Research Square

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

Raman-Chemometric Framework for Rapid Authentication of Edible Oils in Processed Foods Using a Validated One Step Sampling Technique

Amrita Shaw
Sri Sathya Sai Institute of Higher Learning
Chandrasekar SUBRAMANI NARAYANA
Aix-Marseille University
Sai Muthukumar V
Sri Sathya Sai Institute of Higher Learning
Deepak L N Kallepalli
CogniEvolve AI Inc
Bhanu Prakash Rachaiah
Science4u Analytics And Research Solutions Private Limited
Jhinuk Gupta
jhinukgupta@sssihl.edu.in

Sri Sathya Sai Institute of Higher Learning

Research Article

Keywords: Raman spectroscopy, Edible oil authentication, Ratiometric analysis, PCA, ANOVA, MANOVA

Posted Date: February 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8872116/v1

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

Additional Declarations: No competing interests reported.

Abstract

Authentication of cooking oils in processed foods is essential for food safety, quality control, and regulatory compliance, yet routine analysis remains constrained by solvent-intensive extraction procedures and limited applicability to complex food matrices. This study presents a proof-of-concept Raman–chemometric framework for direct oil authentication in processed foods using minimal, solvent-free sample handling. Potato chips were selected as a representative fried food matrix, and five commonly used edible oils (sunflower, soybean, groundnut, palm, and vanaspati) were analyzed in both pure form and corresponding chip matrices to enable systematic cross-matrix evaluation. A one-step tissue paper blotting method was employed for rapid oil recovery, followed by Raman spectroscopic analysis. Fatty-acid-associated Raman bands were identified and systematically combined into chemically interpretable inter-peak intensity ratios reflecting variations in saturation and ester content. A statistically grounded two-stage marker selection workflow integrating Random Forest importance ranking with non-parametric Kruskal–Wallis testing was applied, followed by using one-way analysis of variance and coefficient of variation analysis. Five robust ratiometric markers ($I_{1652}/_{1742}$, $I_{1652}/_{1434}$, $I_{1742}/_{1259}$, $I_{1652}/_{1259}$, and $I_{1296}/_{1434}$) showed strong association with saturated versus unsaturated fatty acid profiles. Multivariate analysis based on these markers revealed pronounced separation among oil types (91% explained variance; F-values up to 22,629; $p < 0.001$) and statistically significant discrimination within chip matrices (77% explained variance), despite attenuation effects from the food matrix. Multivariate analysis of variance confirmed robust separation (Wilks' $\Lambda = 0.0307$; $p < 0.0001$). Overall, this framework establishes an interpretable, extraction-free, and scalable foundation for high-throughput oil authenticity screening in processed foods.

Introduction

Identifying the type of oil used in processed foods is crucial due to its significant impact on consumer health (Foster et al. 2009; Moore et al. 2012). A common form of adulteration involves substituting listed oils with cheaper analogues, locally available alternatives, or repeatedly heated oils, compromising both food quality and safety (Rifna et al. 2022; Haji et al. 2023). Some of these substitutes, such as argemone oil, contain toxins and are banned globally by regulatory agencies (Tan et al. 2021; Begum and Jain 2024). Even legally permitted oils like palm oil and vanaspati pose health risks and are unsuitable for individuals with cardiovascular diseases and various health issues, while soybean and groundnut oils can trigger allergic reactions (Johnson and Saikia 2009; Dorni et al. 2018). Although oil adulteration is a global issue, it becomes a formidable challenge in developing countries like India (Rana et al. 2014). Regulatory enforcement mainly targets the organized food sector, leaving much of the unregulated market vulnerable. As a result, incidents of such adulteration have increased significantly in recent years (Program 2024). Moreover, recent investigations suggest that economically motivated adulteration is no longer confined to unregistered businesses. A nationwide audit by the Food Safety and Standards Authority of India (FSSAI) in 2022, seized thousands of liters of edible oils for suspected mislabeling, fake branding and exceeding levels of trans-fatty acids, highlighting the urgent need for stricter monitoring and legal intervention (Sharma R, 2022; Desk, 2024). Typically, Chromatographic methods like

High-Performance Liquid Chromatography (HPLC) (Cercaci et al. 2003; Navratilova et al. 2022; Khursheed 2024) and Gas Chromatography (GC) (Aparicio and Aparicio-Ruíz 2000; Su et al. 2011; Ün and OK 2018), along with Nuclear Magnetic Resonance (NMR) (Ün and OK 2018; Gu et al. 2021), are used for oil authentication. While these methods provide high accuracy, their complex sample processing protocols, lengthy analysis procedures and high operational costs make them unsuitable for rapid, real time monitoring (Elumalai and Natarajan 2025). Thus, a high-throughput screening technique with minimal sample processing is essential and a desirable alternative.

Vibrational spectroscopic techniques have emerged as a superior alternative to chromatographic methods, overcoming their limitations and enabling real-time monitoring of edible oils (Yadav 2018; Hu et al. 2019; Yao et al. 2020; Forooghi 2023). Among these, Raman spectroscopy stands out for its ability to provide detailed molecular fingerprints *via* Raman scattering (Mendes et al. 2015; Yang et al. 2019; Forooghi 2023). It offers a promising alternative because it directly probes molecular vibrations associated with lipid structure, including saturation, unsaturation and ester functionality (Baeten et al. 1998; Huang et al. 2016; Kwofie et al. 2020; Wang et al. 2025). Characteristic Raman bands at critical wavenumbers like 1650–1660 cm^{-1} (C = C), 1740–1750 cm^{-1} (ester carbonyl), and 1430–1440 cm^{-1} (CH_2 scissoring) have been widely used to differentiate oil types and monitor oxidation (Huang et al. 2022; Wang et al. 2025). Attempts have been made to probe the quality of waste cooking oils (Huang et al. 2016) and the quantification of trans-fat in edible oils (Gong et al. 2019) using Raman spectroscopy. Therefore, the current trends in food quality monitoring clearly indicate the adaption of molecular spectroscopic techniques like Raman Spectrometry for the development of rapid quality monitoring tools.

Ratiometric analysis in spectroscopy involves calculating the ratio of intensities between two characteristic peaks rather than using absolute intensity values, that enhances measurement sensitivity and selectivity (Wu et al. 2024). Ratiometric interpretation of Raman bands is increasingly common due to its robustness against intensity fluctuations, laser defocus, and sample variability (Jamieson et al. 2018). However, the literature lacks ratiometric markers that are systematically derived, reproducible, statistically validated, and correlated with chemical composition. Stringent selection of these ratiometric markers helps to reduce the high dimensionality of spectral data and improve statistical performance (Jamieson et al. 2018). Unfortunately, most existing studies rely only on visual or empirical selection of ratios without rigorous statistical validation. Therefore, it is important to develop a systematic framework for identifying and validating ratiometric markers through quantitative correlation analysis and multivariate statistical testing.

The application of multivariate analysis to spectral data highlights the growing importance of data analysis in food quality monitoring (Vaskova and Buckova 2018). Chemometric analysis combined with spectrometry has been widely recognized for its effectiveness, particularly in oil quality assessment (Gómez-Caravaca et al. 2016). Integrating vibrational spectroscopic techniques with chemometric methods is expected to further enhance the evaluation of edible oil quality (Liu et al. 2020). Rigorous

application of chemometrics for oil authentication in complex food matrices is gap in the scientific domain which demands more efforts.

Quality monitoring in processed food is more challenging than quality monitoring in ingredients due to matrix interference. One such example is interference from starch and seasonings while screening for the oil in potato chips (food matrix). Most of the food quality monitoring studies deal with ingredients testing rather than testing the final food product (Inteaz 1989) as testing of such end products involves elaborate sampling, extraction, isolation, and derivatization. Studies have shown that crucial components, such as fats and oils, which are extensively used in food preparation still need to be isolated from matrices (Bansal et al. 2010) *via* a long, tedious process before impending analysis (Kim et al. 2018; Yao et al. 2020; Kaimal et al. 2021). As per example, the traditional method for oil analysis in fried food matrix involves mechanical pressing to extract oil followed by debris removal, filtration and additional purification steps (Aykas and Rodriguez-Saona 2016). This extended processing time underscores the need for developing a rapid, high throughput, inexpensive and efficient sampling protocol.

