



## Article

# Discrimination/Classification of Edible Vegetable Oils from Raman Spatially Solved Fingerprints Obtained on Portable Instrumentation

Guillermo Jiménez-Hernández<sup>1</sup>, Fidel Ortega-Gavilán<sup>1,2,\*</sup> , M. Gracia Bagur-González<sup>1,\*</sup> and Antonio González-Casado<sup>1</sup> 

<sup>1</sup> Department of Analytical Chemistry, Faculty of Science, University of Granada, C/Fuentenueva w/n, 18071 Granada, Spain; gjh188@correo.ugr.es (G.J.-H.); agcasado@ugr.es (A.G.-C.)

<sup>2</sup> Animal Health Central Laboratory (LCSA), Department of Chemical Analysis of Residues, Ministry of Agriculture, Fisheries and Food, Camino del Jau w/n, 18320 Santa Fe, Spain

\* Correspondence: fog@ugr.es (F.O.-G.); mgbagur@ugr.es (M.G.B.-G.); Tel.: +34-958243327 (M.G.B.-G.)

**Abstract:** Currently, the combination of fingerprinting methodology and environmentally friendly and economical analytical instrumentation is becoming increasingly relevant in the food sector. In this study, a highly versatile portable analyser based on Spatially Offset Raman Spectroscopy (SORS) obtained fingerprints of edible vegetable oils (sunflower and olive oils), and the capability of such fingerprints (obtained quickly, reliably and without any sample treatment) to discriminate/classify the analysed samples was evaluated. After data treatment, not only unsupervised pattern recognition techniques (as HCA and PCA), but also supervised pattern recognition techniques (such as SVM, kNN and SIMCA), showed that the main effect on discrimination/classification was associated with those regions of the Raman fingerprint related to free fatty acid content, especially oleic and linoleic acid. These facts allowed the discernment of the original raw material used in the oil's production. In all the models established, reliable qualimetric parameters were obtained.

**Keywords:** Raman fingerprints; edible vegetable oils; sunflower oil; olive oil; pattern recognition techniques; portable analyser; SORS



**Citation:** Jiménez-Hernández, G.; Ortega-Gavilán, F.; Bagur-González, M.G.; González-Casado, A.

Discrimination/Classification of Edible Vegetable Oils from Raman Spatially Solved Fingerprints Obtained on Portable Instrumentation. *Foods* **2024**, *13*, 183. <https://doi.org/10.3390/foods13020183>

Academic Editor: Zhengjun Qiu

Received: 21 November 2023

Revised: 29 December 2023

Accepted: 29 December 2023

Published: 5 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sunflower oil from the seeds of *Helianthus annuus* L. [1] is produced in several countries, including Russia, Ukraine and China. In the European Union, it represents the second most important source of seed oils after rapeseed oil; the area under sunflower crops will increase up to 1.0 million ha by the end of 2031 [2]. The plant grows best in dry, temperate climates (with an average temperature of 20–25 °C) with high solar radiation, low humidity, and deep soils to spread its roots in search of nutrients and water. Its deshelled seeds are responsible for 80% of the total fruit weight and have a high oil content up to 55% w/w [3,4]. These oils have a high lipid content, most of which are triglycerides composed mainly of long-chain unsaturated fatty acids of different unsaturation, with linoleic acid (59%, polyunsaturated omega 6) being the most important [5,6]. In addition, the presence of other fatty acids such as oleic acid (30%, monounsaturated omega 9), stearic acid (6%, saturated), and palmitic acid (5%, saturated) should be noted.

Like olive oil (OO), there are several commercial categories of sunflower oil, the main ones being as follows: (i) refined sunflower oil (SFO), characterised by a high linoleic acid (LA) content; (ii) high oleic sunflower oil (HOSFO), with an oleic acid (OA) content of at least 75% measured as a percentage of the total fatty acid content; and (iii) medium oleic sunflower oil (MOSFO), with a seed OA content ranged from 50% to 75%. The last two commercial categories come from seeds genetically modified to naturally increase not only the ratio of oleic to linoleic acid but also other monounsaturated fatty acids or vitamins such

as vitamin E. These facts confer to these oils, especially to HOSFO, an overall composition with remarkable similarity to OO, and as an extra factor, greater resistance to oxidation and increased possibilities for use [7,8].

Currently, the main analytical methods used to check authenticity and evaluate possible adulteration of edible vegetable oils, especially OO, are chromatography-based analyses of the presence of triglycerides and fatty acids [9,10]. The most used are Thin-Layer Chromatography (TLC), High-Pressure Liquid Chromatography (HPLC), and Gas Chromatography (GC). TLC is used both qualitatively and quantitatively in official methods such as the analysis of sterol and stanol fractionation in oils [11,12]. HPLC is, among other uses, an International Olive Council reference method commonly used in routine laboratories [13,14]. GC is used not only as the GC-FID (Flame Ionization Detector) official method specified in IOCs guides, but also for characterising the triglyceride profile in vegetable oils, or for the separation and quantification of fatty acids according to their esterified fraction (FAME) by means of a prior hydrolysis and methylation of fatty acids step [15–17].

As an alternative to chromatographic techniques, the spectroscopic ones (Raman, MIR, FIR, NMR, etc.) provide simple, fast and reliable results, with the advantage of being used directly on the sample without the need for any sample pre-treatment stage. Moreover, they appear to be not only economical but also environmentally sustainable analytical methods [17–19]. NMR provides a quick way of measuring the oil content of oil seeds such as sunflower seeds, as well as assessing their resistance to high temperatures by evaluating the degradation of their constituents in their main domestic use, i.e., food frying [20]. Some of these techniques, such as Raman or NIR spectroscopy, can be easily adapted to portable devices, allowing the acquisition of analytical data in situ in real time, although in many cases these techniques have proven to be incapable of measuring through packaging. To overcome this inconvenience, some instrumental improvements have been made to Raman spectroscopy, as in the case of the Spatially Offset Raman spectroscopy (SORS) technique [21]. Despite these advantages, they have a lower resolution capability than chromatographic techniques, as they do not provide information on each individual compound present, but rather on the bonds that occur in the components of a sample as a whole.

