Innovative Food Technologies 13(1) (2025) 1-12

Iranian Research Organization for Science and Technology (IROST)

Innovative Food Technologies

journal homepage: https://jift.irost.ir

Analyzing cinnamon spice adulteration with spectroscopy: the influence of data preprocessing on multivariate prediction models

Mohammad Masoudi[a], Rasool Khodabakhshian [a,*], Mahmood Reza Golzarian[b]

[a] *Department of Biosystem Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*
[b] *School of Computer Science and Information Technology, Murdoch University, Perth, Australia*

ARTICLE INFO

Keywords:
Adulterants,
chemometrics,
data preprocessing,
food safety,
spectral analysis.

ABSTRACT

Detecting fraud in the cinnamon supply chain is critical for ensuring consumer safety and maintaining product integrity. Recent advances in spectral data preprocessing techniques offer enhanced accuracy in identifying adulterants in spices like cinnamon. This study investigates the impact of different spectral preprocessing techniques on predicting adulterants—specifically soybean powder, hazelnut shell powder, and dry bread powder—mixed with cinnamon powder using spectroscopy combined with multivariate analysis. The transmittance spectra were collected across the mid-infrared range of 2–4000 cm^{-1}, and Partial Least Squares Regression (PLSR) was employed to model the adulteration levels based on these spectra. Various preprocessing methods were applied to optimize the spectral data. Among them, orthogonal signal correction (OSC) combined with detrending yielded the highest predictive accuracy, with a coefficient of prediction (R^2p) ranging from 0.900 to 0.981. Conversely, Extended Multiplicative Scatter Correction (EMSC) and Savitzky-Golay second derivative (D^2) were less effective, with R^2p values between 0.115 and 0.931. Soybean powder was the easiest adulterant to detect, with a prediction error range of 5–10%. These findings underscore the importance of selecting appropriate preprocessing techniques to improve the accuracy of fraud detection in cinnamon powder using spectroscopic methods.

1. Introduction

Food fraud, particularly in the spice industry, has emerged as a significant global concern, jeopardizing consumer health and undermining market integrity [1,2]. Spices, including cinnamon, are highly valued not only for their flavor and aroma but also for their numerous health benefits [3,4]. The high market value of authentic spices makes them susceptible to adulteration, where cheaper substances are added to enhance weight or appearance [5,6]. This practice can mislead consumers, posing potential health risks due to the introduction of harmful additives or low-quality materials [7,8]. Furthermore, food fraud can severely impact the economy, as it erodes consumer trust and diminishes the reputation of legitimate producers. As the complexity and prevalence of food fraud continue to rise, effective detection methods become imperative to protect consumer interests and ensure food quality.

Fourier-transform infrared (FTIR) spectroscopy has gained prominence as a powerful tool for detecting food adulteration, thanks to its ability to provide detailed molecular information about food samples [9–11]. FTIR spectroscopy is based on the principle of measuring the absorption of infrared light by a sample, producing a unique spectral fingerprint that corresponds to its molecular composition [12,13]. This technique offers several advantages, including non-destructive analysis, rapid data acquisition, and minimal sample preparation requirements [14]. Additionally, FTIR can analyze solid, liquid, and gas samples across a broad range of wavelengths, making it highly versatile for various applications in food science [15–17]. Its ability to identify specific functional groups in chemical compounds enables researchers to differentiate between pure and adulterated products, making FTIR a valuable method for quality control in the food industry [18].

In recent years, the application of FTIR spectroscopy in food authentication has expanded significantly, particularly in the detection of adulterants in spices [19–21]. Numerous studies have demonstrated its effectiveness in identifying various adulterants in spice products, providing a rapid and reliable means of ensuring product integrity [22–24]. For instance, FTIR has been successfully employed to differentiate between pure spices and those mixed with cheaper fillers or artificial additives, thus safeguarding consumer interests. The integration of FTIR with advanced data analysis techniques, such as multivariate analysis, enhances its capability to classify samples based on spectral data [25–27]. By applying sophisticated algorithms to process the spectral information, researchers can improve the accuracy of fraud detection, enabling the identification of specific adulterants and their concentrations in spice products [28].

Cinnamon, a popular spice derived from the inner

---

* Corresponding author: khodabakhshian@um.ac.ir

DOI: http://dx.doi.org/10.22104/IFT.2025.7722.2223
(Received: 03 July 2025, Received in revised form: 03 August 2025, Accepted: 24 August 2025)

bark of Cinnamomum trees, is not only valued for its culinary uses but also for its health-promoting properties, including antioxidant and anti-inflammatory effects [29]. Unfortunately, the rising demand for cinnamon has led to increased incidents of adulteration, where lower-quality cinnamon varieties, such as Cassia, are often substituted for authentic Ceylon cinnamon [30]. This fraudulent practice can mislead consumers and diminish the health benefits associated with genuine cinnamon. The complexity of cinnamon's chemical composition and the presence of similar aromatic compounds in adulterants complicate detection efforts. Therefore, it is crucial to develop robust analytical methods that can effectively differentiate between pure and adulterated cinnamon, ensuring that consumers receive authentic products and promoting confidence in the spice market [31].

This study aims to investigate the influence of data preprocessing techniques on the predictive accuracy of multivariate models for detecting cinnamon adulteration using FTIR spectroscopy. By employing various preprocessing methods, the research seeks to enhance the reliability of spectral data analysis, enabling more accurate identification of adulterants such as soybean powder, hazelnut shell powder, and dried bread powder in cinnamon products. This research presents a comprehensive evaluation of 14 data preparation methodologies employed in FTIR spectral analysis for the identification of adulteration in cinnamon powder. Furthermore, the inclusion of three adulterants (soybean powder, hazelnut shell powder, and dry bread powder) contributes to the broader generalizability of the results. based on authors' knowledge, these specific adulterants have very limitted previously investigated in the context of spice adulteration specially in cinnamon powder, making this research a novel contribution to the field. The findings of this study will contribute to improving food quality assurance practices, providing insights into the effectiveness of FTIR spectroscopy in combating food fraud, and ultimately supporting consumer health and safety.

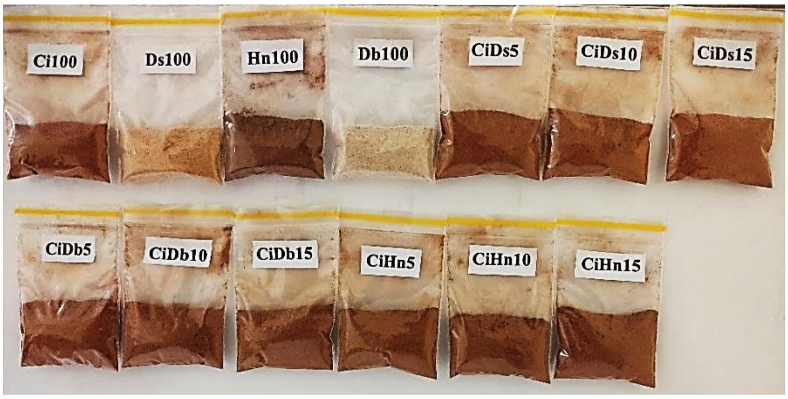


Fig 1. Prepared samples, Ci (cinnamon powder), Ds (soybean powder), Hn (hazelnut shell powder), Db (dry bread powder).

