. Decision Tree
5.1 The code used to build the model
```
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
iris = datasets.load_iris()
X = iris.data
y = iris.target
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
score = clf_score(y_test, y_pred)
```

5.2 The main hyperparameters
The main hyperparameters were set: "random_state" as 42.

6. Random Forest
6.1 The code used to build the model
```
from sklearn.ensemble import
RandomForestClassifier
clf_classifier = RandomForestClassifier(n_estima-
tors=100, random_state=42)
X_train = [...]
y_train = [...]
clf_classifier.fit(X_train, y_train)
X_test = [...]
score = clf_classifier.predict(X_test)
```

6.2 The main hyperparameters
The main hyperparameters were set: "random_state" as 42.

Table S1. Evaluation metrics based on the multi-classification model

Evaluation metrics	Equation	Definition
Accuracy	$\frac{\sum_i^n TP_i}{\sum_i^n (TP_i + TN_i + FP_i + TN_i)}$	The ratio of correctly classified samples out of all samples
$Precision_i$	$\frac{TP_i}{TP_i + FP_i}$	The ratio of true positive samples to the total number of samples predicted as positive
$Recall_i$	$\frac{TP_i}{TP_i + FN_i}$	The ratio of true positive samples to the total number of actual positive samples
$F1_i$	$\frac{2 \times precision_i \times recall_i}{precision_i + recall_i}$	Take into account both precision and recall
Macro-precision	$\frac{\sum_i^n precision_i}{n}$	The average of precision values for all categories
Macro-recall	$\frac{\sum_i^n recall_i}{n}$	The average of recall values for all categories
Macro-F1	$\frac{\sum_i^n F1_i}{n}$	The average of F1 values for all categories

Table S2. The discrimination results of the AdaBoost on the training set

Spectrum	Spectral preprocessing methods			Cross-validation score	Processing time, s
Baseline correction & min-max normalization (raw*)	SNV	None	None	0.929	79.37
		None	Savitzky-Golay	0.958	82.82
		FD	Savitzky-Golay	0.871	81.79
		FD	None	0.844	77.21
		SD	Savitzky-Golay	0.960	76.4
		SD	None	0.900	77.01
	None	None	None	0.904	56.48
		None	Savitzky-Golay	0.938	84.03
		FD	Savitzky-Golay	0.947	74.75
		FD	None	0.851	81.47
		SD	Savitzky-Golay	0.940	80.52
		SD	None	0.836	76.25

All raw spectra were processed by using baseline correction and normalization were defined as raw* spectra in this study.

Table S3. Comparison of different classification models on the training set (n = 540)

Classification models	Full spectra		SPA proceed data		CARS proceed data	
	Cross-validation score	Processing time, s	Cross-validation score	Processing times, s	Cross-validation score	Processing time, s
AdaBoost	0.902	56.87	0.929	0.90	0.960	4.83
KNN	0.918	38.13	0.956	0.78	0.951	2.65
SVC	0.916	48.75	0.933	0.65	0.931	2.71
GNB	0.776	27.26	0.882	0.37	0.878	1.48
RF	0.909	33.29	0.962	0.41	0.947	2.00
DT	0.907	41.54	0.956	0.33	0.933	2.84

Note: Full spectra data were based on the raw* spectra. SPA and CARS proceed data were based on the SNV&SD&SG pre-processed spectral data.

Table S4. Comparison of the prediction results of different models on the test set (n = 90)

Classification models	Accuracy	Macro-precision	Macro-recall	Macro-F1	Processing time, s
SPA-AdaBoost	0.867	0.886	0.867	0.854	1.57
SPA-KNN	0.956	0.964	0.956	0.956	0.90
SPA-SVC	0.944	0.954	0.944	0.944	0.80
SPA-GNB	0.789	0.837	0.789	0.787	0.38
SPA-RF	0.722	0.768	0.722	0.708	0.46
SPA-DT	0.767	0.805	0.767	0.753	0.46