Considering that the final output of an analytical device (bench top, handheld or portable) is an analytical signal, the use of this to obtain information about the properties of a material related to its chemical composition, e.g., fingerprinting methodology, has shown to be a powerful tool to identify, discriminate and authenticate edible oils among other commodities. In this sense, these signals may relate to the identity of a food, to its physico-chemical or other natural properties, as well as to the presence or quantity of compounds in its chemical composition [22,23]. Despite this, instrumental fingerprints contain hidden information, which normally requires the use of chemometric tools in particular, or machine learning methods in general. Multivariate data analysis is essential for handling statistically complex data such as Raman fingerprints. The extraction of valuable information from large and complex chemical measurement data has become a significant research area. Chemometrics and machine learning are interdisciplinary methods that integrate chemistry, mathematics, computer science, and other fields to solve data-processing problems in modern analytical instruments [24–26].

In the food sector, fingerprinting methodology has been widely used to solve different problems, such as the authentication and discrimination of olive oils, margarines, and fat spread, or the evaluation of the quality and production process of spirits, among other foodstuffs [27–31]. Regarding the use of handheld or portable devices based on SORS applications, few studies have been carried out despite the great potential of the technique [21]. Recent examples include the detection of possible adulteration in alcoholic beverages [32]; the study of the evolution of the alcoholic fermentation process in white wine [33]; the authenticity assessment of the origin of animal milk used in the production of cheese; the characterisation of several commercial categories of cheese (Cheddar, Manchego

and Pecorino Romano); the analysis of packaged margarines and fat spreads [34–36]; and the study of the adulteration of extra virgin olive with other edible vegetable oils, as in the case of Varnasseri et al. [37].

Thus, the aim of this work is to propose a more environmentally friendly and sustainable analytical method that combines the Raman fingerprints obtained from a portable instrument with chemometrics/machine learning, to build classification models to reliably differentiate sunflower oils from olive oils of different commercial categories.

## 2. Materials and Methods

### 2.1. Oil Samples

A sample bank consisting of 145 samples from different types of edible vegetable oils, purchased at local supermarkets, was used for this study; these were separated into two main groups: sunflower and olive oil. The first group was constituted by 48 samples where 7 of them were labelled as high oleic acid content. The second group was composed of a total of 97 samples from which 65 were extra virgin olive oil (EVOO), 22 were virgin olive oil (VOO), and 10 were pomace olive oil (POO).

### 2.2. Spectroscopic Analysis

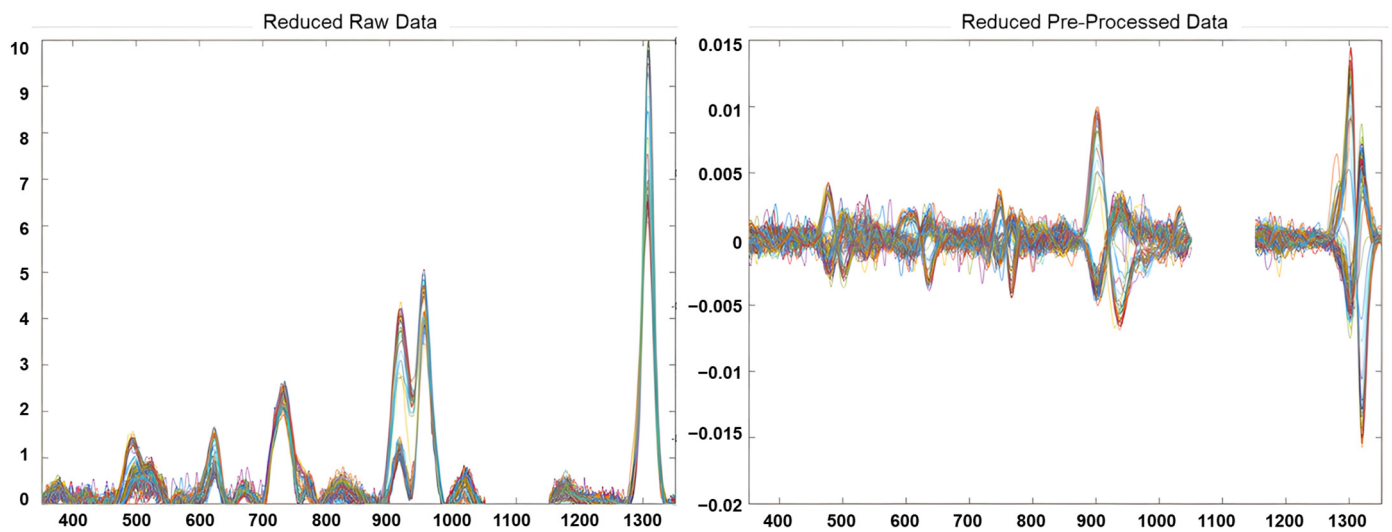
A Vaya Raman portable spectrometer (Agilent Technologies, Santa Clara, CA, USA) was used in this study. It was equipped with a 3B class diode laser operating at 830 nm and a maximum power of 450 mW. The exposure time of the sample to the laser varied between 0.5 and 2 s. The spectral resolution of the device was 12–20  $\text{cm}^{-1}$ , the wavenumber range between 735–1540  $\text{cm}^{-1}$ , and a cooled CCD (charge-coupled device) was used as a detection system. The specific offset length from the incidence point was fixed by equipment at 0.6 cm. For measurement (with a total measurement time between 30 s and 2 min), edible oil samples were introduced into 4 mL borosilicate glass vial of 1mm thickness.