As seen above, most of the reported studies have explored oil type analysis using Raman-PCA (Castro et al. 2022) involving tedious traditional multi-step processes. Little efforts have been made on developing techniques with minimal sample processing, directly from the food matrix, which can actually make the oil quality monitoring highly reliable and reproducible. From a food safety perspective, a framework capable of analyzing oil types from food matrices is inevitably crucial.

In this project, we have developed Raman-PCA frameworks to identify oil types in processed foods (Potato chips as the food matrix) with minimal sample processing. A novel, non-invasive tissue paper-based oil sampling technique for food matrix was developed and validated in the study. Potato chips fried in five different edible oils, including sunflower oil, vanaspati oil, palm oil, groundnut oil, and soybean oil, were explored to design the framework. While these oils are not always classified as adulterants, they can pose health risks to particular individuals with soy or peanut allergies or cardiovascular conditions when added undeclared. For developing the framework, spectral analysis identified fatty acid-associated Raman bands from which inter-peak intensity ratios were systematically constructed. A two-stage selection strategy combining Random Forest marker importance ranking and Kruskal-Wallis statistical testing identified the most discriminatory intensity ratios, which were further validated through one-way ANOVA and coefficient of variance (CV) analysis for marker reproducibility and robustness. The selected ratiometric markers were subjected to multivariate analysis, with principal component analysis (PCA) performed to assess class separability, followed by formal statistical validation using MANOVA to confirm the robustness of multivariate separation beyond visual inspection.

Materials and methods

Procurement and preparation of samples

Five commercially available packaged edible oils (Sunflower-SO, Soybean-SOYO, Groundnut-GNO, Palm-PO, and Vanaspati-VO) with FSSAI certification were procured from retail outlets in Anantapur district, Andhra Pradesh, India. Figure 1 depicts the steps employed for SO, which was systematically replicated across all oil types. The samples, along with their codes, are provided in Table 1. To mimic actual practices followed by small scale potato chips manufacturing units, thin slices of potatoes were fried in each oil over 9 successive batches, with systematic collection of both fried chips and oil samples.

Table 1
List of samples prepared for analysis

	Sample Codes									
Heating Stages	**Sunflower Oil (SO)**		**Palm Oil (PO)**		**Vanaspati (VO)**		**Soya Bean Oil (SOYO)**		**Groundnut Oil (GNO)**	
	Oil	**Food matrix (Chips)**	**Oil**	**Food matrix (Chips)**	**Oil**	**Food matrix (Chips)**	**Oil**	**Food matrix (Chips)**	**Oil**	**Food matrix (Chips)**
0	*SO 0*	-	*PO 0*	-	*VO 0*	-	*SOYO 0*	-	*GNO 0*	-
1	*SO 1*	*SO C 1*	*PO 1*	*PO C 1*	*VO 1*	*VO C 1*	*SOYO 1*	*SOYO C 1*	*GNO 1*	*GNO C 1*
2	*SO 2*	*SO C 2*	*PO 2*	*PO C 2*	*VO 2*	*VO C 2*	*SOYO 2*	*SOYO C 2*	*GNO 2*	*GNO C 2*
3	*SO 3*	*SO C 3*	*PO 3*	*PO C 3*	*VO 3*	*VO C 3*	*SOYO 3*	*SOYO C 3*	*GNO 3*	*GNO C 3*
4	*SO 4*	*SO C 4*	*PO 4*	*PO C 4*	*VO 4*	*VO C 4*	*SOYO 4*	*SOYO C 4*	*GNO 4*	*GNO C 4*
5	*SO 5*	*SO C 5*	*PO 5*	*PO C 5*	*VO 5*	*VO C 5*	*SOYO 5*	*SOYO C 5*	*GNO 5*	*GNO C 5*
6	*SO 6*	*SO C 6*	*PO 6*	*PO C 6*	*VO 6*	*VO C 6*	*SOYO 6*	*SOYO C 6*	*GNO 6*	*GNO C 6*
7	*SO 7*	*SO C 7*	*PO 7*	*PO C 7*	*VO 7*	*VO C 7*	*SOYO 7*	*SOYO C 7*	*GNO 7*	*GNO C 7*
8	*SO 8*	*SO C 8*	*PO 8*	*PO C 8*	*VO 8*	*VO C 8*	*SOYO 8*	*SOYO C 8*	*GNO 8*	*GNO C 8*
9	*SO 9*	*SO C 9*	*PO 9*	*PO C 9*	*VO 9*	*VO C 9*	*SOYO 9*	*SOYO C 9*	*GNO 9*	*GNO C 9*

Tissue-paper based oil sampling for food matrix (potato chips)

A novel, simplistic but promising tissue paper–based sampling technique tailored for food matrices was developed and evaluated in this study, with potato chips used as a representative model system as a food matrix. The proposed extraction approach is intended as a rapid and facile screening tool for oil authentication. Single-ply, laboratory-grade, non-fragrant tissue paper (Kimwipes®, Kimberly-Clark) was cut into uniform sections and placed between two fried potato chips with consistent pressure to facilitate oil absorption through capillary action. This method enabled efficient transfer of surface oil into the tissue paper *via* blotting. These tissue papers which assimilated oil from the fried chips were subsequently analyzed directly by Raman spectroscopy without any further processing. A pictorial depiction of the developed extraction technique is presented in Fig. 2.

Data acquisition

Raman spectroscopic measurements of all oil and chip samples were performed using a dispersive Raman microscope system (DXR, Thermo Fisher Scientific) equipped with a 785 nm excitation laser. After preliminary spectral acquisition survey, the instrumental parameters were optimized as follows: laser power, 20 mW; exposure time, 30 s; number of exposures, 2; confocal aperture, 25 μm slit; diffraction grating, 400 lines mm^{-1}; spectral resolution, 4 cm^{-1}. Raman spectra were acquired over the typical molecular finger printing range of 200–3000 cm^{-1}. To ensure statistical reliability and account for sample heterogeneity, 20 different spectra were collected by focusing on different locations on each sample. This spatially resolved acquisition strategy enhances measurement robustness and reduces local sampling bias. The resulting spectra were treated as technical replicates representative of each experimental condition, and all subsequent analyses were performed accordingly under controlled laboratory settings. The complete dataset comprised 1000 Raman spectra from oil samples (200 spectra per oil type across 10 heating stages, including pristine) and 900 spectra from chip samples (180 spectra per oil type across 9 frying stages, starting from 1st frying as chips cannot exist in pristine form), totaling 1900 spectra across five oil types.

Data preprocessing

The spectral dataset was subjected to a systematic preprocessing pipeline comprising three sequential steps (Fig. 3). (i) Baseline correction was performed using the asymmetric least squares (ALS) (Korepanov 2020) algorithm with optimized parameters: asymmetry parameter (p) = 0.01, smoothing parameter (λ) = 10^5, and 10 iterations. (ii) Spectral smoothing was applied using the Savitzky–Golay filter (Barton et al. 2018) with an 11-point window and third-order polynomial fitting to reduce noise while preserving spectral features. (iii) Finally, the entire dataset was normalized to the 2720 cm^{-1} band, which corresponds to C–H stretching vibrations found commonly in all fatty acid chains across all samples. Specifically, 2720 cm^{-1} was chosen as the wavenumber for normalization as it will not hamper the relative intensities of other functionally important peaks (Baeten et al. 1998; Kwofie et al. 2020).

Figure 3 Schematic flow for spectral data processing used in the proposed framework

Construction of ratiometric markers

The key component of identifying and differentiating various oil in the proposed framework is attempted by Ratiometric analysis. For ratiometric analysis, spectral peaks corresponding to the following major vibrational modes of fatty acids and esters were systematically identified; for example, Carbonyl stretching of carboxyl groups (–COOH) vibration at 1742 cm^{-1}, methylene bending (δ-CH_2) at 1434 cm^{-1}, C-C skeletal vibrations at 1123 cm^{-1} and 1296 cm^{-1}, C-H stretching at 2720 cm^{-1}, and -CH_2 stretching at 3010 cm^{-1}. Unsaturated fatty acids exhibited additional characteristic bands at 1259 cm^{-1} (= C–H bending vibrations) and 1652 cm^{-1} (C = C stretching vibrations), with peak intensities directly proportional to the degree of unsaturation. All possible pairwise intensity ratios (Ratiometric markers) were computed using these key spectral bands. Optimal ratiometric markers were identified using statistical analysis and validation, as described below.