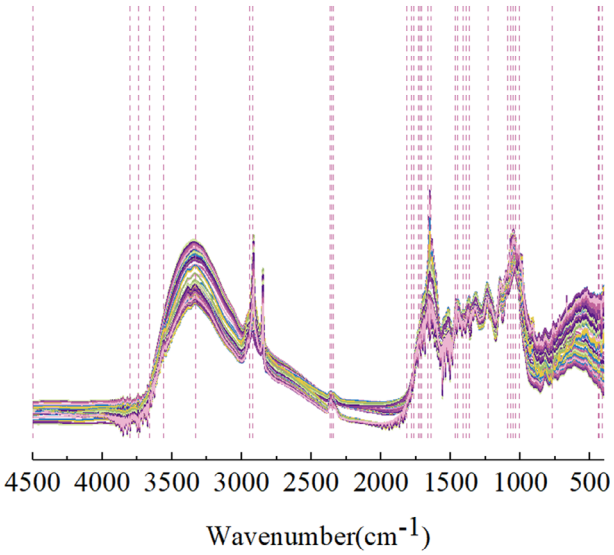


Fig. S1. The wavenumbers selected by SPA: tea + white sugar + paraffin wax

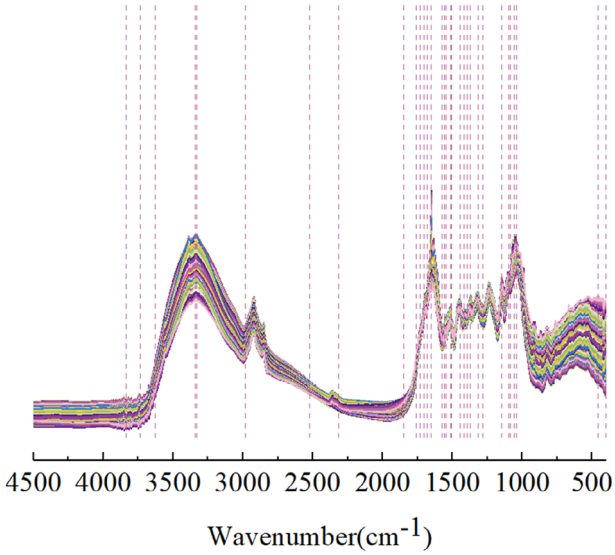


Fig. S2. The wavenumbers selected by SPA: tea + white sugar

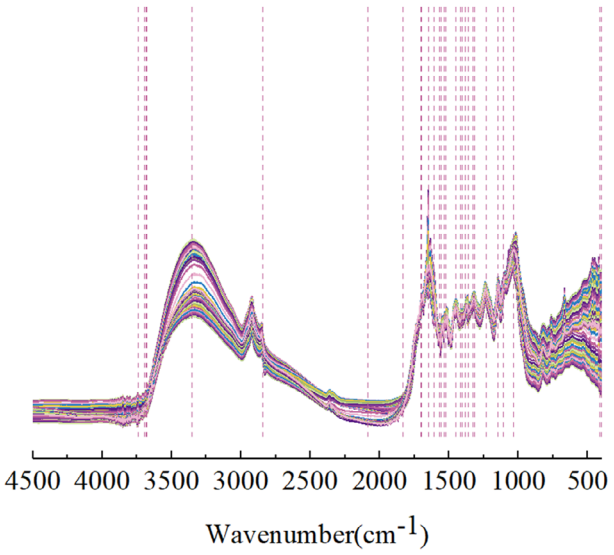


Fig. S3. The wavenumbers selected by SPA: tea + talcum powder