After carrying out the measurement, the equipment software performed the following: (i) a correction from the information was obtained to eliminate any possible influence of the container, (ii) a baseline adjustment, and finally (iii) a normalisation of the intensity values. Therefore, the result of each analysed sample was a Raman spectrum from 350 to 2000  $\text{cm}^{-1}$  in which the intensity values were between 0 and 1. Thus, the final spectrum was a normalised spatially resolved Raman spectrum (NSR Raman Spectrum).

### 2.3. Data Treatment

Each NSR RAMAN spectrum was collected in .CSV format (comma separated value) and exported to .mat format for the subsequent elaboration of the fingerprints matrix within the MATLAB environment (version 9.3, Mathworks Inc., Natick, MA, USA). As a result, a comprehensive 145  $\times$  1651 data matrix consisting of 145 NSR RAMAN fingerprints (rows), each of which was composed of 1651 normalised intensity values (columns), was obtained as a raw data matrix and once pre-processed, used for the different chemometric studies.

The preprocessing stage included the following: a selection of interval of interest for each fingerprint; a filtering smoothing of the signals using a Savitzky-Golay filter (1st derivate, 2nd order polynomial and a filter width of 21 points); and a mean centre, obtaining a reduced-preprocessed NSR RAMAN fingerprints matrix of 145 samples  $\times$  902 variables. Figure 1 shows the NSR RAMAN fingerprint of edible vegetable oils analysed for this study before and after the pre-processing step. Unsupervised and supervised pattern recognition techniques were explored using PLS\_Toolbox, version 8.6.1 (Eigenvector Research Inc., Manson, WA, USA) [38].



**Figure 1.** Reduced RAMAN fingerprints of edible vegetable oils before and after data pre-processing.

### 3. Results and Discussion

#### 3.1. Unsupervised Pattern Recognition Methods

The analysis of natural grouping trends could allow for establishing a correlation between the data in the NSR RAMAN fingerprints and their impact on the vegetable origin of the edible oils analysed. This would lead to evaluating the use of green analytical technology such as a ‘Vaya Raman’ portable spectrometer, to distinguish between vegetable oils obtained from different types of raw materials such as olive and sunflower.

##### 3.1.1. Hierarchical Cluster Analysis of Raman Edible Vegetable Oil Fingerprints

An HCA was performed using the previously defined data matrix. Ward’s method and Manhattan distance were used as the linkage criterion and measure of distance between pairs of observations, respectively. The number of clusters was selected using  $D_{linkage} = 2/3$  of  $D_{max}$  as the internal criterion. The edible oils clustered naturally, based on the raw material used and the oleic acid content. Thus, two main clusters could be observed. Group I was constituted by oils obtained from sunflower seeds, while group II included all the olive oils (extra virgin, virgin, and pomace) and high oleic sunflower oils (Figure 2).

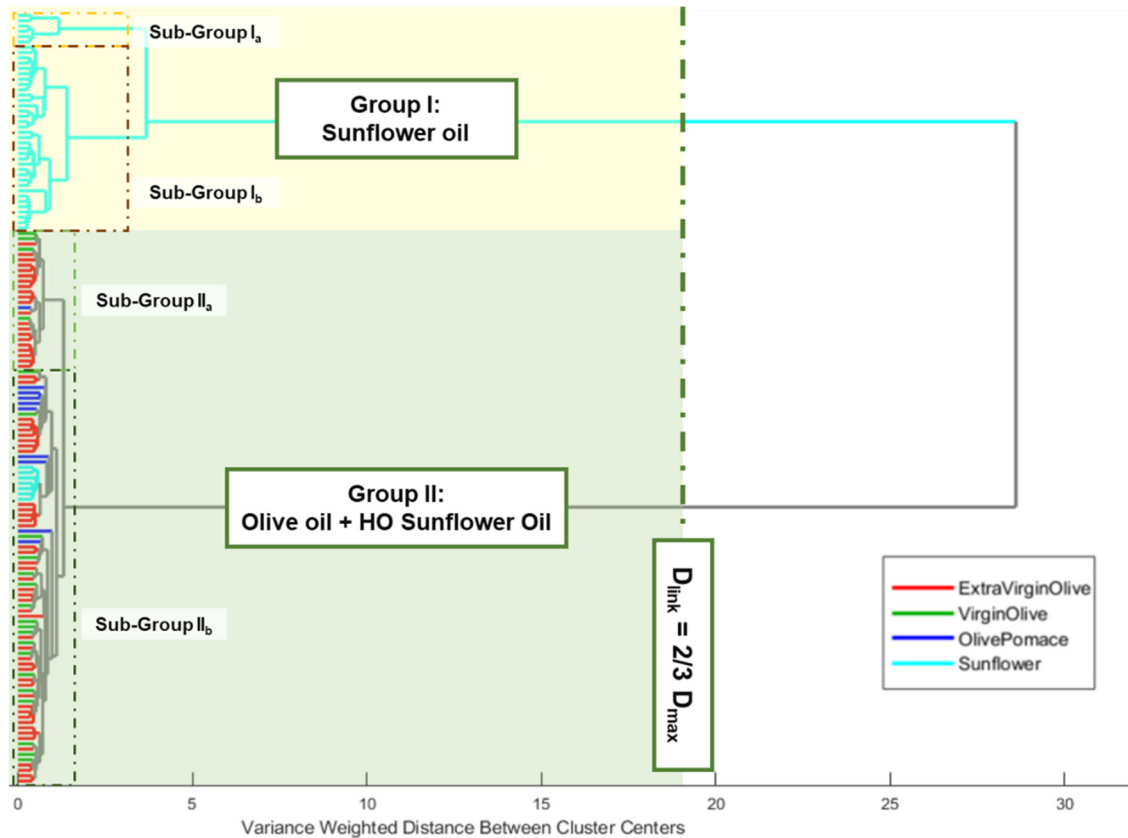
Regarding group I, it also could be observed that a previous partial clustering took place according to the percentage of oleic acid shown on the label of the commercial container, where natural nesting distinguished those sunflower oils with oleic acid, locating them in the upper part of the group. Although a similar behaviour could be observed when group II was analysed, in this case the nesting order was the opposite to the previous one, due to the upper partial cluster groups of olive oil samples with low acidity, which implied a lower free oleic acid content. Meanwhile, those samples with a higher content of this acid appear grouped in the second partial cluster (some EVOO, refined and pomace samples). Thus, the appearance of those sunflower oil samples (commercially labelled as high oleic acid content sunflower oils) in this last sub-group could be justified.