2. Materials and methods

2.1. Collection and preparation of samples

In this study, various ingredients were acquired from local markets in Mashhad, Iran, specifically focusing on their authenticity and quality for food analysis. The primary ingredient used was Cinnamomum verum, commonly known as cinnamon, which was selected for its exceptional quality. The preparation involved a meticulous cleaning process to eliminate contaminants, followed by an air-drying phase that preserved its natural characteristics. After drying, the cinnamon was finely processed using a high-precision grinder to obtain a standard cinnamon powder, which served as the baseline for comparison in our experiments.

To simulate common adulteration practices, this study also utilized several potential adulterants, including roasted soybean grains, cleaned and ground hazelnut shells, and powdered dried bread. Each adulterant was processed according to established methods documented in previous studies, including those by Hashemi-Nasab et al. (2023) and Khodabakhshian et al. (2021), with necessary adaptations made to suit the parameters of our investigation [32,33]. The soybean was lightly toasted before grinding to reflect typical processing conditions, while the hazelnut shells were dried and pulverized into a consistent powder form. The dried bread was ground and sieved to achieve a uniform texture compatible with the other ingredients.

All the selected samples underwent a standard grinding procedure to ensure uniform particle size. Following this, they were passed through a mesh sieve to maintain consistent particle distribution, which is critical for accurate analysis. To create mixed samples, the cinnamon powder was combined with varying proportions of the adulterants—5%, 10%, and 15% by weight. Fig 1. presents the various prepared mixtures. The composition of the sample groups, including pure and adulterated powders, along with the number of replicates, is summarized in Table 1. This blending was executed thoroughly to achieve a homogeneous mixture, which is representative of actual adulteration practices. Each adulterated mixture was prepared in triplicate, including controls of pure cinnamon powder. The final samples were securely packaged in airtight bags to shield them from external factors like moisture and light. They were stored in a controlled environment to preserve their integrity until the spectroscopic evaluations were conducted.

Table 1. Composition of sample groups and number of replicates

Sample NO.	Description	Adulterant Level (% w/w)	No. of Unique Samples	Replicates per Sample	Weight of per sample	Total Measurements
1	Pure cinnamon powder (Ci)	0%	1	5	10g	5
2	Pure soybean powder (Ds)	100%	1	5	10g	5
3	Pure hazelnut shell powder (Hn)	100%	1	5	10g	5
4	Pure dried bread powder (Db)	100%	1	5	10g	5
5	Cinnamon + Soybean powder (CiDs)	5%, 10%, 15%	3	5	10g	15
6	Cinnamon + Hazelnut shell powder (CiHn)	5%, 10%, 15%	3	5	10g	15
8	Cinnamon + Dried bread powder (CiDb)	5%, 10%, 15%	3	5	10g	15
Total			13			65

2.2. Spectra data acquisition

In this study, an Avatar 370 FTIR spectrometer (Thermo Nicolet Corporation, Madison, WI, USA) equipped with a deuterated triglycine sulfate (DTGS) detector and a KBr beamsplitter was employed to detect adulteration in cinnamon powder. The spectrometer operates in the mid-infrared range (4000–400 cm^{-1}) with a spectral resolution of 4 cm^{-1} and a wavenumber accuracy of ±0.01 cm^{-1}. The system includes an attenuated total reflectance (ATR) accessory with a diamond crystal, allowing for minimal sample preparation and direct analysis of powdered samples.

Each spectrum was obtained by averaging 32 scans per sample to enhance the signal-to-noise ratio (SNR). The instrumental setup and spectral acquisition process are illustrated in Fig 2. To ensure data reproducibility, the spectrometer was calibrated weekly using a polystyrene standard, and background scans were performed before each measurement session.

For data acquisition and initial preprocessing, OMNIC spectroscopy software (Thermo Nicolet Corporation, Madison, WI, USA) was used. This software supports real-time spectral correction, baseline adjustment, and noise reduction. Further, advanced preprocessing was conducted using Unscrambler X V10.4 (Camo Analytics, Oslo, Norway), as detailed in the following sections.

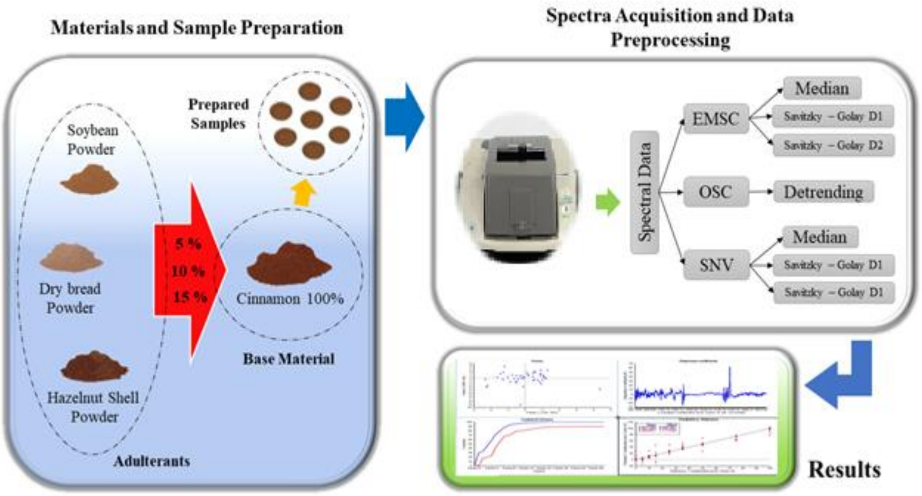
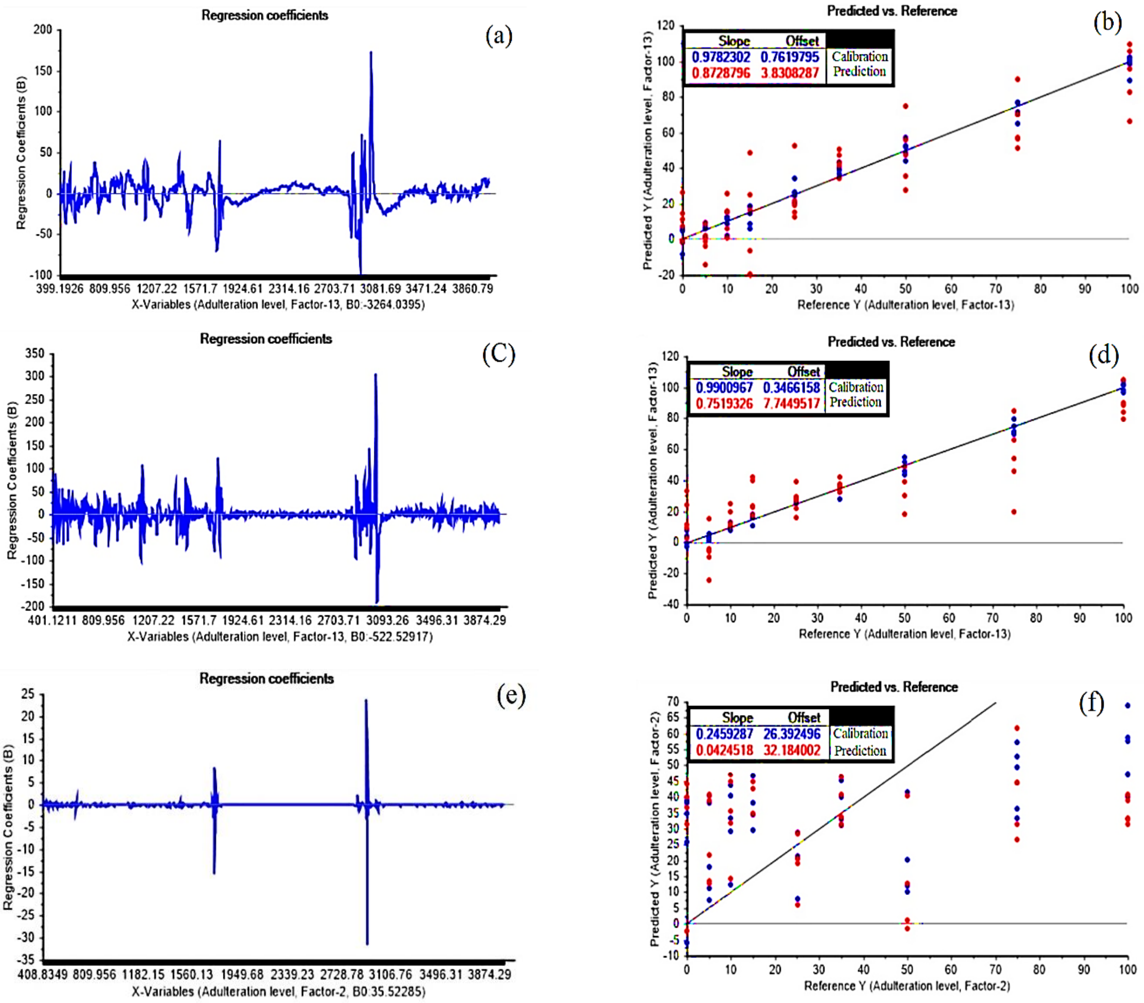