Statistical selection, marker validation, and multivariate modeling

Statistical identification of ratio markers

Optimization and validation of the ratiometric markers were carried out following the analytical workflow shown in Fig. 4. All analyses were performed using open source Python package 3.12.9. After constructing the comprehensive set of ratiometric markers, a two-stage selection strategy were employed to identify the most discriminatory markers. First, a Random Forest (RF) classifier was applied to rank all ratiometric markers based on their marker importance scores. Random Forest creates an ensemble of decision trees trained on randomized data subsets and assigns feature importance based on their contribution to classification accuracy across the forest (Menze et al. 2009; Shehata et al. 2024). The RF model provided an independent, model-based assessment of each marker. It should be noted that feature importance rankings obtained from Random Forest models may be influenced by feature correlations and dataset characteristics and should therefore be interpreted in conjunction with complementary statistical analyses. The markers exhibiting the highest importance scores were chosen as the most effective markers and selected for further statistical validation using the Kruskal-Wallis test (Kaur et al. 2016; Pereira et al. 2021), a non-parametric method for assessing group-wise differences. The ratios demonstrating statistically significant variations across oil classes ($p < 0.05$) in the Kruskal-Wallis analysis were selected for further screening and validation through one-way analysis of variance (ANOVA) (Davari et al. 2024) to confirm discriminatory power. This validation step was further aided by coefficient of variance (CV) analysis to assess marker reproducibility and robustness. Ultimately, Principal Component Analysis (PCA) (Jolliffe and Cadima 2016) was performed on the final set of the ratiometric markers thus obtained to assess their effectiveness in clustering different oil classes in multivariate space. The variance explained by each principal component were calculated, and score plots were generated to visualize clustering. To check whether the clustering observed in PCA is statistically significant, MANOVA (Zhu et al. 2022) was performed on the ratios. Four complementary

test statistics were computed: Wilks' lambda (Λ), Pillai's trace (V), Hotelling–Lawley trace (T), and Roy's greatest root (Θ) (Alkarkhi and Alqaraghuli 2019) to confirm the robustness of multivariate separation. The marker selection and validation procedures were designed to be systematic and statistically grounded within the available dataset and within the scope of the present proof-of-concept study. Nevertheless, further evaluation using independent validation datasets will be valuable for confirming long-term marker stability and generalizability under broader application scenarios.

Results and discussion

Interpretation of Raman spectral peaks related to fatty acids and esters

Spectral analysis was performed on the baseline-corrected, smoothened and normalized Raman spectra of five oil types and their corresponding chip samples (Fig. 5). Both oils and chips samples were found to contain a few strong peaks at 1123 cm^{-1} [(C–C) stretching of the $(CH2)_n$ group], 1259 cm^{-1} (= C–H bending), 1296 cm^{-1} [(= C–H) deformation of cis(R–HC = CH–R)], 1434 cm^{-1} (methylene bending (δ-CH_2)), 1652 cm^{-1} [(C = C) of cis(R-HC = CH-R)], 1742 cm^{-1} [carbonyl stretching of ester carboxyl groups (-COOH)], 2720 cm^{-1} (C–H stretching) and 3010 cm^{-1} (methyl group stretching). As mentioned in materials and methods section, the chips spectra were collected after blotting the oil from the chips using tissue paper. Therefore, presence of similar signature peaks in chips and oils suggests that the dominant molecular signatures of the oils are largely preserved under the proposed sampling conditions, despite effects from both the food matrix and the tissue paper (cellulose) matrix. It suggests tissue paper based sampling to be an effective alternative for tedious extraction based sampling which are most commonly used during food ingredient analysis. Overall, the preliminary spectral analysis of the data acquired from the oil laden tissue paper confirmed that Raman spectroscopy captured chemically meaningful lipid signatures in both oil and chips, paving way for a robust physicochemical foundation for subsequent ratiometric analysis and multivariate discrimination.

After assigning the peaks to specific vibrational groups, the peak intensities were systematically evaluated to determine their correlation with variations in saturated and unsaturated fatty acid composition across the studied oils. CV analysis was done on the intensity of the above mentioned peaks. Unfortunately, CV obtained, as low as 0.2–0.3%, indicated their incapability to be used as markers. In order to find suitable markers, instead of relying on the individual peak intensities, intensity ratios of two chemically meaningful peaks (ratiometric markers) were explored.

Figure 5 Baseline corrected, smoothened and normalized Raman spectra collected from oils and chips

Identifying ratiometric markers suitable for oil type classification

To identify the suitable ratiometric markers from the extensive pool of all potential markers, three crucial criteria were evaluated; chemical relevance, model-based marker-ranking approach and statistical significance. Only those ratios were retained which (i) were constructed from chemically interpretable Raman bands, (ii) showed high significance in Random forest markers importance and (iii) statistically significant as per Kruskal–Wallis analysis. Ratios failing to meet these criteria were excluded.

Random Forest (RF) classifier was employed on the chemically relevant markers as an independent, model-based marker-ranking approach. According to the RF model (Table 2), five ratiometric markers $I_{1652}/_{1742}$, $I_{1652}/_{1434}$, $I_{1742}/_{1259}$, $I_{1652}/_{1259}$, and $I_{1296}/_{1434}$ were identified as the most informative. Their importance values ranged from ~ 0.300 to 0.116, while all other ratios showed much lower contributions. The statistical significance ($p < 0.005$) of these highly ranked ratios were further assessed through Kruskal-Wallis group-wise testing, which revealed the same ratios to consistently show highly significant differences ($p << 10^{-6}$) across oil classes. This systematic selection resulted in a compact, chemically significant ratiometric marker set, which formed the basis for all subsequent multivariate classification.

Table 2
Statistical Significance by Kruskal–Wallis analysis and Random Forest Importance of Ratiometric markers in Oils and Chips

Ratio markers	Kruskal–Wallis analysis		RF importance	
	Oils	Chips	Oils	Chips
	p value	p value		
$I_{1652/1434}$	$p < 1 \times 10^{-22}$	$p < 1 \times 10^{-15}$	0.300	0.417
$I_{1296/1434}$	$p < 1 \times 10^{-6}$	$p < 1 \times 10^{-9}$	0.116	0.139
$I_{1652/1742}$	$p < 1 \times 10^{-10}$	$p < 1 \times 10^{-8}$	0.208	0.138
$I_{1742/1259}$	$p < 1 \times 10^{16}$	$p < 1 \times 10^{-13}$	0.212	0.131
$I_{1652/1259}$	$p < 1 \times 10^{-12}$	$p < 1 \times 10^{-16}$	0.123	0.103

Statistical screening and reproducibility validation of ratiometric markers

A cross-validated marker selection strategy was employed to systematically validate the ratiometric markers identified in previous section. Based on the ANOVA results (F-statistic) and CV values (Table 3), all five ratiometric markers: I_{1652}/I_{1434}, I_{1742}/I_{1259}, and I_{1652}/I_{1742}, $I_{1652/1259}$ and $I_{1296/1434}$ were identified as suitable for optimal oil type classification. For oil samples, these ratios exhibited strong statistical separation across oil types (F: 22629.79, 1267.22, and 11697.20 and CV: 43.77%, 43.86% and 38.23%, respectively). However, for chips samples, the F-statistics and CV values turned out to be lower (F: 2568.23, 331.46 and 282.45; CV: 43.64%, 38.82% and 33.09%). This reduction is most likely attributable

to the increased heterogeneity and compositional complexity of the chips (food matrix). Overall, the results from the ratio markers analysis demonstrated that oil type classification can be potentially achieved using this identified and statistically significant set of five robust, chemically interpretable ratio markers and suitable for chemometric analysis for the final framework development.

Table 3
Statistical Significance of selective ratiometric markers assessed by ANOVA (F_statistic) and Coefficient of Variance (CV) values for oils and chips.

Ratiometric Markers	Oil Samples		Chips Samples	
	F_statistic	CV Values (%)	F_statistic	CV Values (%)
$I_{1652/1434}$	22629.79	43.77	2568.23	43.64
$I_{1652/1742}$	11697.20	43.86	331.46	38.82
$I_{1742/1259}$	1267.22	38.23	282.45	33.09
$I_{1652/1259}$	666.20	34.52	55.76	10.58
$I_{1296/1434}$	637.99	33.28	189.10	20.48

Multivariate classification and statistical validation using optimized ratiometric markers

The classification efficiency of the selected ratio markers was further evaluated using multivariate analysis to assess whether oil types can form distinct clustering corresponding to different types of oils. Multivariate separation was explored using Principal Component Analysis (PCA), followed by formal statistical validation using MANOVA.