##### 3.1.2. Principal Component Analysis

When PCA was applied to the data matrix, five principal components (PCs) were obtained which explained 92.42% of the variance of the model for the edible vegetable oils analysed. PC1 explained 88.99% of the total variance of the system, while the other four principal components explained the remaining 3.43% of the variance (PC2: 1.06%; PC3 0.94%; PC4: 0.73% and PC5: 0.70%).

In Figure 3a, the scores received by the analysed oil samples in the first two components (PC2 vs. PC1) are illustrated. The edible vegetable oils are grouped according to the raw material used in the elaboration process, similar to what occurs when HCA is applied.

Most of the sunflower oils received negative scores for PC1 (Group I), while the olive oils scored positively for this component (Group II). Additionally, the high oleic sunflower oils included in this group received negative scores for PC2.

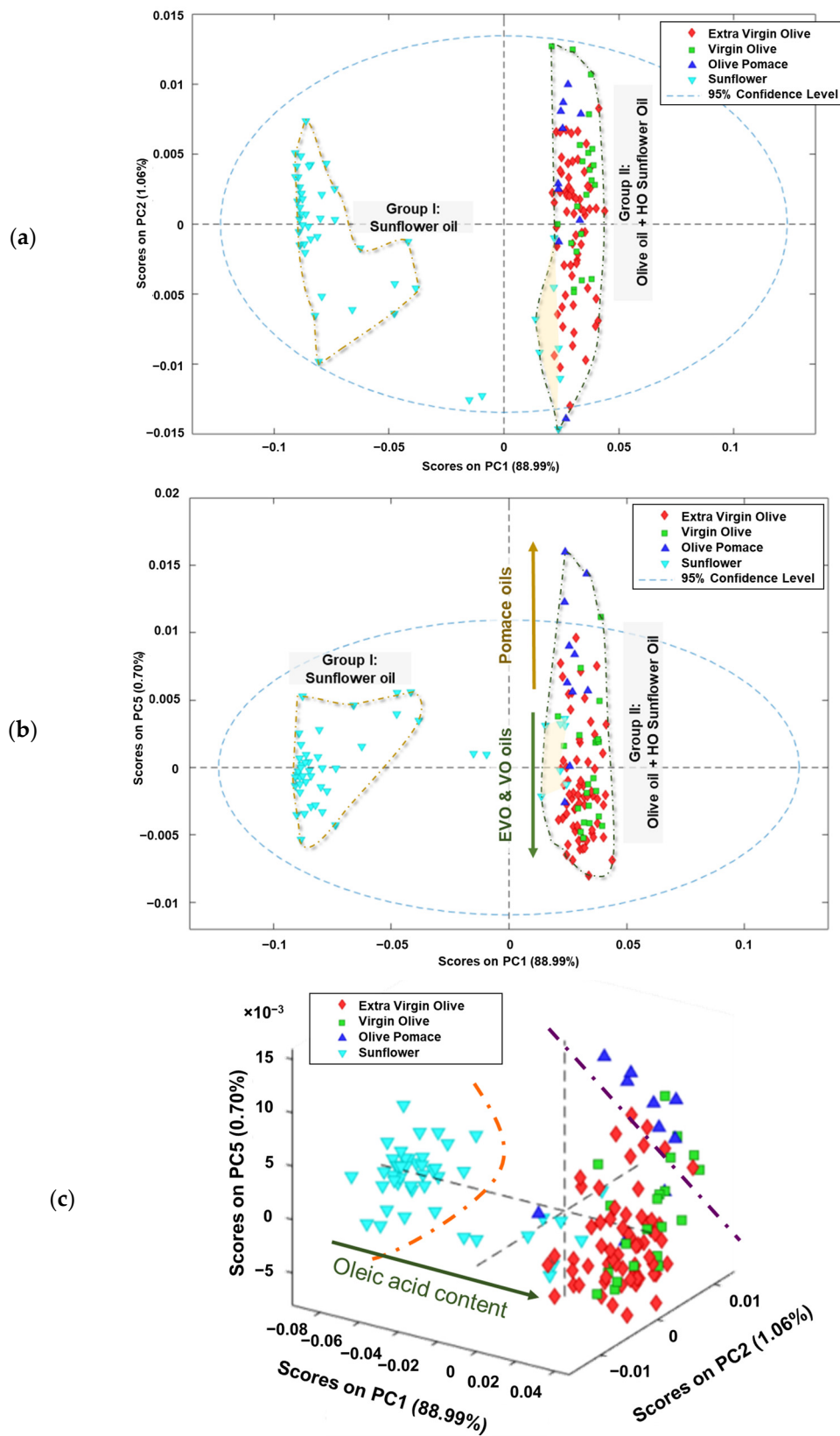


**Figure 2.** Dendrogram from HCA of analysed edible vegetable oils.

Figure 3b presents a graphical representation of the scores received by the analysed oil samples in the PC5 vs. PC1 space. The distribution of the edible oils' scores along the PC1 space reveals the same main grouping as the raw material from which the oils are obtained (Group I and II). Additionally, a new tendency is observed in Group II, which can be explained by the oleic acid content. The oils with a high oleic acid content received positive scores on PC5, while the others were in the negative area of this component. In addition, it was observed that higher oleic acid content in the oil samples resulted in higher positive scores on PC1. These trends can be explained by examining the fingerprint regions with minimal variations in total variance, indicating minimal differences between fingerprints.