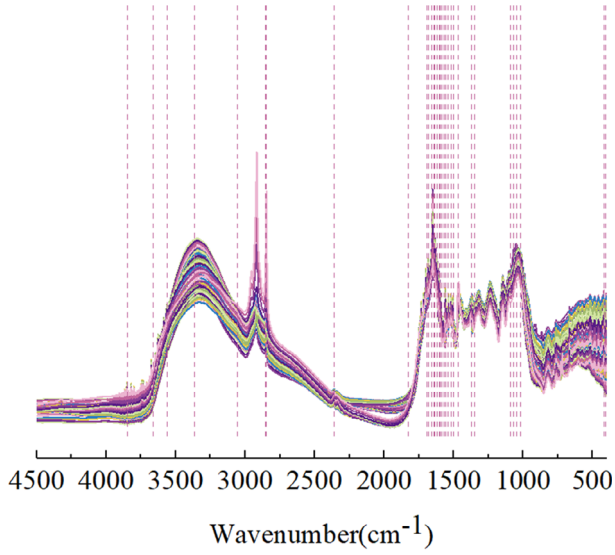


Fig. S4. The wavenumbers selected by SPA: tea + sulphur

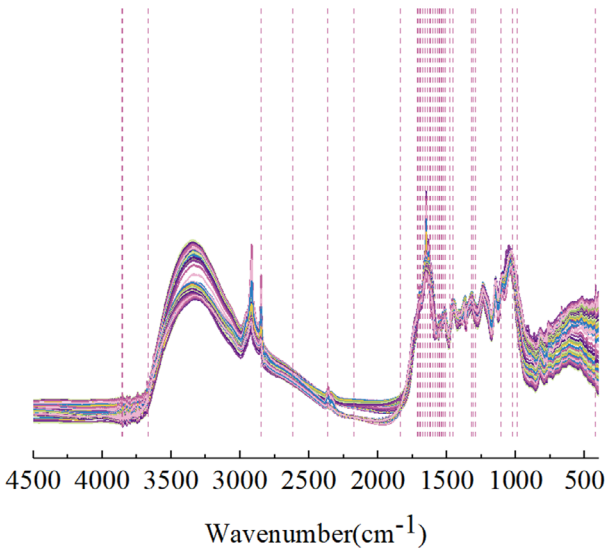


Fig. S5. The wavenumbers selected by SPA: tea + food colouring

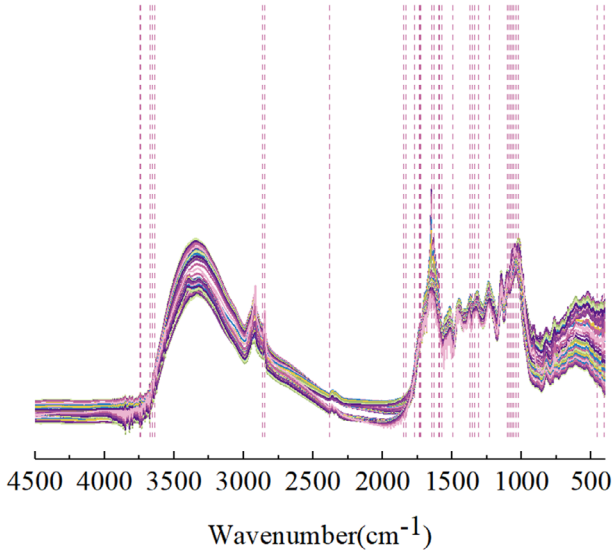


Fig. S6. The wavenumbers selected by SPA: tea + tea flavouring

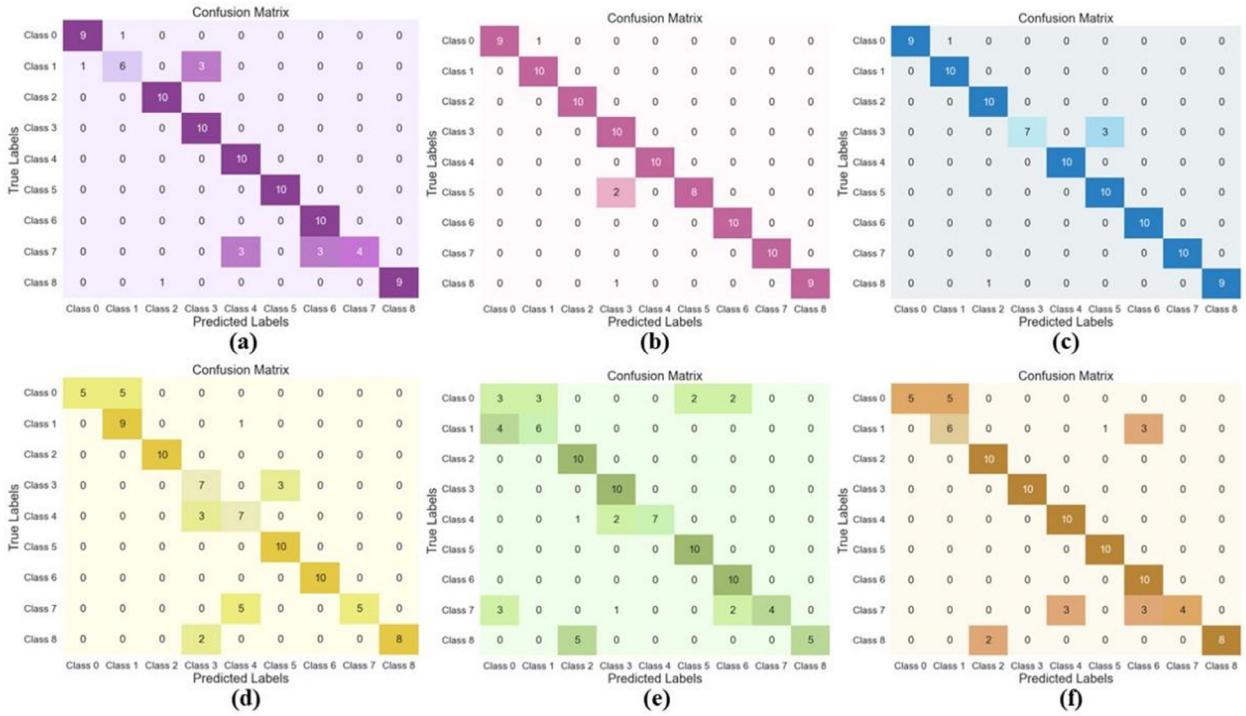


Fig. S7. Confusion matrix of different models on the test set: (a) SPA-AdaBoost, (b) SPA-KNN, (c) SPA-SVC, (d) SPA-GNB, (e) SPA-RF, (f) SPA-DT. Class 0 is pure tea, Class 1 ~ 8 are adulterated with white sugar, talcum powder, sulphur, paraffin wax, food colouring, tea flavouring, white sugar + paraffin wax and talcum powder + paraffin wax + tea flavouring