Oil data analysis

As seen in Fig. 6, PCA of oil samples based on the selected ratiometric markers revealed discernible clustering, with the first two principal components explaining 76% and 15% of the total variance, respectively (91% cumulative variance). The grouping of oil types within the principal component space is evident in the corresponding score plot. Quantitative evaluation yielded a silhouette coefficient of 0.36 (Table 4), indicating weak-to-moderate cluster separation. In general silhouette coefficients range from −1 to 0 to 1, with values close to zero reflecting overlapping clusters and values approaching unity indicating well-separated groups. In addition, the ratio of between-class to within-class scatter (B/W = 3.34) demonstrated that inter-class variability substantially exceeded intra-class variability in the PCA space. A permutation test on the B/W statistic (p_perm = 0.005) further confirmed that the observed clustering is unlikely to arise from random variation. Together, these complementary metrics indicate statistically meaningful separation between oil classes.

Chips data analysis

For chips, PCA revealed a reduced but still interpretable multivariate separation, with the first two principal components accounting for 57% and 20% of the variance, respectively (77% cumulative variance). Although cluster separation was less pronounced than in oils, quantitative measures still indicated non-random structure (silhouette = 0.12; B/W = 0.88; p_perm = 0.005). This reduced separation is expected to result from food-matrix interference, yet oil-dependent trends remain detectable in multivariate space which is valuable from screening purpose.

While PCA and associated clustering metrics provide geometric and distributional measures of class separability, they do not constitute formal hypothesis tests. Therefore, multivariate analysis of variance (MANOVA) was employed to rigorously assess whether the observed group separation is statistically significant across all ratiometric variables simultaneously. Results obtained from the MANOVA analysis are presented in Table 4.

Table 4
Evaluation metrics by PCA and MANOVA

PCA metrics						
Oils				**Chips**		
Features			**Values**	**Features**		**Values**
Silhouette			0.361	**Silhouette**		0.116
Between_within_ratio			3.33	**Between_within_ratio**		0.87
p_perm			0.005	**p_perm**		0.005
MANOVA metrics						
Oils				**Chips**		
Features	**Values**	**F_value**	**p**	**Values**	**F_value**	**p**
Wilks' lambda	0.0008	984.83	< 0.001	0.0307	221.75	< 0.001
Pillai's trace	2.2241	207.28	< 0.001	1.5068	89.94	< 0.001
Hotelling-Lawley trace	184.68	7609.35	< 0.001	15.4828	573.41	< 0.001
Roy's greatest root	180.84	29929.62	< 0.001	14.4794	2155.02	< 0.001

For oils, Wilks’ lambda indicated strong multivariate separation among oil classes ($\Lambda = 0.0008$, $F = 984.83$, $p < 0.001$). Wilks’ lambda ranges from 0 to 1, with values close to unity indicating no group separation and values approaching zero reflecting near-complete discrimination. The observed value therefore indicates that only a negligible fraction of the total multivariate variance remains unexplained by oil type. Consistent results were obtained from Pillai’s trace ($V = 2.2241$, $F = 207.28$, $p < 0.001$), which ranges from 0 to minimum $(p, k - 1)$, where $(p = 5)$ is the number of variables and $(k = 5)$ the number of groups, and quantifies the cumulative proportion of variance explained across multiple discriminant

dimensions. In the present study, the observed Λ and V values correspond to more than half of the theoretical maximum, indicating strong multivariate separation. Similarly, the Hotelling–Lawley trace ($T = 184.68$, $F = 7609.35$, $p < 0.001$), which reflects the overall signal-to-noise ratio, and Roy's greatest root ($\Theta = 180.84$, $F = 29929.62$, $p < 0.001$), which measures separation along the strongest discriminant axis, both indicate highly pronounced group differences. For chips, despite reduced discriminatory power due to food-matrix interference, significant multivariate separation was likewise observed ($\Lambda = 0.0307$, $V = 1.5068$, $T = 15.48$, $\Theta = 14.48$; all $p < 0.0001$). Although these values are smaller than those obtained for oils, they remain well within ranges associated with statistically meaningful discrimination. In particular, Wilks' lambda remains close to zero and Pillai's trace indicates substantial explained variance, demonstrating that oil-dependent spectral signatures are retained despite interference from non-lipid components.

Taken together, the consistent significance across all four complementary statistics confirms robust and method-independent group separation. The close agreement between exploratory chemometric visualization (PCA) and formal multivariate inference (MANOVA) demonstrates that oil-type classification is driven by genuine multivariate oil-type-dependent compositional differences encoded in the ratiometric markers, establishing the statistical robustness and chemical relevance.

Conclusion

This study establishes a proof-of-concept Raman–chemometric framework for rapid authentication of edible oils in processed foods using chemically interpretable ratiometric markers. Key markers were systematically identified through a two-stage selection strategy combining Random Forest importance ranking and Kruskal–Wallis testing across all inter-peak combinations. Subsequent ANOVA and coefficient of variance (CV) analyses validated a final set of five robust markers exhibiting strong associations with saturated and unsaturated fatty acid content. PCA based chemometric analysis using these markers revealed pronounced clustering for oils (91% cumulative variance; silhouette = 0.36; B/W = 3.34) and moderate yet statistically significant separation for chips (77% cumulative variance; silhouette = 0.12; B/W = 0.88), with permutation testing confirming non-random structure ($p = 0.005$). Multivariate analysis of variance further validated group separation for both oils (Wilks' $\Lambda = 0.0008$, $p < 0.001$) and chips (Wilks' $\Lambda = 0.0307$, $p < 0.0001$), with consistent significance across complementary test statistics, demonstrating that the observed discrimination reflects genuine oil-type-dependent multivariate compositional differences rather than sampling variability.

In parallel, this study introduces a novel tissue paper–based minimal sampling strategy tailored for complex food matrices like fried chips, providing a practical and environmentally sustainable alternative to conventional solvent-intensive extraction protocols. This approach facilitates rapid, low-cost sample preparation while preserving analytical sensitivity in multi-component systems.

Overall, the proof-of-concept framework provides a foundation for automated and AI-assisted oil authentication workflow in processed foods. While the present study was conducted under controlled

laboratory conditions using a limited number of oil types and a single food matrix, the robustness and chemical interpretability of the proposed framework provide a strong basis for extension to diverse industrial processing conditions, oil blends, varied adulteration scenarios, and commercial products. Future work will focus on expanding the methodology to a broader range of oil varieties and diverse food matrices. In addition, integration with advanced machine learning models will be explored to further enhance classification accuracy and enable scalable, real-time food authenticity monitoring and quality control.

Declarations

AcknowledgementsThe authors would like to express their sincere gratitude to the Founder Chancellor and the Institute Management of Sri Sathya Sai Institute of Higher Learning, Anantapur, Andhra Pradesh, India, for their support in facilitating this research. The authors also acknowledge the Central Research Instrumentation Facility (CRIF) for providing the essential resources and facilities necessary for the instrumental analysis. Additionally, the authors extend their thanks to the Department of Food and Nutritional Sciences for their valuable resources and support throughout the course of this work.

Author contributionJ.G. conceptualization, supervision, formal analysis, review & editing, approval of final version of manuscript. A.S. conceptualization, methodology, investigation, software, data curation, formal analysis, visualization, writing – original draft, approval of final version of manuscript. C.S.N. software, data curation, formal analysis, visualization, review & editing, approval of final version of manuscript. S.M.V. supervision, review & editing, approval of final version of manuscript. D.L.N.K. software, formal analysis, review & editing, approval of final version of manuscript. B.P.R. data curation, formal analysis, approval of final version of manuscript.

Data Availability The datasets and analysis scripts generated and used in the present study are available from the corresponding author upon reasonable request and will be shared on a case-by-case basis, subject to ongoing related research activities and institutional policies.

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding The authors did not receive support from any organization for the submitted work.