When the 3D space is defined as PC5 vs. PC2 vs. PC1 (Figure 3c), a clearer grouping can be observed. Thus, sunflower oils received increasingly positive PC1 scores, as the oleic acid content was higher. In addition, the olive pomace oils received positive values from PC1, PC2, and PC5, which could allow it to distinguish between sunflower, virgin/extra virgin olive oils, and olive pomace oils. These results are consistent with those found by several authors, indicating that olive oil and sunflower oil can be distinguished regardless of whether the study is to identify adulterations or to authenticate samples [39,40]. However, it is worth noting that when the oleic acid content is taken into account, sunflower oil can resemble olive oil in cases where the seed has been genetically modified to have this characteristic.





**Figure 3.** Score plots of analysed edible vegetable oils in: (a) plane PC2 vs. PC1, (b) plane PC5 vs. PC1 and (c) the 3D space defined by PC1 vs. PC2 vs. PC5.

If the loading plots of each component (Figure 4a–c) are analysed, based on Figure 4a, it can be observed that the PC1 loadings would allow the identification of those variables

that explain the natural grouping according to the raw material. The variables (wave numbers) associated with the bands in “conventional” Raman spectra associated with free fatty acid (mainly oleic and linoleic acids) [41] are as follows: (i)  $\nu(\text{C}=\text{C})$  vibration of *cis*-alkene, stretching of double bond ( $\approx 1650 \text{ cm}^{-1}$ ); (ii)  $\delta(\text{C}-\text{H})$  twisting of  $\text{CH}_2$  ( $\approx 1300 \text{ cm}^{-1}$ ); (iii)  $\delta(\text{C}=\text{H})$  scissoring ( $\approx 1250 \text{ cm}^{-1}$ ); (iv)  $\nu(\text{C}-\text{C})$  and  $\delta(\text{C}=\text{C})$  twisting of *trans*-alkene ( $\approx 1000 \text{ cm}^{-1}$ ); and (v)  $\nu(\text{C}-\text{H})$  symmetric and/or asymmetric stretching ( $\approx 800 \text{ cm}^{-1}$ ). Regarding Figure 4b, it can be observed that for the PC2 loadings the wavenumbers in which the free fatty acid information appears are as follows: (i)  $\nu(\text{C}=\text{C})$  vibration of *cis*-alkene ( $\approx 1650 \text{ cm}^{-1}$ ); and (ii)  $\nu(\text{C}-\text{H})$  symmetric and/or asymmetric stretching ( $\approx 800 \text{ cm}^{-1}$ ), the first one being the most influential in the natural grouping observed. When the PC5 loadings are considered (Figure 4c), the natural grouping could once again be associated with the same wavenumbers in PC2. The main difference between PC2 and PC5 loadings arose from the  $\nu(\text{C}-\text{H})$  symmetric and/or asymmetric stretching ( $\approx 800 \text{ cm}^{-1}$ ). Thus, slight variations on these Raman modes provoke the separation between the edible olive oils analysed. This is consistent with previous publications that have characterised the main Raman bands of EVOO and SFO, where the different vibrational modes of  $\text{C}=\text{C}$  and  $\text{C}-\text{C}$  can distinguish between the spectra of EVOO and SFO. This fact confirms that Raman spectroscopy is a reliable analytical technique not only for adulteration studies, as Philippidis et al. quoted in 2017, but also for authentication purposes [42–45].

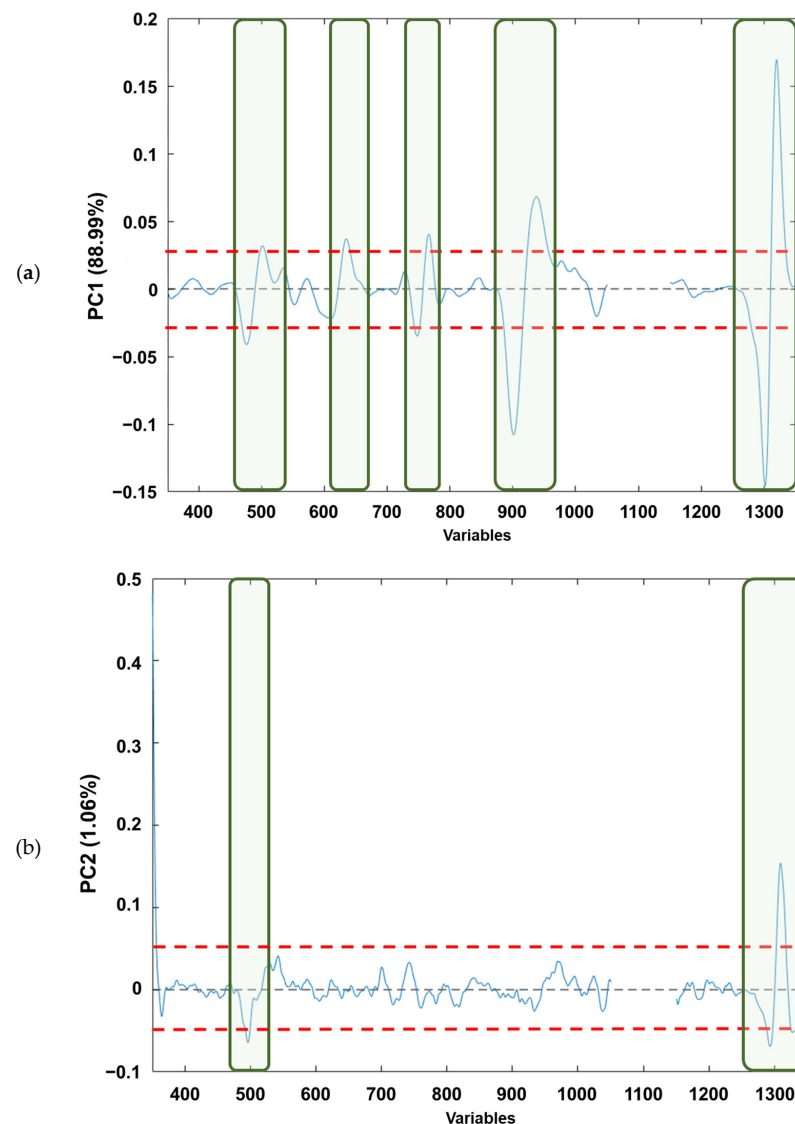
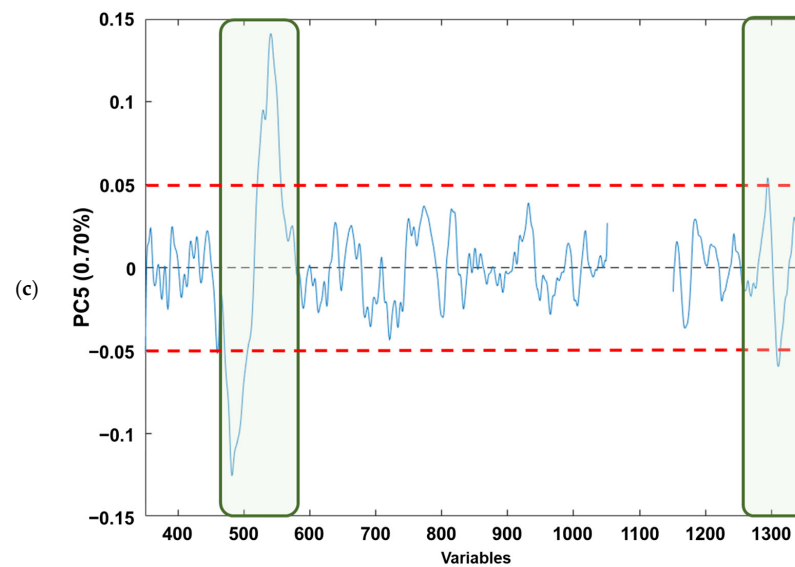