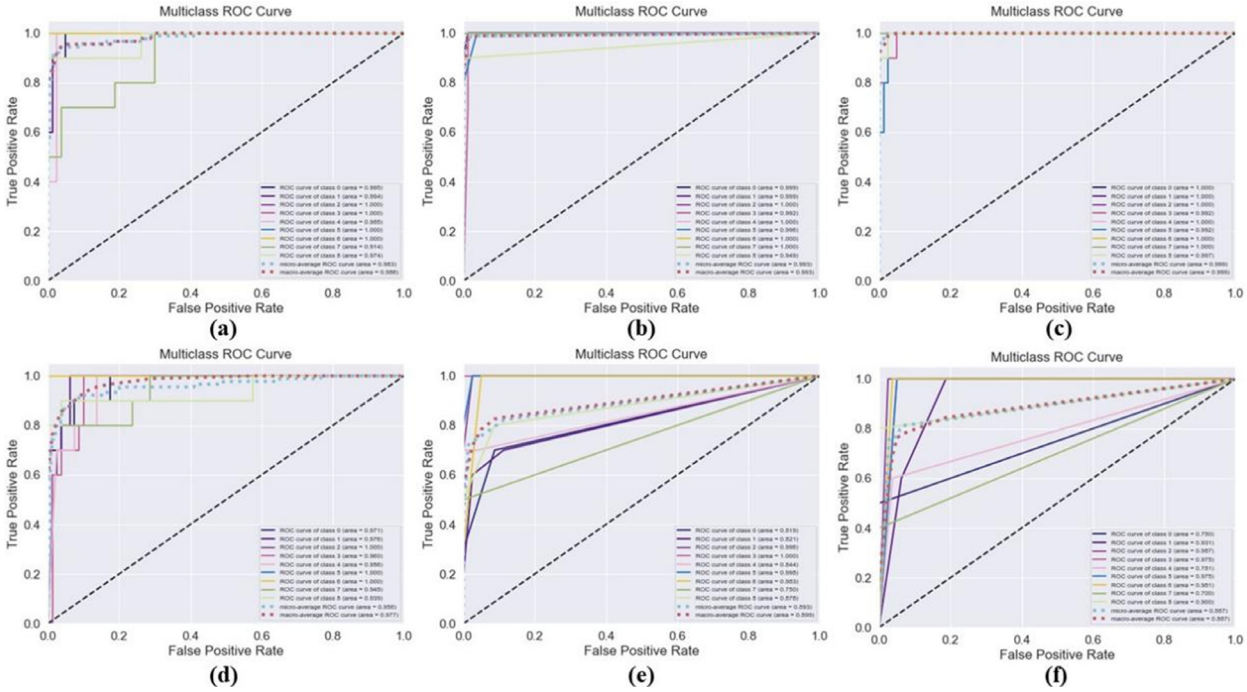


Fig. S8. The ROC curve of different models on the test set: (a) SPA-AdaBoost, (b) SPA-KNN, (c) SPA-SVC, (d) SPA-GNB, (e) SPA-RF, (f) SPA-DT. Class 0 is pure tea, Class 1 ~ 8 are tea adulterated with white sugar, talcum powder, sulphur, paraffin wax, food colouring, tea flavouring, white sugar + paraffin wax and talcum powder + paraffin wax + tea flavouring

DECLARATIONS

Data statement

All data supporting this study has been included in this manuscript.

Ethical Approval

Not applicable.

Competing Interests

The authors declare that they have no conflicts of interest.

OPEN ACCESS

REFERENCES

Amsaraj, R., Mutturi, S. (2023). Rapid detection of sunset yellow adulteration in tea powder with variable selection coupled to machine learning tools using spectral data. J. Food Sci. Technol., 60(5), 1530–1540. http://doi.org/10.1007/s13197-023-05694-3

Barbosa, R. M., Nacano, L. R., Freitas, R., Batista, B. L., Barbosa Jr, F. (2014). The use of decision trees and naïve Bayes algorithms and trace element patterns for

controlling the authenticity of free-range-pastured hens' eggs. J. Food Sci., 79(9), C1672–C1677. https://doi.org/10.1111/1750-3841.12577

Barnes, R., Dhanoa, M. S., Lister, S. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc., 43(5), 772–777. https://doi.org/10.1366/0003702894202201

Borkin, D., Némethová, A., Michaľčonok, G., Maiorov, K. (2019). Impact of data normalization on classification model accuracy. Res. Pap. – Fac. Mater. Sci. Technol., Slovak Univ. Technol., 27(45), 79–84. https://doi.org/10.2478/rput-2019-0029

Brereton, R. G., Lloyd, G. R. (2010). Support vector machines for classification and regression. Analyst, 135(2), 230–267. https://doi.org/10.1039/B918972F

Chen, S., Wang, Y., Zhu, Q., Ni, H., Cai, H. (2023). Fast recognition of the harvest period of Porphyra haitanensis based on mid-infrared spectroscopy and chemometrics. J. Food Meas. Charact., 17(5), 5487–5496. https://doi.org/10.1007/s11694-023-01999-1