References

1. Alkarkhi, A. F. M., Alqaraghuli, W. A. A. (2019). Comparing several population means. Easy Statistics for Food Science with R. Academic Press, pp 81-105. https://doi.org/10.1016/B978-0-12-814262-2.00006-6
2. Aparicio R, Aparicio-Ruíz R (2000) Authentication of vegetable oils by chromatographic techniques. Journal of Chromatography A 881:93–104. https://doi.org/10.1016/S0021-9673(00)00355-1

3. Aykas DP, Rodriguez-Saona LE (2016) Assessing potato chip oil quality using a portable infrared spectrometer combined with pattern recognition analysis. Anal Methods 8:731–741. https://doi.org/10.1039/C5AY02387D
4. Baeten V, Hourant P, Morales MT, Aparicio R (1998) Oil and Fat Classification by FT-Raman Spectroscopy. J Agric Food Chem 46:2638–2646. https://doi.org/10.1021/jf9707851
5. Bansal G, Zhou W, Barlow PJ, et al (2010) Performance of palm olein in repeated deep frying and controlled heating processes. Food Chemistry 121:338–347. https://doi.org/10.1016/j.foodchem.2009.12.034
6. Barton SJ, Ward TE, Hennelly BM (2018) Algorithm for optimal denoising of Raman spectra. Anal Methods 10:3759–3769. https://doi.org/10.1039/C8AY01089G
7. Begum A, Jain BP (2024) Adulteration in edible oil (mustard oil) and ghee; detection and their effects on human health. Int J Biochem Mol Biol 15:141–148. https://doi.org/10.62347/VVCI2092
8. Comparing Several Population Means. https://www.sciencedirect.com/science/chapter/monograph/abs/pii/B9780128142622000066. Accessed 4 Feb 2026
9. Castro RC, Ribeiro DSM, Santos JLM, Páscoa RNMJ (2022) The use of in-situ Raman spectroscopy to monitor at real time the quality of different types of edible oils under frying conditions. Food Control 136:108879. https://doi.org/10.1016/j.foodcont.2022.108879
10. Cercaci L, Rodriguezestrada M, Lercker G (2003) Solid-phase extraction–thin-layer chromatography–gas chromatography method for the detection of hazelnut oil in olive oils by determination of esterified sterols. Journal of Chromatography A 985:211–220. https://doi.org/10.1016/S0021-9673(02)01397-3
11. Davari M, Bahreini M, Sabzevari Z (2024) Developing a non-destructive method for the detection of egg quality and freshness using micro-Raman spectroscopy. Applied Food Research 4:100453. https://doi.org/10.1016/j.afres.2024.100453
12. Desk, T. L. (2024, June 6). How to identify fake cooking oil? (FSSAI recommended tips). The Times of India. https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/how-to-identify-fake-cooking-oil-fssai-recommended-tips/articleshow/110694766.cms
13. Dorni C, Sharma P, Saikia G, Longvah T (2018) Fatty acid profile of edible oils and fats consumed in India. Food Chemistry 238:9–15. https://doi.org/10.1016/j.foodchem.2017.05.072
14. Elumalai A, Natarajan V (2025) Advancements in analytical technologies for ensuring food quality and authentication: A comprehensive review. Journal of Food Composition and Analysis 139:107075. https://doi.org/10.1016/j.jfca.2024.107075
15. Fatima S, Kumar V, Bhadauria G, Verma H (2023) Quality indicators based rapid test kits for detection of frying oil quality: A review. Food Chemistry Advances 2:100305. https://doi.org/10.1016/j.focha.2023.100305
16. Forooghi E (2023) Applications of Raman spectroscopy in food safety and quality control. HHHM 4:2. https://doi.org/10.30502/jhhhm.2024.430203.1082

17. Foster R, Williamson CS, Lunn J (2009) BRIEFING PAPER: Culinary oils and their health effects. Nutrition Bulletin 34:4–47. https://doi.org/10.1111/j.1467-3010.2008.01738.x
18. Gómez-Caravaca AM, Maggio RM, Cerretani L (2016) Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. Analytica Chimica Acta 913:1–21. https://doi.org/10.1016/j.aca.2016.01.025
19. Gong W, Shi R, Chen M, et al (2019) Quantification and monitoring the heat-induced formation of trans fatty acids in edible oils by Raman Spectroscopy. Food Measure 13:2203–2210. https://doi.org/10.1007/s11694-019-00140-5
20. Gu H, Huang X, Sun Y, et al (2021) Intelligent evaluation of total polar compounds (TPC) content of frying oil based on fluorescence spectroscopy and low-field NMR. Food Chemistry 342:128242. https://doi.org/10.1016/j.foodchem.2020.128242
21. Haji A, Desalegn K, Hassen H (2023) Selected food items adulteration, their impacts on public health, and detection methods: A review. Food Sci Nutr 11:7534–7545. https://doi.org/10.1002/fsn3.3732
22. Hu R, He T, Zhang Z, et al (2019) Safety analysis of edible oil products via Raman spectroscopy. Talanta 191:324–332. https://doi.org/10.1016/j.talanta.2018.08.074
23. Huang F, Li Y, Guo H, et al (2016) Identification of waste cooking oil and vegetable oil via Raman spectroscopy. J Raman Spectroscopy 47:860–864. https://doi.org/10.1002/jrs.4895
24. Huang X, Li H, Ruan Y, et al (2022) An integrated approach utilizing raman spectroscopy and chemometrics for authentication and detection of adulteration of agarwood essential oils. Front Chem 10:1036082. https://doi.org/10.3389/fchem.2022.1036082
25. Inteaz A (1989) Food Quality Assurance: Elements of Raw Materials Control. Canadian Institute of Food Science and Technology Journal 22:113–115. https://doi.org/10.1016/S0315-5463(89)70337-0
26. Jamieson LE, Li A, Faulds K, Graham D (2018a) Ratiometric analysis using Raman spectroscopy as a powerful predictor of structural properties of fatty acids. R Soc Open Sci 5:181483. https://doi.org/10.1098/rsos.181483
27. Jamieson LE, Li A, Faulds K, Graham D (2018b) Ratiometric analysis using Raman spectroscopy as a powerful predictor of structural properties of fatty acids. R Soc Open Sci 5:181483. https://doi.org/10.1098/rsos.181483
28. Johnson DS, Saikia DN (2009) Fatty acids profile of Edible Oils and Fats in India
29. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Phil Trans R Soc A 374:20150202. https://doi.org/10.1098/rsta.2015.0202
30. Kaimal A, Dhingra M, Singhal R (2021) Monitoring of oil quality from commercial fried foods-A case study from India. Journal of Food Processing and Preservation 46:. https://doi.org/10.1111/jfpp.16138
31. Kaur A, Scarborough P, Hieke S, et al (2016) The nutritional quality of foods carrying health-related claims in Germany, The Netherlands, Spain, Slovenia and the United Kingdom. European Journal of

Clinical Nutrition 70:114. https://doi.org/10.1038/ejcn.2016.114

32. Khursheed M (2024) Chromatographic Techniques for the Detection and Identification of Olive Oil Adulteration. ReiDoCrea: Revista electrónica de investigación Docencia Creativa. https://doi.org/10.30827/Digibug.86578
33. Kim N, Yu KS, Kim J, et al (2018) Chemical characteristics of potato chips fried in repeatedly used oils. Food Measure 12:1863–1871. https://doi.org/10.1007/s11694-018-9800-x
34. Korepanov VI (2020) Asymmetric least-squares baseline algorithm with peak screening for automatic processing of the Raman spectra. Journal of Raman Spectroscopy 51:2061–2065. https://doi.org/10.1002/jrs.5952
35. Kwofie F, Lavine BK, Ottaway J, Booksh K (2020) Differentiation of Edible Oils by Type Using Raman Spectroscopy and Pattern Recognition Methods. Appl Spectrosc 74:645–654. https://doi.org/10.1177/0003702819888220
36. Liu H, Chen Y, Shi C, et al (2020) FT-IR and Raman spectroscopy data fusion with chemometrics for simultaneous determination of chemical quality indices of edible oils during thermal oxidation. LWT 119:108906. https://doi.org/10.1016/j.lwt.2019.108906
37. Mendes TO, Da Rocha RA, Porto BLS, et al (2015) Quantification of Extra-virgin Olive Oil Adulteration with Soybean Oil: a Comparative Study of NIR, MIR, and Raman Spectroscopy Associated with Chemometric Approaches. Food Anal Methods 8:2339–2346. https://doi.org/10.1007/s12161-015-0121-y
38. Menze BH, Kelm BM, Masuch R, et al (2009) A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics 10:213. https://doi.org/10.1186/1471-2105-10-213
39. Moore JC, Spink J, Lipp M (2012) Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. Journal of Food Science 77:R118–R126. https://doi.org/10.1111/j.1750-3841.2012.02657.x
40. Navratilova K, Hurkova K, Hrbek V, Uttl L (2022) Metabolic fingerprinting strategy: Investigation of markers for the detection of extra virgin olive oil adulteration with soft-deodorized olive oils. Food Control 134: 108649. https://doi.org/10.1016/j.foodcont.2021.108649
41. Pereira FJ, López R, Ferrer N, et al (2021) A comparative appraisal of Raman band ratioing and chemometric analysis for classification of ancient papyri. Journal of Cultural Heritage 52:55–64. https://doi.org/10.1016/j.culher.2021.09.003
42. Program HF (2024) Economically Motivated Adulteration (Food Fraud) Research Publications. FDA. https://www.fda.gov/food/compliance-enforcement-food/economically-motivated-adulteration food-fraud
43. Rana R, Kumar N, Pandit A, et al (2014) Effect of various socio-economic factors on the consumption of processed potato products in Punjab. Indian Journal of Agricultural Marketing 28:24–35. https://www.researchgate.net/profile/Rajesh-Rana 3/publication/264889204_Effect_of_various_socio