Figure 4. Cont.



**Figure 4.** Loading plots from analysed edible vegetable oils in: (a) PC1, (b) PC2 and (c) PC5. The red dashed lines indicate the loading thresholds above which their contribution to natural sample grouping is considered.

### 3.2. Supervised Pattern Recognition Methods

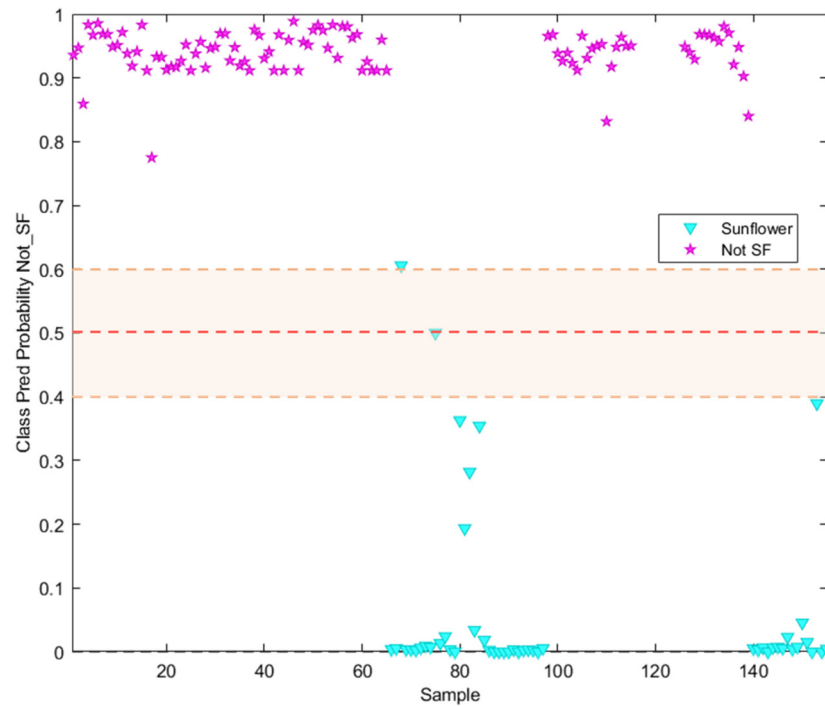
In order to study the capability of the NSR RAMAN fingerprint to discriminate/classify between the edible vegetable oils analysed, several one-input class classification/discrimination models, as SVM, kNN and SIMCA, were developed. For these purposes, the target class was assigned to sunflower oils, with the not-SF (olive oils) considered the non-target class. The two necessary sets (calibration and prediction) were selected using the Kennard–Stone algorithm [38,39,46], keeping 66.6% of the original samples for calibration/cross-validation purposes and the remaining 33.4% for external prediction purposes.

The SVM model was developed without reducing the data dimensionality (uncompressed). The Kernel algorithm used a radial basis function (RBF) with the default values for gamma and cost parameters in the PLS\_Toolbox™.