Fig 2. A schematic diagram of the proposed procedure.

2.3. Spectral preprocessing

To ensure accurate multivariate modeling for detecting adulteration in cinnamon powder, several spectral preprocessing techniques were applied. These preprocessing steps were conducted using the Unscrambler X software (version 10.4, CAMO Software AS, Bedford, Massachusetts, United States). Initially, orthogonal signal correction (OSC) was used to remove variations unrelated to the cinnamon adulteration while preserving the primary information relevant to the spectral analysis. OSC is particularly effective in eliminating noise or irrelevant variability in spectral data, which could obscure the detection of adulterants such as dry bread powder. Following OSC, standard normal variate (SNV) transformation was implemented. SNV reduces scattering effects that may arise from inconsistencies in sample morphology or particle size, ensuring that the spectral data are more uniform and interpretable across different samples [37]. Additionally, extended multiplicative scatter correction (EMSC) was applied to further correct for scatter variations while incorporating adjustments for physical and chemical interactions affecting the absorbance spectra [38].

The preprocessing strategy also included smoothing techniques to reduce noise and enhance signal clarity. A median filter was first employed to remove outlier spikes, followed by Savitzky-Golay smoothing, which was applied in both the first (D^1) and second (D^2) derivative forms to further improve peak resolution. The Savitzky-Golay algorithm, which fits successive polynomial functions to localized spectral segments, is particularly useful for reducing random noise while preserving important features such as peak intensity and width. By applying D^1 and D^2 derivatives, baseline shifts were effectively corrected, and overlapping spectral peaks were resolved, making it easier to distinguish between pure and adulterated cinnamon samples [34]. The combination of these preprocessing methods enabled the generation of a clean and reliable dataset, suitable for subsequent multivariate analysis, particularly in the context of the Partial Least Squares Regression (PLSR) model [39].



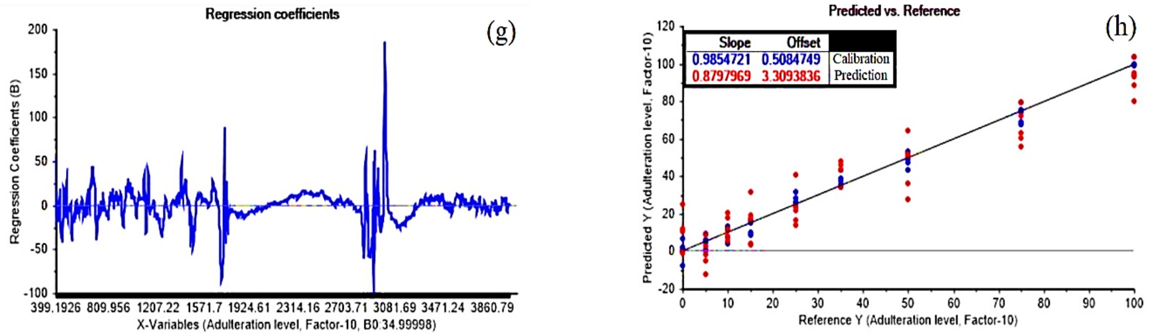


Fig 3. Spectra and quantification models (pure cinnamon and adulterated with dry bread powder) developed by PLSR, with (a, b) no preprocessing, (c, d) SNV with Savitzky-Golay D^1, (e, f) EMSC with Savitzky-Golay D^2, (g, h) OSC with Detrending.

In addition to scatter correction and smoothing, the preprocessing approach involved detrending to address any remaining baseline drifts. Detrending removes systematic baseline variations that may occur over the course of spectral data collection, particularly when dealing with complex matrices like powdered food products. This step was essential for ensuring that the PLSR model could accurately predict adulteration levels without interference from baseline noise. The final preprocessing pipeline consisted of a combination of OSC, SNV, EMSC, and derivative transformations, followed by mean-centering of the spectral data. These preprocessing techniques were chosen based on their proven effectiveness in previous studies and were fine-tuned to achieve optimal performance for this specific dataset [40–42].

Fig 3. illustrates the effect of preprocessing methods on pure cinnamon and cinnamon adulterated with dry bread powder. The dataset was divided into 70% calibration (n=84 samples) and 30% validation (n=36 samples) sets to ensure robust model training and testing. The grouping of data points in the scatter plot reflects the distinct adulteration levels (0–30% w/w), with each cluster representing a specific concentration. The PLSR model utilized 10 latent variables, selected based on minimizing the root mean square error (RMSE) of cross-validation.

The processed data were validated using k-fold cross-validation (k=5) to assess the predictive capability of the PLSR model, confirming the preprocessing approach's impact on enhancing model accuracy and robustness. The regression coefficients (slope = 0.98, offset = 0.51) indicate strong agreement between predicted and reference values ($R^2 > 0.90$).

2.4. Partial least squares regression (PLSR)

In this study, Partial Least Squares Regression (PLSR) was employed to establish a relationship between the spectral data and the level of adulteration in cinnamon powder samples. PLSR is widely used in chemometrics due to its ability to handle collinear and high-dimensional spectral data while predicting the concentration of target compounds. By projecting both the spectral matrix (X) and the adulteration levels (Y) into a new latent space, PLSR simplifies the complex relationship between the two variables, allowing for robust calibration models. This approach was chosen because of its effectiveness in handling multivariate data with overlapping spectral regions, as well as its ability to deal with noise and variability in spectral measurements [43]. One key factor in improving the accuracy of PLSR models is variable selection, which focuses on identifying the most relevant spectral regions. By excluding non-informative variables, the model is refined, improving both its predictive power and stability [44].