economic_factors_on_the_consumption_of_processed_potato_products_in_Punjab/links/5771046808ae6219474a3263/Effect-of-various-socio-economic-factors-on-the-consumption-of-processed potato-products-in-Punjab.pdf

44. Rifna EJ, Pandiselvam R, Kothakota A, et al (2022) Advanced process analytical tools for identification of adulterants in edible oils – A review. Food Chemistry 369:130898. https://doi.org/10.1016/j.foodchem.2021.130898
45. Sharma R (2022). Food Safety and Standards Authority of India Press release, India. Retrieved from https.://www.fssai.gov.in/upload/press_release/2022/08/62fcb62aa2949Press_Release_Edible_oil_17_08_2022.pdf
46. Shehata M, Dodd S, Mosca S, et al (2024) Application of Spatial Offset Raman Spectroscopy (SORS) and Machine Learning for Sugar Syrup Adulteration Detection in UK Honey. Foods 13:2425. https://doi.org/10.3390/foods13152425
47. Su R, Xu X, Wang X, et al (2011) Determination of organophosphorus pesticides in peanut oil by dispersive solid phase extraction gas chromatography–mass spectrometry. Journal of Chromatography B 879:3423–3428. https://doi.org/10.1016/j.jchromb.2011.09.016
48. Tan C-H, Kong I, Irfan U, et al (2021) Edible Oils Adulteration: A Review on Regulatory Compliance and Its Detection Technologies. Journal of Oleo Science 70:1343–1356. https://doi.org/10.5650/jos.ess21109
49. Ün İ, OK S (2018) Analysis of olive oil for authentication and shelf life determination. J Food Sci Technol 55:2476–2487. https://doi.org/10.1007/s13197-018-3165-3
50. Vaskova H, Buckova M (2018) Multivariate study of Raman spectral data of edible oils. 14: 02022. https://doi.org/10.1051/matecconf/201712502022
51. Wang J, Qian J, Xu M, et al (2025) Adulteration detection of multi-species vegetable oils in camellia oil using Raman spectroscopy: Comparison of chemometrics and deep learning methods. Food Chemistry 463:141314. https://doi.org/10.1016/j.foodchem.2024.141314
52. Wu Y, Liu J, Xu R, et al (2024) Recent advances in ratiometric surface-enhanced Raman spectroscopy sensing strategies. Microchemical Journal 199:110127. https://doi.org/10.1016/j.microc.2024.110127
53. Yadav S (2018) Edible oil adulterations: Current issues, detection techniques, and health hazards. Int J Chem Stud 6:1393–1397
54. Yang T, Zhao B, He L (2019) Raman instruments for food quality evaluation. In: Evaluation Technologies for Food Quality. Elsevier, pp 119–143. https://www.chemijournal.com/archives/2018/vol6issue2/PartT/6-2-102-397.pdf
55. Yao S, Aykas DP, Rodriguez-Saona L (2020) Rapid Authentication of Potato Chip Oil by Vibrational Spectroscopy Combined with Pattern Recognition Analysis. Foods 10:42. https://doi.org/10.3390/foods10010042
56. Zhu P, Yang Q, Zhao H (2022) Identification of peanut oil origins based on Raman spectroscopy combined with multivariate data analysis methods. Journal of Integrative Agriculture 21:2777–

2785. https://doi.org/10.1016/j.jia.2022.07.026

Figures

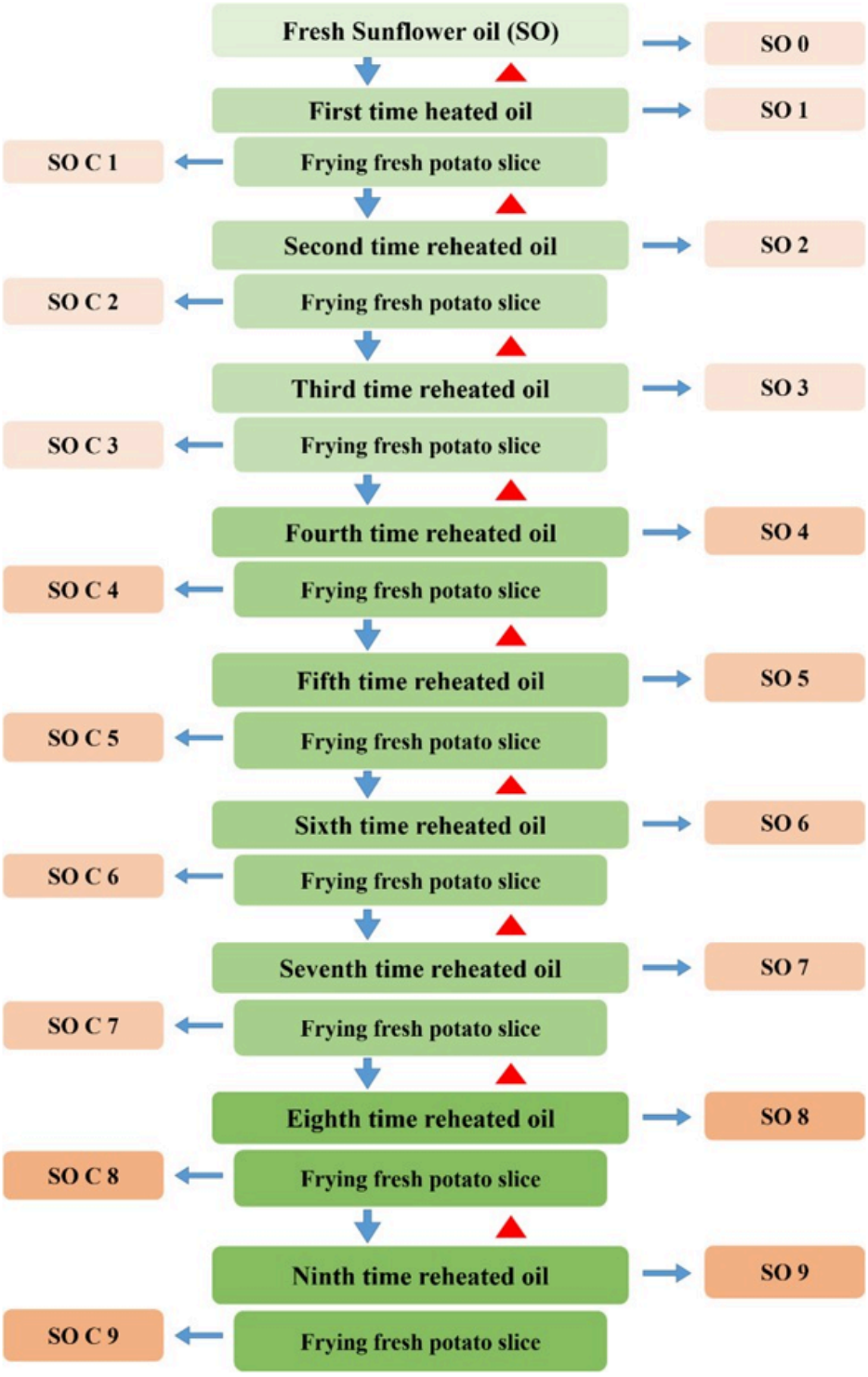

Figure 1

Steps followed in sample preparation (with Sunflower Oil as an example).300 mL of Sunflower oil was heated to frying temperature (195˚ C). Thin potato slices were added to the oil, and fried until golden brown. The fried chips were cooled to room temperature and collected in an airtightpouch, marked with the sample codes. After cooling the oil in the pan, aliquots of 5 mL (used oil) were collected in glass vials and were sequentially labelled in both oil and chips category (additional suffix ‘C’ for chips category).