Da Costa Filho, P. A., Chen, Y., Cavin, C., Galluzzo, R. (2022). Mid-infrared spectroscopy: Screening method for analysis of food adulterants in reconstituted skimmed milk powder. Food Control, 136, 108884. https://doi.org/10.1016/j.foodcont.2022.108884

Ding, Y., Yan, Y., Li, J., Chen, X., Jiang, H. (2022). Classification of Tea Quality Levels Using Near-Infrared Spectroscopy Based on CLPSO-SVM. Foods, 11(11), 1658. https://doi.org/10.3390/foods11111658

Donald, I., Lee, S. (1954). Infrared analysis of solids by potassium bromide pellet technique. Anal. Chem., 26(11), 1765–1768. https://doi.org/10.1021/ac60095a023

Ellis, D. I., Broadhurst, D., Goodacre, R. (2004). Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform infrared spectroscopy and machine learning. Anal. Chim. Acta, 514(2), 193–201. https://doi.org/10.1016/j.aca.2004.03.060

Ellis, D. I., Broadhurst, D., Kell, D. B., Rowland, J. J., Goodacre, R. (2002). Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. Appl. Environ. Microbiol., 68(6), 2822–2828. https://doi.org/10.1128/AEM.68.6.2822-2828.2002

Esteki, M., Nouroozi, S., Shahsavari, Z. (2016). A fast and direct spectrophotometric method for the simultaneous determination of methyl paraben and hydroquinone in cosmetic products using successive projections algorithm. Int. J. Cosmet. Sci., 38(1), 25–34. https://doi.org/10.1111/ics.12241

Gao, Z., Chen, S., Huang, J., Cai, H. (2024a). Real-time quantitative detection of hydrocolloid adulteration in meat based on Swin Transformer and smartphone. J. Food Sci., 89(7), 4359–4371. https://doi.org/10.1111/1750-3841.17159

Gao, Z., Huang, J., Chen, J., Shao, T., Ni, H., Cai, H. (2024b). Deep transfer learning-based computer vision for real-time harvest period classification and impurity detection of Porphyra haitnensis. Aquacult. Int., 32, 5171–5198. https://doi.org/10.1007/s10499-024-01422-6

Gao, Z., Lin, Q., He, Q., Liu, C., Cai, H., Ni, H. (2024c). Rapid Detection of Spoiled Apple Juice Using Electrical Impedance Spectroscopy and Data Augmentation-Based Machine Learning. Chiang Mai J. Sci., 51(5), e2024071. https://doi.org/10.12982/CMJS.2024.071

He, W., Zhou, J., Cheng, H., Wang, L., Wei, K., Wang, W., Li, X. (2012). Validation of origins of tea samples using partial least squares analysis and Euclidean distance method with near-infrared spectroscopy data. Spectrochim. Acta, Part A, 86, 399–404. https://doi.org/10.1016/j.saa.2011.10.056

Huang, Z., Xiao, Y., Xiao, Y., Cai, H., Ni, H. (2023). Rapid recognition of processed milk type using electrical impedance spectroscopy and machine learning. Int. J. Food Sci. Technol., 58(6), 3121–3134. https://doi.org/10.1111/ijfs.16440

Jun, Y., Chunling, Y., Xu, M., Xiangru, M., Zhimin, L., Leqian, H. (2021). Rapid discrimination of adulteration in Radix Astragali combining diffuse reflectance mid-infrared Fourier transform spectroscopy with chemometrics. Spectrochim. Acta, Part A, 248, 119251. https://doi.org/10.1016/j.saa.2020.119251

Li, H., Liang, Y., Xu, Q., Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta, 648(1), 77–84. https://doi.org/10.1016/j.aca.2009.06.046

Li, X., Zhang, Y., He, Y. (2016). Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics. Sci. Rep., 6(1), 30313. https://doi.org/10.1038/srep30313