The performance of the developed PLSR models was evaluated using several statistical metrics. The calibration model's accuracy was assessed through the correlation coefficient (R^2), which measures the goodness of fit between the predicted and actual adulteration levels. Additionally, the root mean square error (RMSE) was calculated for both the calibration (RMSEC) and prediction (RMSEP) datasets. Lower RMSE values indicate a more accurate and reliable model, while higher R^2 values suggest better predictive capability. Cross-validation techniques were also applied to minimize overfitting, ensuring that the PLSR models would generalize well to unseen data. This combination of metrics provided a comprehensive evaluation of model performance, confirming the effectiveness of the PLSR approach in predicting adulteration levels based on spectral information [45].

3. Results and discussion

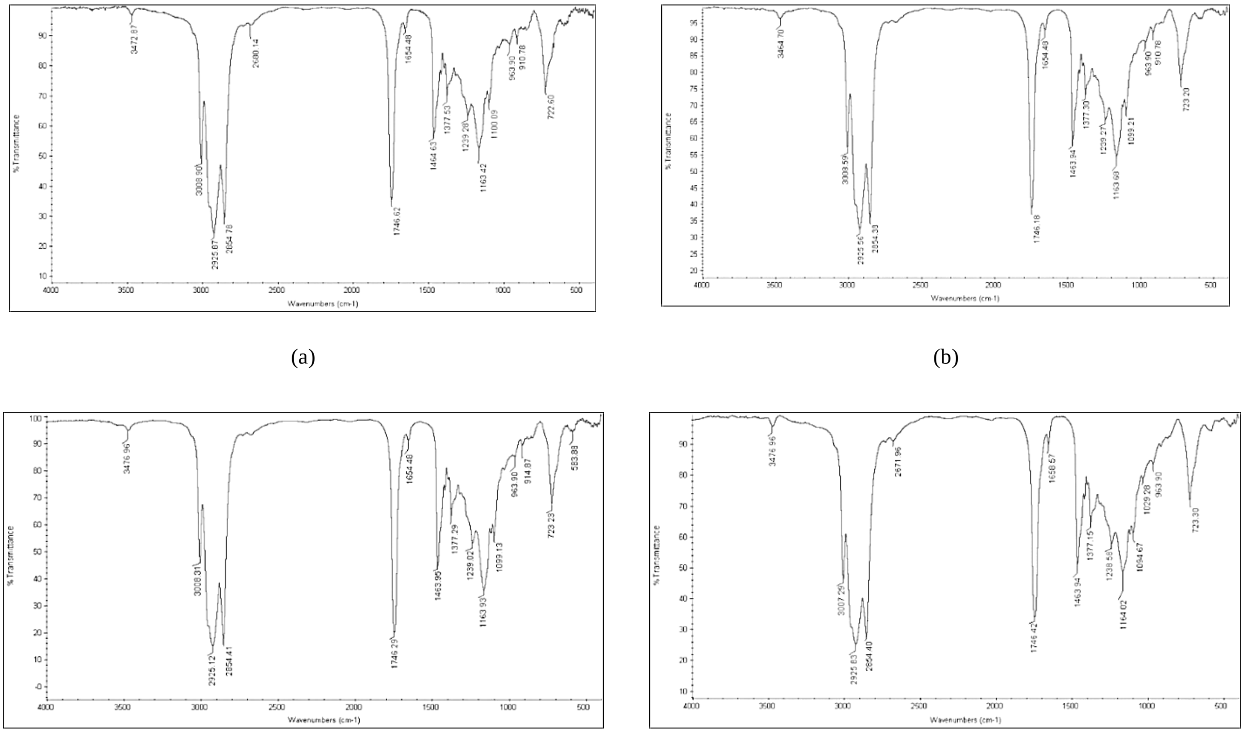
3.1. Spectral features of the samples

In Fig 4., raw spectra of samples including pure cinnamon powder, soybean powder, hazelnut shell powder, and dry bread powder in the range of 4000 – 400 cm-1 are shown. The differences between the spectra are attributed to the transmittance associated with various functional groups, which can increase or decrease depending on the sample composition. The absorption bands of water, carbohydrates, proteins, and lipids can explain this spectral treatment [46]. The known troughs in cinnamon spectra, corresponding to the stretching vibrations of C=O, C=C, and OH in aromatic compounds are approximately located at 1678 cm−1, 1626 cm−1, 1450 cm−1, and 1289 cm−1 [47]. Two prominent transmittance troughs in the nearby region of 2800 cm−1 in dry soybean suggest the symmetric stretching of fatty acid CH2 groups [48]. For hazelnut shell powder, valleys at ~2924 cm−1, ~2845 cm−1, ~1709 cm−1, and ~1613 cm−1 correspond to the vibrations of the alpha-hydroxy group (Alpha-CH), aromatic carbonyl/carboxyl C=O stretching, and aromatic C=C ring stretching, respectively [49]. The valley near 1373 cm−1 in dry bread powder is attributed to sulfone, a derivative of wheat flour [50]. As shown in Fig 4., the chemical composition of the investigated samples makes it difficult to distinguish between pure cinnamon powder and adulterated samples with the naked eye. This is because the spectra of all the samples appear very similar, with nearly parallel peaks.

To build upon this analysis, it is important to compare these findings with similar studies on food adulteration and spectral analysis. A similar study observed distinct spectral features in various food powders, demonstrating that FTIR (Fourier-transform infrared) spectroscopy can effectively identify adulterants in food products [46]. In the current study, while the transmittance spectra for adulterated and pure cinnamon powder appear nearly identical, small differences in peak intensities and locations indicate the potential for advanced multivariate analysis techniques, such as PCA (Principal Component Analysis) or PLS-DA (Partial Least Squares Discriminant Analysis), to distinguish these samples more effectively.

Another study confirmed that the presence of C=O and OH stretching vibrations in cinnamon is a reliable marker for the identification of pure samples [47]. However, the overlapping peaks with adulterated samples, as seen in this study, indicate that visual differentiation is insufficient. Similarly, a research observed that hazelnut shell powder contains strong aromatic C=C and carbonyl groups, which can interfere with identifying certain adulterants [51].

In conclusion, while the spectral analysis shows similarities across samples, more refined data analysis methods can provide higher accuracy in distinguishing pure and adulterated samples. This comparison also highlights the need for complementary techniques like NIR (near-infrared) spectroscopy, which could further improve adulteration detection.



(a)

(b)

(c)

(d)

Fig 4. Transmittance spectrum of (a) pure cinnamon powder, (b) soybean powder, (c) hazelnut shell powder, and (d) dry bread powder samples.

3.2. Evaluation of PLSR model performance

Based on a literature review, a PLSR model is considered suitable when it has a high coefficient of determination, low RMSE (<10%), and an R^2 value greater than 0.9 [52,53]. Based on the results (shown in Tables 2-4), the PLSR model demonstrated a high value of correlation coefficient of calibration (R2c) and a low root mean squared error of calibration (RMSEC) for all adulterations (R2c > 0.97 and RMSEC< 5%). However, it performed poorly in predicting adulterations, except for soybean powder (with an R^2p value of 0.938 and an error of approximately 9%). For other adulterants, the R2p value was below 0.900 and the RMSEP was above 14%. These results highlight the importance of using preprocessing methods to enhance the accuracy of fraud prediction. In the following section, the effect of preprocessing methods on performance of the PLSR model for predicting the level of frauds is investigated.