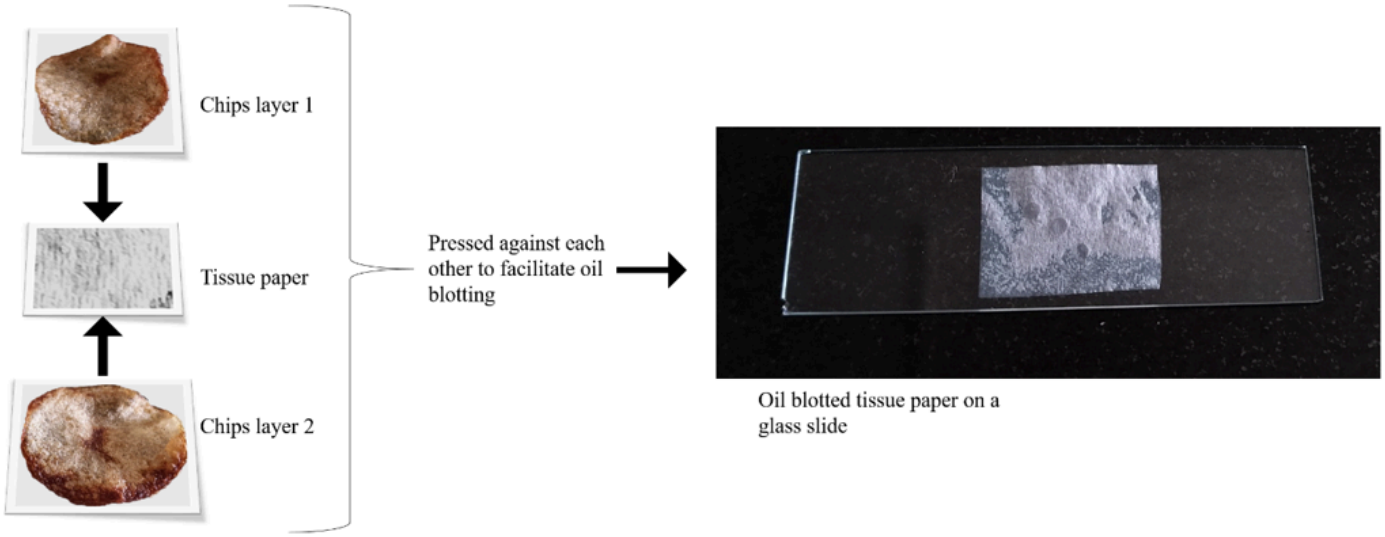

Figure 2

Schematic illustration of the developed tissue paper-based extraction method for the food matrix.