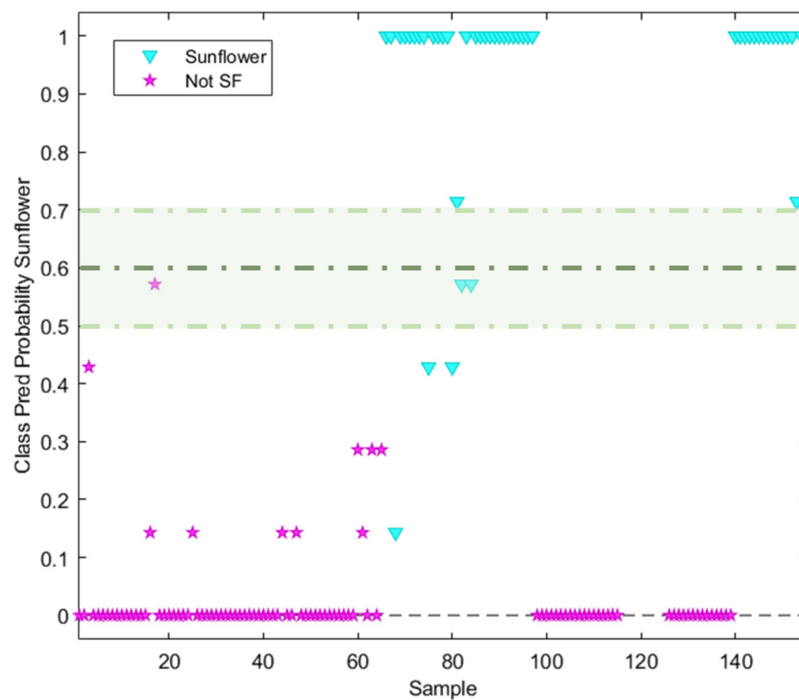
It can be seen in Figure 5 how only one sample labelled as sunflower oil was classified as not-SF. In addition, other samples labelled as Sunflower appeared closer to the threshold, which should be considered as badly classified despite appearing in the sunflower prediction probability area. Regarding the information included in the commercial label of the analysed samples, all of them appear described as high oleic sunflower oils with different content percentages of this acid, which explains the behaviour illustrated in this Figure.

The discrimination/classification power of the NSR RAMAN fingerprint was evaluated using neighbour distances. A kNN with  $k = 7$  was found to be the best value for deciding the neighbour distance in the model. The sunflower class was defined by a predicted probability value of 1, while the non-target class (olive oil) was defined by a probability of 0. Considering the two hard modelling techniques used, a similar behaviour can be observed, the main difference being the probability prediction value belonging to each class. All the sunflower samples labelled without high oleic acid content were well classified (probability = 1), with the exception of those samples labelled as containing oleic acid (7 samples). These had an assigned probability lower than 1 (Figure 6), which is why these samples were classified by the model as not-SF. When the non-targeted class was analysed, only 10 samples had an assigned probability value higher than 0. These samples corresponded mostly to pomace olive oil, for which the NSR RAMAN fingerprint should be different from the remaining olive oil fingerprints.





**Figure 5.** Probability prediction plot obtained from the SVM classification model. Dashed lines indicate class thresholds separating misclassified samples.



**Figure 6.** Probability prediction plot obtained from kNN ( $k = 7$ ) classification model. Dashed lines indicate class thresholds separating misclassified samples.

Finally, a SIMCA model as a soft modelling technique was performed. Three PCs were selected for the target and non-target classes, which explains the 93.4% and 30.5% of the cumulative variance, respectively.

Figure 7 shows the Cooman’s classification plot for the analysed samples. It can be seen that some samples appeared as not conclusive (SFO samples labelled as containing

a high oleic acid content); as outliers (three SFO samples and most of the POO samples); and misclassified (a HOSFO sample assigned to the non-target class). Compared with the non-supervised pattern recognition techniques, HCA and PCA, these samples appear grouped in the olive oil samples cluster or region.

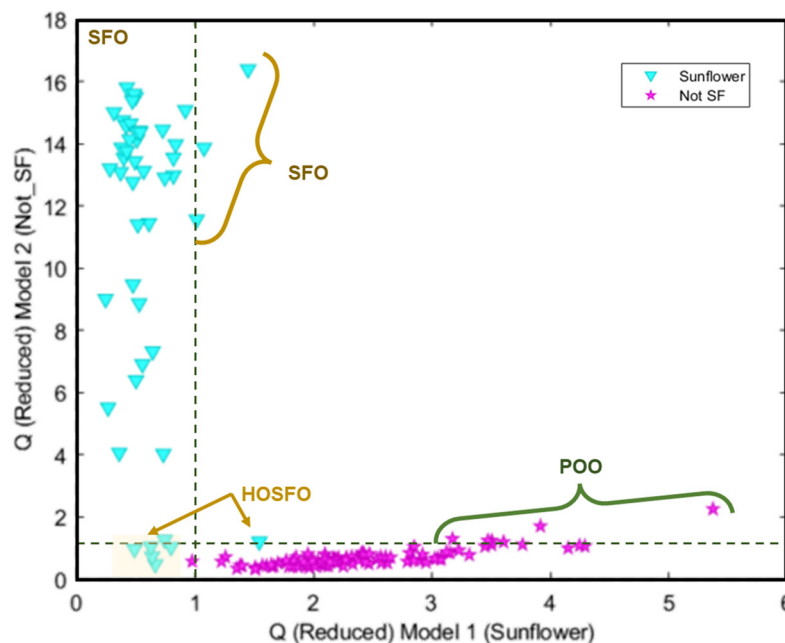


Figure 7. Cooman’s classification plots obtained for SIMCA model.

Tables 1–3 show the quality parameters of these three models. In general, the SVM model produced the best result, followed by the kNN and SIMCA. Thus, it can be considered that the hard models correctly discriminate the samples according to the raw material. Nevertheless, the SIMCA soft model showed a better capacity to classify those sunflower oil samples with high oleic acid content, considering that they appear not only as sunflower but also as not-sunflower i.e., olive oil.

Table 1. Summary of discrimination/classification performance metrics obtained for the SVM-DA one-input class model.