These findings are consistent with other studies that have highlighted the challenge of achieving robust prediction for multiple adulterants without preprocessing techniques. For instance, Tan et al. (2024) reported similar difficulties in predicting adulteration in cumin powder without applying spectral preprocessing methods [54]. Studies like the one conducted by Khodabakhshian et al. (2023) also emphasize that while PLSR models can yield high calibration accuracy, the prediction accuracy often depends heavily on the use of preprocessing techniques to reduce noise and enhance the quality of the spectral data [34]. Thus, as seen in the present study, the relatively low predictive performance for most adulterants supports the argument that preprocessing methods are crucial for improving fraud detection accuracy in spectroscopy-based models. This will be further explored in the following section.

3.3. Effect of preprocessing methods on PLSR model

In this section, various preprocessing techniques – including 14 methods – and their effects on the PLSR model were investigated. The evaluation results of the PLSR model and the PLSR model combined with preprocessing methods for adulterated samples are presented in Tables 2-4.

Table 2. The results of PLSR for different preprocessing techniques (cinnamon powder adulterated with hazelnut shell powder)

NO.	Preprocessing	Factor	R^2_c	RMSEC (%)	R^2_p	RMSEP (%)
1	None	15	0.993	2.823	0.867	14.800
2	OSC	5	0.995	2.355	0.971	7.571
3	SNV	18	0.996	1.562	0.841	15.827
4	EMSC	15	0.995	2.042	0.873	14.026
5	Detrending	15	0.995	2.212	0.881	14.063
6	Median filter	17	0.997	2.062	0.832	16.146
7	Savitzky-Golay D^1	16	0.999	0.945	0.782	19.642
8	Savitzky-Golay D^2	15	0.999	1.258	0.632	25.531
9	OSC, Detrending	6	0.995	2.756	0.978	7.683
10	SNV, Median filter	16	0.998	1.563	0.846	15.892
11	SNV, Savitzky-Golay D^1	15	0.999	1.405	0.768	18.376
12	SNV, Savitzky-Golay D^2	15	0.997	1.782	0.551	25.340
13	EMSC, Median filter	16	0.997	1.335	0.893	13.006
14	EMSC, Savitzky-Golay D^1	16	0.998	1.582	0.773	18.621
15	EMSC, Savitzky-Golay D^2	17	0.998	1.593	0.571	25.367

With this vision, in cinnamon powder adulterated with hazelnut shell powder, the orthogonal signal correction (OSC) method with detrending (Fig 3g, h) yielded the most accurate prediction results, with an R^2_p value of 0.978. It was able to estimate the level of hazelnut shell powder in cinnamon powder with an error of approximately 7%. Following OSC with detrending, OSC alone produced the second-best result with, an R^2_p value of 0.971. The other results can be seen in Table 2.

It is worth noting that although models with a higher number of latent factors often show better fitting performance (higher R^2c and lower RMSEC), they are more prone to overfitting and multicollinearity, especially when dealing with high-dimensional spectral data. In contrast, reducing the number of factors generally leads to higher prediction errors (e.g., increased RMSEP), yet it also simplifies the model and reduces the dimensionality of the data matrix, which is desirable for robust model interpretation and generalizability. Therefore, a trade-off must be made between model complexity and predictive performance. For instance, in Table 2, while the raw data (no preprocessing) model used 15 factors and yielded a relatively high R^2c (0.993), its prediction performance (R^2p = 0.867) was lower than that of OSC with 5 factors (R^2p = 0.971). This indicates that using fewer factors with effective preprocessing (like OSC or OSC with detrending) can lead to more accurate and stable predictions despite lower apparent model fit during calibration.

Consequently, researchers must strike a balance between the number of latent variables and prediction accuracy. A model using 5–6 factors with lower RMSEP and higher R^2p (such as OSC and OSC+Detrending) is often preferable over a highly fitted model using 15–17 factors with worse prediction errors, due to the risk of overfitting and reduced generalizability.

Table 3. The results of PLSR for different preprocessing techniques (cinnamon powder adulterated with soybean powder)

NO.	Preprocessing	Factor	R^2_c	RMSEC (%)	R^2_p	RMSEP (%)
1	None	13	0.991	3.095	0.938	8.537
2	OSC	5	0.985	3.978	0.971	6.169
3	SNV	11	0.993	2.917	0.963	7.352
4	EMSC	8	0.989	3.411	0.951	7.638
5	Detrending	10	0.988	3.587	0.932	8.922
6	Median filter	10	0.992	3.064	0.951	7.992
7	Savitzky-Golay D^1	12	0.995	2.440	0.967	7.265
8	Savitzky-Golay D^2	12	0.991	3.101	0.942	8.230
9	OSC, Detrending	5	0.987	3.714	0.981	5.773
10	SNV, Median filter	9	0.993	3.023	0.951	7.393
11	SNV, Savitzky-Golay D^1	11	0.995	2.698	0.963	6.696
12	SNV, Savitzky-Golay D^2	15	0.999	0.853	0.941	8.680
13	EMSC, Median filter	10	0.993	2.858	0.969	6.597
14	EMSC, Savitzky-Golay D^1	11	0.994	2.672	0.964	6.782
15	EMSC, Savitzky-Golay D^2	16	0.999	1.015	0.931	9.279

Table 4. The results of PLSR for different preprocessing techniques (cinnamon powder adulterated with dry bread powder)

NO.	Preprocessing	Factor	R^2_c	RMSEC (%)	R^2_p	RMSEP (%)
1	None	12	0.972	4.488	0.811	14.159
2	OSC	10	0.971	4.518	0.882	11.287
3	SNV	15	0.994	1.767	0.878	11.500
4	EMSC	12	0.990	2.971	0.886	11.069
5	Detrending	11	0.981	4.010	0.864	12.109
6	Median filter	13	0.984	3.653	0.858	12.360
7	Savitzky-Golay D^1	15	0.997	1.650	0.552	22.897
8	Savitzky-Golay D^2	3	0.321	25.880	0.212	30.996
9	OSC, Detrending	11	0.982	3.708	0.900	9.972
10	SNV, Median filter	12	0.988	3.098	0.874	11.626
11	SNV, Savitzky-Golay D^1	12	0.987	3.100	0.771	15.716
12	SNV, Savitzky-Golay D^2	2	0.332	25.678	0.195	31.107
13	EMSC, Median filter	12	0.989	3.169	0.889	10.826
14	EMSC, Savitzky-Golay D^1	15	0.989	1.122	0.793	14.796
15	EMSC, Savitzky-Golay D^2	3	0.351	25.316	0.115	31.266

As emphasized in section 2.3, we employed preprocessing methods to increase the accuracy of the prediction models. After examining the results shown in Table 2, we discovered that certain preprocessing methods, such as SNV, median filter, and SNV with median filter, underperformed when compared to the un-preprocessed data of adulterated cinnamon powder and hazelnut shell powder. Furthermore, some preprocessing methods, such as Savitzky-Golay D^2, EMSC with Savitzky-Golay D^2, and SNV with Savitzky-Golay D^2, had a significantly negative impact on prediction accuracy (Table 2). The poorest result was obtained using SNV Savitzky-Golay D^2 which yielded an R^2_p value of 0.551.

Regarding cinnamon powder adulterated with soybean powder, almost all of preprocessing methods demonstrated suitable performance and improved the R^2_p value of the PLSR models. In some cases, such as detrending and EMSC with Savitzky-Golay D^2, the accuracy was lower compared to the un-preprocessed data (Table 3). However, in this case, OSC with detrending showed the highest accuracy in prediction, with an R^2_p value of 0.981. It was able to estimate the level of soybean powder in cinnamon powder with an error of approximately 6%. Additionally, OSC produced the second-best results with an R^2_p value of 0.971. Furthermore, several methods including EMSC, median filter, and SNV with median filter illustrated similar results ($R^2_p = 0.951$) (Table 3).