Li, Y., Yang, X. (2023). Quantitative analysis of near infrared spectroscopic data based on dual-band transformation and competitive adaptive reweighted sampling. Spectrochim. Acta, Part A, 285, 121924. https://doi.org/10.1016/j.saa.2022.121924

Lin, M., Sun, L., Kong, F., Lin, M. (2021). Rapid detection of paraquat residues in green tea using surface-enhanced Raman spectroscopy (SERS) coupled with

gold nanostars. Food Control, 130, 108280. https://doi.org/10.1016/j.foodcont.2021.108280

Liu, H., Liu, H., Li, J., Wang, Y. (2023). Rapid and Accurate Authentication of Porcini Mushroom Species Using Fourier Transform Near-Infrared Spectra Combined with Machine Learning and Chemometrics. ACS Omega, 8(22), 19663–19673. https://doi.org/10.1021/acsomega.3c01229

Ma, D., Liu, G., Ou, Q., Yu, H., Li, H., Shi, Y.-m. (2018). Discrimination of common wild mushrooms by FTIR and two-dimensional correlation infrared spectroscopy. Spectrosc. Spect. Anal., 38(7), 2113–2122. DOI: 10.3964/j.issn.1000-0593(2018)07-2113-10

Maraschin, M., Somensi-Zeggio, A., Oliveira, S. K., Kuhnen, S., Tomazzoli, M. M., …, Rocha, M. (2016). Metabolic Profiling and Classification of Propolis Samples from Southern Brazil: An NMR-Based Platform Coupled with Machine Learning. J. Nat. Prod., 79(1), 13–23. https://doi.org/10.1021/acs.jnatprod.5b00315

Mohammadi, M., Khorrami, M. K., Vatanparast, H., Karimi, A., Sadrara, M. (2022). Classification and determination of sulfur content in crude oil samples by infrared spectrometry. Infrared Phys. Technol., 127, 104382. https://doi.org/10.1016/j.infrared.2022.104382

Ng, K. H., Heng, A., Osborne, M. (2012). Quantitative analysis of perfumes in talcum powder by using headspace sorptive extraction. J. Sep. Sci., 35(5-6), 758–762. https://doi.org/10.1002/jssc.201100918

Nnorom, I. (2011). Trace metals in cosmetic facial talcum powders marketed in Nigeria. Toxicol. Environ. Chem., 93(6), 1135–1148. https://doi.org/10.1080/02772248.2011.577075

Özdemir, İ. S., Firat, E. Ö., Özturk, T., Zomp, G., Arici, M. (2024). Geographical origin determination of the PDO hazelnut (cv. Giresun Tombul) by chemometric analysis of FT-NIR and Raman spectra acquired from shell and kernel. J. Food Sci., 89(8), 4806–4822. https://doi.org/10.1111/1750-3841.17214

Qiu, J., Lin, Y., Wu, J., Xiao, Y., Cai, H., Ni, H. (2024). Rapid beef quality detection using spectra pre-processing methods in electrical impedance spectroscopy and machine learning. Int. J. Food Sci. Tech., 59(3), 1624–1634. https://doi.org/10.1111/ijfs.16915

Rinnan, Å., Van Den Berg, F., Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. TrAC, Trends Anal. Chem., 28(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Shao, L., Griffiths, P. R. (2007). Automatic baseline correction by wavelet transform for quantitative open-path Fourier transform infrared spectroscopy. Environ. Sci. Technol., 41(20), 7054–7059. https://doi.org/10.1021/es062188d

Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. Comput. Electron. Agr., 160, 153–159. https://doi.org/10.1016/j.compag.2019.03.004

Sun, X., Wu, H., Liu, Z., Chen, Y., Liu, Q., Ding, Y.-J., Ru, R.-Q. (2019). Rapid and Sensitive Detection of Multi-Class Food Additives in Beverages for Quality Control by Using HPLC-DAD and Chemometrics Methods. Food Anal. Methods, 12(2), 381–393. https://doi.org/10.1007/s12161-018-1370-3