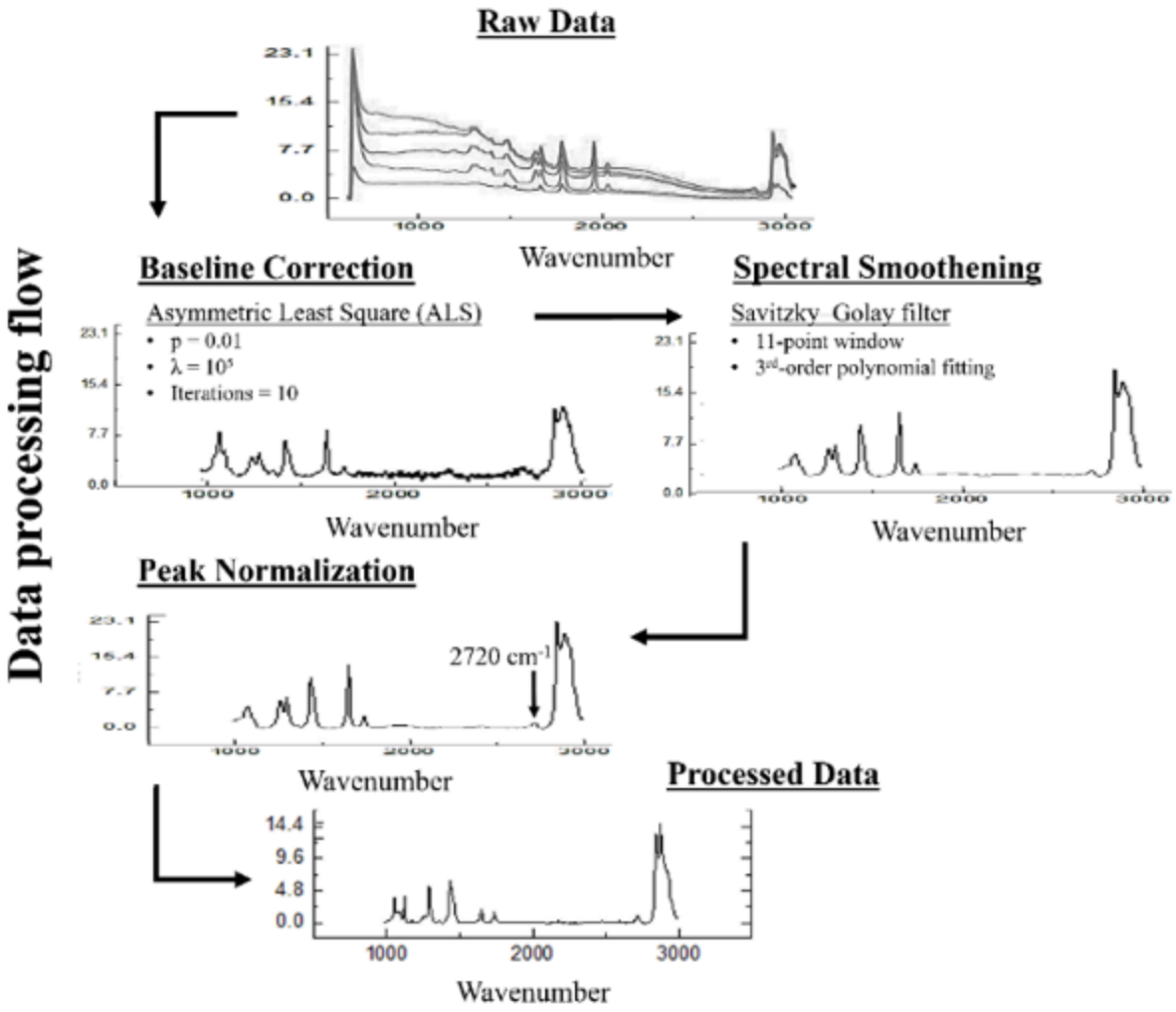

Figure 3

Schematic flow for spectral data processing used in the proposed framework

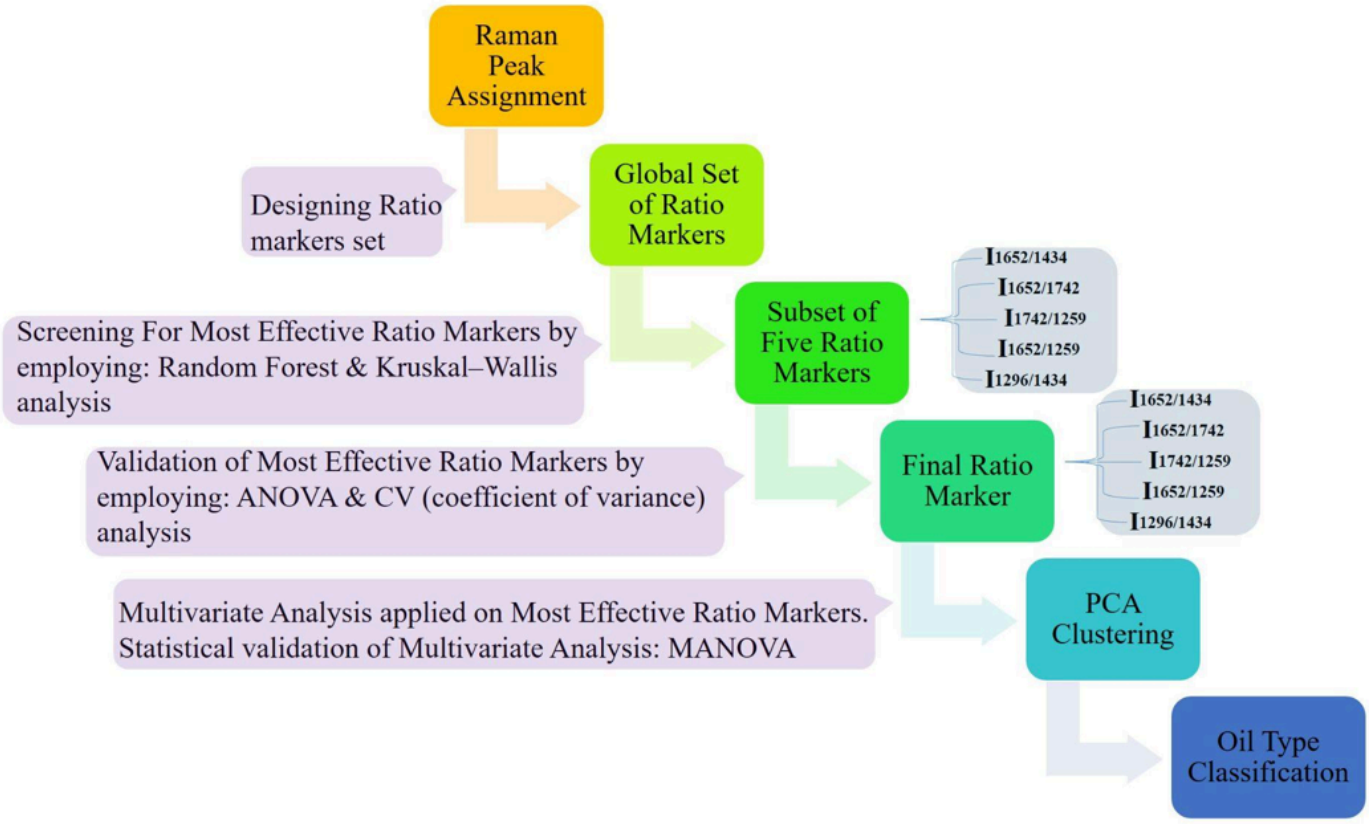

Figure 4

Schematic workflow illustrating the development and validation of the ratiometric markers in the current framework.

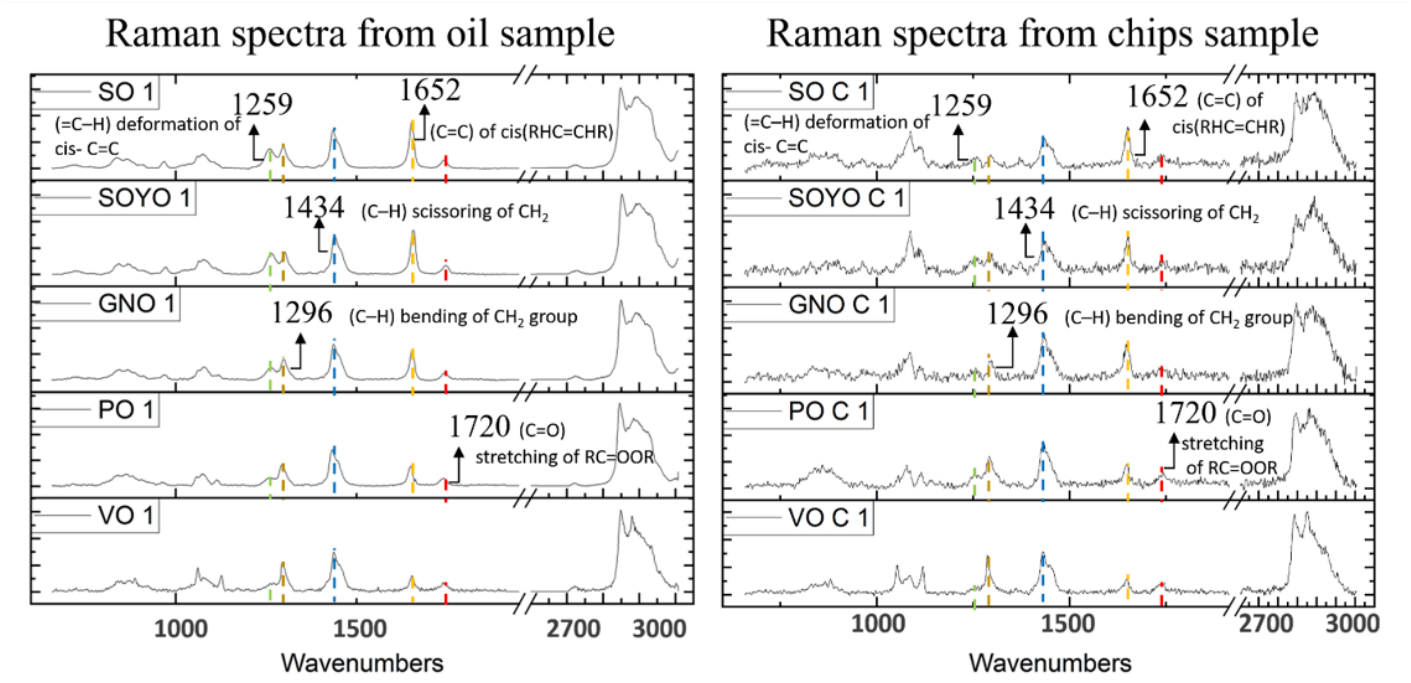

Figure 5

Baseline corrected, smoothened and normalized Raman spectra collected from oils and chips

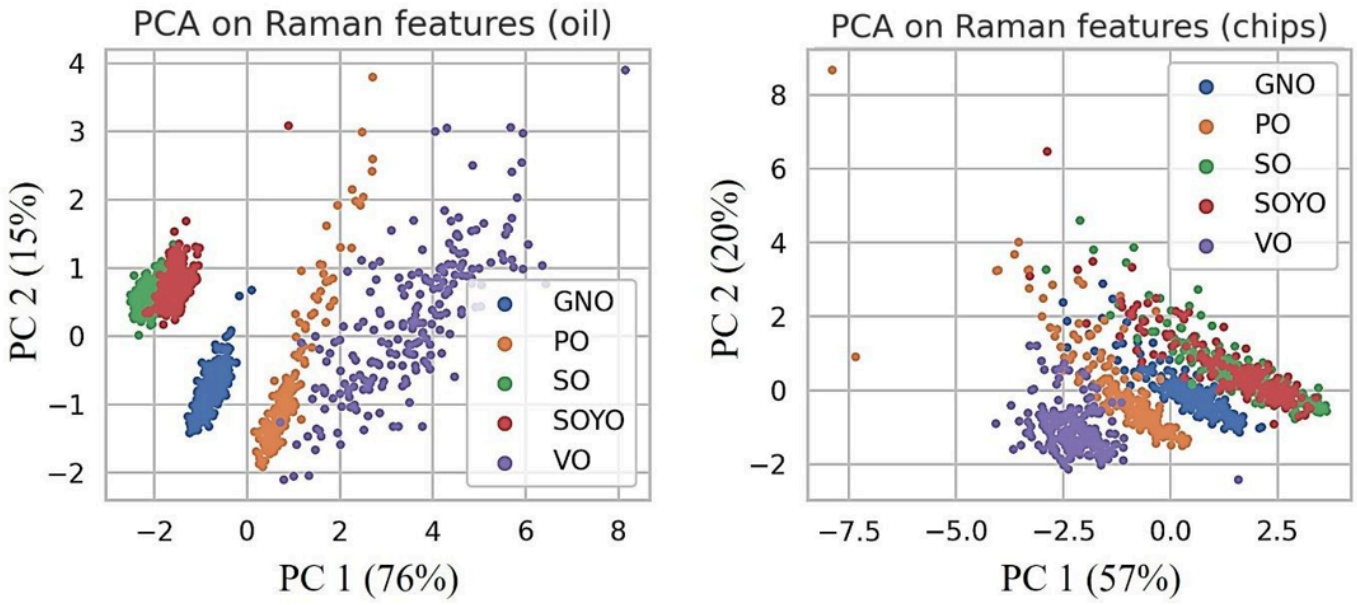

Figure 6

PCA plot from oil and chips samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicalabstract.docx