For cinnamon powder adulterated with dry bread powder (Table 4), approximately half of the preprocessing methods had a negative effect, with some cases showing a significant negative impact (such as Savitzky-Golay D^2, SNV with Savitzky-Golay D^2, and

EMSC with Savitzky-Golay D^2, resulting in an $R^2_p < 0.22$). The poorest prediction accuracy was obtained using EMSC with Savitzky-Golay D^2 with an R^2_p value of 0.115 and an RMSEP of 31.266% (Fig 3e, f). However, in other cases, the preprocessing methods improved prediction accuracy compared to the un-preprocessed spectra of adulterated samples. The OSC with detrending method yielded an acceptable with a R^2_p value of 0.900 and an RMSEP=9.972%).

3.4. Comparison of preprocessing methods

It is important to keep in mind that the ultimate goal of a chemometrician is to develop a reliable model that can accurately predict results. As observed earlier, most preprocessing methods had a positive effect on prediction results and improved prediction accuracy. However, in some cases, this effect was negligible, and in some cases, it was significant. Some methods showed worse results than un-preprocessing data. Our results indicated that preprocessing techniques play a crucial role in enhancing the predictive performance of the model, but the effectiveness of each method varied. The OSC (Orthogonal Signal Correction) combined with detrending proved to be the most effective technique, reducing the prediction error to a range of 5-10%. This high accuracy suggests that OSC can effectively remove uninformative variation from the spectral data, making it highly suitable for this specific application. On the contrary, methods such as EMSC (Extended Multiplicative Signal Correction) with Savitzky-Golay D^2 did not perform well across the board. This result aligns with other studies that indicate the potential limitations of certain preprocessing techniques in handling complex datasets. For instance, a study by Liu et al. (2018) also reported poor results using EMSC in the context of powdered food adulteration, attributing the failure to the method's inability to adequately account for baseline shifts and noise [55].

When comparing our findings with those of Wu et al. (2023) and other related studies, we found that preprocessing methods such as MSC (Multiplicative Scatter Correction) and SNV (Standard Normal Variate) tend to enhance predictive accuracy, particularly in cases involving low levels of adulteration [56]. However, in our study, the application of MSC was not as effective as OSC, which might be due to the specific characteristics of cinnamon powder and the adulterants used (e.g., soybean powder, hazelnut shell powder). Similar conclusions were drawn by Khodabakhshian et al. (2022), who emphasized that the performance of preprocessing methods depends on the type of adulterant and the specific spectral properties of the sample [43].

This variability in effectiveness highlights the importance of selecting the appropriate preprocessing technique based on the characteristics of both the sample and the adulterant [57]. While prior studies offer valuable insights, relying solely on their results can lead to suboptimal outcomes. Our study confirms that preprocessing techniques must be evaluated and tailored to the specific adulteration scenario at hand, as supported by findings in related research. In summary, although there is a general consensus on the positive impact of preprocessing on prediction accuracy, our study demonstrates that not all methods are equally effective. Therefore, it is critical for researchers to carefully assess and select the most suitable preprocessing techniques for their specific case studies.

4. Conclusions

In this study, we employed FTIR spectroscopy combined with multivariate analysis to predict adulteration levels in cinnamon powder and evaluated the impact of different data preprocessing methods on prediction accuracy. We evaluated the quantitative performance of 14 preprocessing techniques and assessed their efficacies on the prediction accuracy of the PLSR model. According to the results, OSC with detrending demonstrated the best results for all of the adulterants (soybean powder, hazelnut shell powder, and dry bread powder), with an R^2_p value of greater than 0.90 and an RMSEP value of under 10% for predicting cinnamon powder adulteration. On the other hand, EMSC with Savitzky-Golay D^2 showed the lowest accuracy in predicting adulterants with an RMSEP range of 9-32%. However, in comparison with the other adulterants, the accuracy of soybean powder prediction was slightly superior. In summary, the PLSR model based on OSC with detrending could predict cinnamon powder adulteration with high accuracy and this combination performed best in all adulterants, while other integrations reduce the model’s performance compared to using un-preprocessed data. Furthermore, since the effectiveness of preprocessing approaches can vary across different adulterants, selecting an appropriate technique is essential; otherwise, it may compromise the accuracy of fraud prediction. Given the nature of soft computing techniques, it is recommended that future research investigate the use of artificial neural networks and other machine learning methods on predictive accuracy.

Conflict of interest: No conflict of interest has been declared by the authors.

Acknowledgment

The authors would like to thank Ferdowsi University of Mashhad for equipping the lab resources and pecuniary assistance.

References

[1] Gerardi, A. (2023). Global Food Safety Initiative (GFSI): underpinning the safety of the global food chain, facilitating regulatory compliance, trade, and consumer trust. In *Present knowledge in food safety* (pp. 1089-1098). Academic Press. mailto: https://doi.org/10.1016/B978-0-12-819470-6.00058-5

[2] Gwenzi, W., Makuvara, Z., Marumure, J., Simbanegavi, T. T., Mukonza, S. S., & Chaukura, N. (2023). Chicanery in the food supply chain! Food fraud, mitigation, and research needs

in low-income countries. *Trends Food Sci. Technol.*, *136*, 194-223. https://doi.org/10.1016/j.tifs.2023.03.027

[3] Chen, J., Pan, B. (2023). Food Flavors, Chem. *Funct. Prop. Food Components*, 363–400.

[4] Handayani, A., Lailaty, I. Q., Rosyidah, A. L., Sari, D. R. T., Yunarto, N., & Suherman, D. (2024). Indonesian Cinnamon (Cinnamomum burmanni (Nees & T. Nees) Blume) as promising medicinal resources: a review. *Jurnal Sylva Lestari*, *12*(3), 610-633. https://doi.org/10.23960/jsl.v12i3.929

[5] Negi, A., & Meenatchi, R. (2023). Herbs and Spices. In *Emerging Food Authentication Methodologies Using GC/MS* (pp. 253-279). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-30288-6_9

[6] Kumar, P., Tripathi, S., Islam, Z., & Shaida, B. (2023). Detection of Adulteration in Spices. Int. J. Med. Toxicol. Leg. Med., 26 (3and4), 138-142. https://doi.org/ 10.5958/0974-4614.2023.00061.X

[7] Sawyer, W. E., & Izah, S. C. (2024). Unmasking food adulteration: public health challenges, impacts and mitigation strategies. *ES General*, *4*, 1091. https://doi.org/ 10.30919/esg1091

[8] Vasu, P., & Martin, A. (2023). Chemical Adulterants in Food: Recent Challenges. In *Engineering Aspects of Food Quality and Safety* (pp. 31-52). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-30683-9_2

[9] Koczoń, P., Hołaj-Krzak, J. T., Palani, B. K., Bolewski, T., Dąbrowski, J., Bartyzel, B. J., & Gruczyńska-Sękowska, E. (2023). The analytical possibilities of FT-IR spectroscopy powered by vibrating molecules. *Int. J. Mol. Sci.*, *24*(2), 1013. https://doi.org/10.3390/ijms24021013

[10] Saji, R., Ramani, A., Gandhi, K., Seth, R., & Sharma, R. (2024). Application of FTIR spectroscopy in dairy products: A systematic review. *Food Humanity*, *2*, 100239. https://doi.org/10.1016/j.foohum.2024.100239