Tang, R., Chen, X., Li, C. (2018). Detection of Nitrogen Content in Rubber Leaves Using Near-Infrared (NIR) Spectroscopy with Correlation-Based Successive Projections Algorithm (SPA). Appl. Spectrosc., 72(5), 740–749. https://doi.org/10.1364/AS.72.000740

Triguero, I., García-Gil, D., Maillo, J., Luengo, J., García, S., Herrera, F. (2019). Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery, 9(2), e1289. https://doi.org/10.1002/widm.1289

Valand, R., Tanna, S., Lawson, G., Bengtström, L. (2020). A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. Food Addit. Contam. A., 37(1), 19–38. https://doi.org/10.1080/19440049.2019.1675909

Wang, S., Gao, R., Wang, L. (2016). Bayesian network classifiers based on Gaussian kernel density. Expert Syst. Appl., 51, 207–217. https://doi.org/10.1016/j.eswa.2015.12.031

Whitbeck, M. R. (1981). Second Derivative Infrared Spectroscopy. Appl. Spectrosc., 35(1), 93–95. https://doi.org/10.1366/0003702814731851

Wu, X., Zhou, J., Wu, B., Sun, J., Dai, C. (2019). Identification of tea varieties by mid-infrared diffuse reflectance spectroscopy coupled with a possibilistic fuzzy c-means clustering with a fuzzy covariance matrix. J. Food Process Eng., 42(8), e13298. https://doi.org/10.1111/jfpe.13298

Xia, H., Chen, W., Hu, D., Miao, A., Qiao, X., …, Ma, Ch. (2024). Rapid discrimination of quality grade of black tea based on near-infrared spectroscopy (NIRS), electronic nose (E-nose) and data fusion. Food Chem., 440, 138242. https://doi.org/10.1016/j.foodchem.2023.138242

Xiang, Z. (2011). Isolation and identification of cytoderm peptidoglycan from Lactobacillus plantarum. J. Nat.

Sci. Heilongjiang Univ., 28(1), 113–116. https://doi.org/10.13482/j.issn1001-7011.2011.01.029

Xiao, Y., Cai, H., Ni, H. (2024). Identification of geographical origin and adulteration of Northeast China soybeans by mid-infrared spectroscopy and spectra augmentation. J. Consum. Prot, Food Saf., 19(1), 99–111. https://doi.org/10.1007/s00003-023-01471-8

Ying, C., Qi-Guang, M., Jia-Chen, L., Lin, G. (2013). Advance and prospects of AdaBoost algorithm. Acta Autom. Sin., 39(6), 745–758. https://doi.org/10.1016/S1874-1029(13)60052-X

Zhang, J., Wu, X., He, C., Wu, B., Zhang, S., Sun, J. (2024a). Near-Infrared Spectroscopy Combined with Fuzzy Improved Direct Linear Discriminant Analysis for Nondestructive Discrimination of Chrysanthemum Tea Varieties. Foods, 13(10), 1439. https://doi.org/10.3390/foods13101439

Zhang, W., Sanaeifar, A., Ji, X., Luo, X., Guo, H., …, He, Y. (2024b). Data-driven optimization of nitrogen fertilization and quality sensing across tea bud varieties using near-infrared spectroscopy and deep learning. Comput. Electron. Agric., 222, 109071. https://doi.org/10.1016/j.compag.2024.109071

Zhang, X., Diao, M., Zhang, Y. (2023). A review of the occurrence, metabolites and health risks of butylated hydroxyanisole (BHA). J. Sci. Food Agric., 103(13), 6150–6166. https://doi.org/10.1002/jsfa.12676

Zhuang, X., Wang, L., Chen, Q., Wu, X., Fang, J. (2017). Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools. Sci. China: Technol. Sci., 60, 84–90. https://doi.org/10.1007/s11431-016-0464-0

Zimmermann, B., Kohler, A. (2013). Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. Appl. Spectrosc., 67(8), 892–902. https://doi.org/10.1366/12-06723