[11] Jamwal, R., Kumari, S., Sharma, S., Kelly, S., Cannavan, A., & Singh, D. K. (2021). Recent trends in the use of FTIR spectroscopy integrated with chemometrics for the detection of edible oil adulteration. *Vib. Spectrosc.*, *113*, 103222. https://doi.org/10.1016/j.vibspec.2021.103222

[12] Mangam, V. T., Narla, D., Konda, R. K., & Sarella, P. N. K. (2024). Beyond the spectrum: Exploring unconventional applications of fourier transform infrared (FTIR) spectroscopy. *Asian J. Pharm. Anal.*, *14*(2), 86-94. https://doi.org/ 10.52711/2231-5675.2024.00016

[13] Guerrero-Pérez, M. O., & Patience, G. S. (2020). Experimental methods in chemical engineering: Fourier transform infrared spectroscopy—FTIR. *Can. J. Chem. Eng.*, *98*(1), 25-33. https://doi.org/10.1002/cjce.23664

[14] Barnes, M., Sulé-Suso, J., Millett, J., & Roach, P. (2023). Fourier transform infrared spectroscopy as a non-destructive method for analysing herbarium specimens. *Biol. Lett.*, *19*(3), 20220546. https://doi.org/10.1098/rsbl.2022.0546

[15] Rohman, A., Ghazali, M. A. I. B., Windarsih, A., Riyanto, S., Yusof, F. M., & Mustafa, S. (2020). Comprehensive review on application of FTIR spectroscopy coupled with chemometrics for authentication analysis of fats and oils in the food products. *Molecules*, *25*(22), 5485. https://doi.org/10.3390/molecules25225485

[16] Dashti, A., Weesepoel, Y., Müller-Maatsch, J., Parastar, H., Kobarfard, F., Daraei, B., & Yazdanpanah, H. (2022). Assessment of meat authenticity using portable Fourier transform infrared spectroscopy combined with multivariate classification techniques. *Microchem. J.*, *181*, 107735. https://doi.org/10.1016/j.microc.2022.107735

[17] Apolonski, A., Roy, S., Lampe, R., & Sankar Maiti, K. (2020). Molecular identification of bio-fluids in gas phase using infrared spectroscopy. *Appl. Opt.*, *59*(17), E36-E41. https://doi.org/10.1364/AO.388362

[18] 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. Part A*, *37*(1), 19-38. https://doi.org/10.1080/19440049.2019.1675909

[19] Saadi, S., Nacer, N. E., Ariffin, A. A., Ghazali, H. M., Abdulkarim, S. M., Boo, H. C., ... & Anwar, F. (2023). Discrimination of food adulteration by means of PCR and FTIR. *Food Humanity*, *1*, 1362-1378. https://doi.org/10.1016/j.foohum.2023.10.008

[20] Ahmad, A., & Ayub, H. (2022). Fourier transform infrared spectroscopy (FTIR) technique for food analysis and authentication. In *Nondestructive quality assessment techniques for fresh fruits and vegetables* (pp. 103-142). Singapore: Springer Nature Singapore.

[21] Kaavya, R., Pandiselvam, R., Mohammed, M., Dakshayani, R., Kothakota, A., Ramesh, S. V., ... & Ashokkumar, C. (2020). Application of infrared spectroscopy techniques for the assessment of quality and safety in spices: a review. *Appl. Spectrosc. Rev.*, *55*(7), 593-611. https://doi.org/10.1080/05704928.2020.1713801

[22] Shannon, M., Lafeuille, J. L., Frégière-Salomon, A., Lefevre, S., Galvin-King, P., Haughey, S. A., ... & Elliott, C. T. (2022). The detection and determination of adulterants in turmeric using fourier-transform infrared (FTIR) spectroscopy coupled to chemometric analysis and micro-FTIR imaging. *Food Control*, *139*, 109093. https://doi.org/10.1016/j.foodcont.2022.109093

[23] Galvin-King, P., Haughey, S. A., & Elliott, C. T. (2021). Garlic adulteration detection using NIR and FTIR spectroscopy and chemometrics. *J. Food Compos. Anal.*, *96*, 103757. https://doi.org/10.1016/j.jfca.2020.103757

[24] Lixourgioti, P., Goggin, K. A., Zhao, X., Murphy, D. J., van Ruth, S., & Koidis, A. (2022). Authentication of cinnamon spice samples using FT-IR spectroscopy and chemometric classification. *LWT*, *154*, 112760. https://doi.org/10.1016/j.lwt.2021.112760

[25] Indrayanto, G., & Rohman, A. (2020). The use of FTIR spectroscopy combined with multivariate analysis in food composition analysis. Spectrosc. *Tech. Artif. Intell. Food Bever. Anal.*, 25-51. https://doi.org/10.1007/978-981-15-6495-6_2

[26] Damto, T., Zewdu, A., & Birhanu, T. (2023). Application of Fourier transform infrared (FT-IR) spectroscopy and multivariate analysis for detection of adulteration in honey markets in Ethiopia. *Curr. Res. Food Sci.*, *7*, 100565. https://doi.org/10.1016/j.crfs.2023.100565

[27] Masoudi, M., & Khodabakhshian, R. (2025). Genetic algorithm-optimized PLS for detecting adulteration in cinnamon powder via FT-IR spectroscopy. *Expert Syst. Appl.*, 128522. https://doi.org/10.1016/j.eswa.2025.128522

[28] Abbasi-Tarighat, M., Abdi, G., Heidari Ghorghosheh, F., & Shahmohammadi Bayatiyani, K. (2023). Multivariate Authentication of Herbs and spices through UV-Vis and FT-IR fingerprint. *Anal. Bioanal. Chem. Res.*, *10*(3), 301-317. https://doi.org/ 10.22036/abcr.2023.366412.1847

[29] Pagliari, S., Forcella, M., Lonati, E., Sacco, G., Romaniello, F., Rovellini, P., ... & Bruni, I. (2023). Antioxidant and anti-inflammatory effect of cinnamon (Cinnamomum verum J. Presl) bark extract after in vitro digestion simulation. *Foods*, *12*(3), 452. https://doi.org/10.3390/foods12030452

[30] Feltes, G., Ballen, S. C., Steffens, J., Paroul, N., & Steffens, C. (2023). Differentiating True and False Cinnamon: Exploring Multiple Approaches for Discrimination. *Micromachines*, *14*(10), 1819.

https://doi.org/10.3390/mi14101819

[31] Vinothkanna, A., Dar, O. I., Liu, Z., & Jia, A. Q. (2024). Advanced detection tools in food fraud: A systematic review for holistic and rational detection method based on research and patents. *Food Chem.*, 138893. https://doi.org/10.1016/j.foodchem.2024.138893

[32] Hashemi-Nasab, F. S., Talebian, S., & Parastar, H. (2023). Multiple adulterants detection in turmeric powder using Vis-SWNIR hyperspectral imaging followed by multivariate curve resolution and classification techniques. *Microchem. J.*, *185*, 108203. https://doi.org/10.1016/j.microc.2022.108203

[33] Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2021). An evaluation of IR spectroscopy for authentication of adulterated turmeric powder using pattern recognition. *Food Chem.*, *364*, 130406. https://doi.org/10.1016/j.foodchem.2021.130406

[34] Khodabakhshian, R., Lavasani, H. S., & Weller, P. (2023). A methodological approach to preprocessing FTIR spectra of adulterated sesame oil. *Food Chem.*, *419*, 136055. https://doi.org/10.1016/j.foodchem.2023.136055

[35] da Silva Bruni, A. R., de Oliveira, V. M. A. T., Fernandez, A. S. T., Sakai, O. A., Março, P. H., & Valderrama, P. (2021). Attenuated total reflectance Fourier transform (ATR-FTIR) spectroscopy and chemometrics for organic cinnamon evaluation. *Food Chem.*, *365*, 130466. https://doi.org/10.1016/j.foodchem.2021.130466

[36] Wen, Y., Zhou, S., Wang, L., Li, Q., Gao, Y., & Yu, X. (2022). New method for the determination of the induction period of walnut oil by fourier transform infrared spectroscopy. *Food Anal. Methods*, 1-11. https://doi.org/10.1007/s12161-021-02170-6

[37] Fattahi, S. H., Kazemi, A., Khojastehnazhand, M., Roostaei, M., & Mahmoudi, A. (2024). The classification of Iranian wheat flour varieties using FT-MIR spectroscopy and chemometrics methods. *Expert Syst. Appl.*, *239*, 122175. https://doi.org/10.1016/j.eswa.2023.122175

[38] Jiang, D., Zhang, Y., Ge, Y., & Wang, K. (2023). Fusion Recalibration Method for Addressing Multiplicative and Additive Effects and Peak Shifts in Analytical Chemistry. *Chemosensors*, *11*(9), 472. https://doi.org/10.3390/chemosensors11090472

[39] Parsaei-Khomami, A., Badiei, A., Ghavami, Z. S., & Ghasemi, J. B. (2022). A new fluorescence probe for simultaneous determination of Fe2+ and Fe3+ by orthogonal signal correction-principal component regression. *J. Mol. Struct.*, *1252*, 131978. https://doi.org/10.1016/j.molstruc.2021.131978

[40] Ji, Q., Li, C., Fu, X., Liao, J., Hong, X., Yu, X., ... & Qiu, Y. (2023). Protected Geographical Indication Discrimination of Zhejiang and Non-Zhejiang Ophiopogonis japonicus by Near-Infrared (NIR) Spectroscopy Combined with Chemometrics: The Influence of Different Stoichiometric and Spectrogram Pretreatment Methods. *Molecules*, *28*(6), 2803. https://doi.org/10.3390/molecules28062803

[41] Reddy, P., Panozzo, J., Guthridge, K. M., Spangenberg, G. C., & Rochfort, S. J. (2023). Single seed near-infrared hyperspectral imaging for classification of perennial ryegrass seed. *Sensors*, *23*(4), 1820. https://doi.org/10.3390/s23041820

[42] Qian, S., Wang, Z., Chao, H., Sheng, X., Zhao, X., Lu, Z., ... & Chen, K. (2024). Development of near-infrared spectroscopy calibration model and monitoring software: For monitoring hexamethylenetetramine concentration in hexamethylenetetramine–acetic acid solution. *Infrared Phys. Technol.*, *139*, 105286. https://doi.org/10.1016/j.infrared.2024.105286

[43] Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2022). Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: The role of preprocessing methods in analysis. *Vib. Spectrosc.*, *120*, 103372. https://doi.org/10.1016/j.vibspec.2022.103372

[44] Yang, W., Xiong, Y., Xu, Z., Li, L., & Du, Y. (2022). Piecewise preprocessing of near-infrared spectra for improving prediction ability of a PLS model. *Infrared Phys. Technol.*, *126*, 104359. https://doi.org/10.1016/j.infrared.2022.104359

[45] Xie, L., Zhu, J., Wang, Y., Wang, N., Liu, F., Chen, Z., ... & Shen, X. (2022). Rapid and accurate determination of prohibited components in pesticides based on near infrared spectroscopy. *Infrared Phys. Technol.*, *121*, 104038. https://doi.org/10.1016/j.infrared.2022.104038

[46] de Lima, A. B. S., Batista, A. S., de Jesus, J. C., de Jesus Silva, J., de Araújo, A. C. M., & Santos, L. S. (2020). Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control*, *107*, 106802. https://doi.org/10.1016/j.foodcont.2019.106802

[47] Hu, J., Zhang, Y., Xiao, Z., & Wang, X. (2018). Preparation and properties of cinnamon-thyme-ginger composite essential oil nanocapsules. *Ind. Crops Prod.*, *122*, 85-92. https://doi.org/10.1016/j.indcrop.2018.05.058

[48] Wu, Y., Xian, Y., Guo, X., Chen, L., Zhao, X., Wang, B., & Wang, L. (2018). Development and validation of a screening and quantification method for simultaneous determination of seven fluorescent whitening agents in commercial flour using UPLC–MS/MS. *Food chem.*, *243*, 162-167. https://doi.org/10.1016/j.foodchem.2017.09.110

[49] Özçimen, D., & Ersoy-Meriçboyu, A. (2010). Adsorption of copper (II) ions onto hazelnut shell and apricot stone activated carbons. *Adsorp. Sci. Technol.*, *28*(4), 327-340. https://doi.org/10.1260/0263-6174.28.4.327

[50] Dave, G., & Modi, H. (2018). FT-IR method for estimation of phytic acid content during bread-making process. *J. Food Meas. Charact.*, *12*(3), 2202-2208. https://doi.org/10.1007/s11694-018-9836-y

[51] Özçimen, D., & Ersoy-Meriçboyu, A. (2010). Characterization of biochar and bio-oil samples obtained from carbonization of various biomass materials. *Renew. Energy*, *35*(6), 1319-1324. https://doi.org/10.1016/j.renene.2009.11.042

[52] Moghimi, A., Aghkhani, M. H., Sazgarnia, A., & Sarmad, M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosyst. Eng.*, 106(3), 295-302. https://doi.org/10.1016/j.biosystemseng.2010.04.002

[53] Khodabakhshian, R., Emadi, B., Khojastehpour, M., & Golzarian, M. R. (2016). Visible-NIR infrared spectroscopy for pomegranate fruit quality assessment: chemometrics and common preprocessing methods. *Ann. Food Sci. Technol.*, 17(1).

[54] Tan, J., Liu, J. Y., Su, H., Yang, X. H., & Li, H. F. (2024). Detection of adulteration of cumin powder by front-face synchronous fluorescence spectroscopy: The influence of the natural variation of adulterants. *Food Control*, 158, 110228. https://doi.org/10.1016/j.foodcont.2023.110228

[55] Liu, Y., Finley, J., Betz, J. M., & Brown, P. N. (2018). FT-NIR characterization with chemometric analyses to differentiate goldenseal from common adulterants. *Fitoterapia*, 127, 81-88. https://doi.org/10.1016/j.fitote.2018.02.006

[56] Wu, S., Wang, L., Zhou, G., Liu, C., Ji, Z., Li, Z., & Li, W. (2023). Strategies for the content determination of capsaicin and the identification of adulterated pepper powder

using a hand-held near-infrared spectrometer. *Food Res. Int.*, 163, 112192. https://doi.org/10.1016/j.foodres.2022.112192

[57] Masoudi, M., & Khodabakhshian, R. (2025). Genetic algorithm-optimized PLS for detecting adulteration in cinnamon powder via FT-IR spectroscopy. *Expert Syst. Appl.*, 128522. https://doi.org/10.1016/j.eswa.2025.128522