TARGET Class (TC): Sunflower Oil		
Features:		
X Block: [reduced RAMAN instrumental fingerprints]		
Y Block: [TC (sunflower oil, SFO); NTC (not sunflower oil, not SFO)]		
Preprocessing: 1st Derivative (order 2, window; 21 pt, tails: polyinterp) + Mean center		
Training Set: [97 × 902]		
Prediction Set: [48 × 902] See Confusion Table below		
Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS -prediction stage)	0.98	1.00
Specificity (SPEC-prediction stage)	1.00	0.98
False positive rate (FPR)	0.00	0.02
False negative rate (FNR)	0.02	0.00
Positive predictive value (precision) (PPV)	1.00	0.99
Negative predictive value (NPV)	0.99	1.00
Youden index (YOUD)	0.98	0.98
F-measure (F)	0.99	0.99
Discriminant power (DP)	-	-
Efficiency (or accuracy) (EFFIC)	0.99	0.99
Misclassification rate (MR)	0.01	0.01

**Table 1.** *Cont.*

TARGET Class (TC): Sunflower Oil		
AUC (correctly classified rate)	0.99	0.99
Gini coefficient (Gini)	0.98	0.98
G-mean (GM)	0.99	0.99
Matthews correlation coefficient (MCC)	0.98	0.98
Chance agreement rate (CAR)	0.56	0.56
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.98	0.98

**Confusion Table:**

<b>Assigned class</b>		48	97	145
	<b>Not SF</b>	1	97	98
	<b>Sunflower</b>	47	0	47
		<b>Sunflower</b>	<b>Not SF</b>	<b>T</b>
	<b>Reference class</b>			

Note. The hyphen “-” refers to metrics that cannot be determined since a division by zero is involved.

**Table 2.** Summary of discrimination/classification performance metrics obtained for kNN-DA one-input class model.

TARGET Class (TC): Sunflower Oil		
Features:		
X Block: [reduced RAMAN instrumental fingerprints]		
Y Block: [TC (sunflower oil, SFO); NTC (not sunflower oil, not SFO)]		
Preprocessing: 1st Derivative (order 2, window; 21 pt, tails: polyinterp) + Mean center		
Training Set: [97 × 902]		
Prediction Set: [48 × 902] See Confusion Table below		
Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS-prediction stage)	0.90	1.00
Specificity (SPEC-prediction stage)	1.00	0.90
False positive rate (FPR)	0.00	0.10
False negative rate (FNR)	0.10	0.00
Positive predictive value (precision) (PPV)	1.00	0.95
Negative predictive value (NPV)	0.95	1.00
Youden index (YOUJ)	0.90	0.90
F-measure (F)	0.95	0.97
Discriminant power (DP)	-	-
Efficiency (or accuracy) (EFFIC)	0.97	0.97
Misclassification rate (MR)	0.03	0.03
AUC (correctly classified rate)	0.95	0.95
Gini coefficient (Gini)	0.90	0.90
G-mean (GM)	0.95	0.95
Matthews correlation coefficient (MCC)	0.92	0.92
Chance agreement rate (CAR)	0.57	0.57
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.92	0.92

**Confusion Table:**

<b>Assigned class</b>		97	48	145
	<b>Sunflower</b>	0	43	43
	<b>Not SF</b>	97	5	102
		<b>Not SF</b>	<b>Sunflower</b>	<b>T</b>
	<b>Reference class</b>			

Note. The hyphen “-” refers to metrics that cannot be determined since a division by zero is involved.

**Table 3.** Summary of discrimination/classification performance metrics obtained for the SIMCA one-input class model.

TARGET Class (TC): Sunflower Oil		
Features:		
X Block: [reduced RAMAN instrumental fingerprints]		
Y Block: [TC (sunflower oil, SFO); NTC (not sunflower oil, not SFO)]		
Preprocessing: 1st Derivative (order 2, window; 21 pt, tails: polyinterp) + Mean center		
Training Set: [97 × 902]		
Prediction Set: [48 × 902] See Confusion Table below		
Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS -prediction stage)	0.93	0.81
Specificity (SPEC-prediction stage)	0.81	0.93
False positive rate (FPR)	0.19	0.07
False negative rate (FNR)	0.07	0.19
Positive predictive value (precision) (PPV)	0.99	1.00
Negative predictive value (NPV)	1.00	0.99
Youden index (YOUd)	0.74	0.74
F-measure (F)	0.96	0.90
Discriminant power (DP)	0.96	0.96
Efficiency (or accuracy) (EFFIC)	0.89	0.89
Misclassification rate (MR)	0.11	0.11
AUC (correctly classified rate)	0.87	0.87
Gini coefficient (Gini)	0.74	0.74
G-mean (GM)	0.87	0.87
Matthews correlation coefficient (MCC)	0.86	0.86
Chance agreement rate (CAR)	0.51	0.51
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.78	0.78

**Confusion Table:**

<b>Assigned class</b>		<b>97</b>	<b>48</b>	<b>145</b>
	<b>Nr</b>	1	5	6
	<b>I</b>	6	3	9
	<b>Sunflower</b>	0	39	39
	<b>Not SF</b>	90	1	91
		<b>Not SF</b>	<b>Sunflower</b>	<b>T</b>
		<b>Reference class</b>		

Note. Nr: not recognised; I: inconclusive.

#### 4. Conclusions

The chemometric/machine learning study of the spectroscopic instrumental fingerprints of edible vegetable oils (sunflower and olive oils), obtained using a highly versatile portable analyser based on Spatially Offset Raman Spectroscopy (SORS), showed that they are capable of distinguishing the analysed samples according to the original raw material used in the oil's production.

Additionally, the obtained results are consistent with the revised bibliography. The normalised spatially resolved Raman fingerprints of different zones not only allowed for grouping of samples based on raw material, but also for differentiation within the groups based on oil types. Therefore, despite using a shorter wavelength range, the unsupervised techniques employed (HCA and PCA) demonstrated that the raw material, such as sunflower or olive oil, was the most influential attribute followed by the types (commercial categories) of each type of oil, both determined by the oleic acid content and the oleic/linoleic acid ratio in the analysed samples. This highlights the functional qualities of this type of sunflower oil and its importance in the sunflower oil market.

By applying supervised techniques, different models to discriminate (SVM and kNN) and classify (SIMCA) were obtained. Although reliable quality metrics of the models were satisfactory, as far as the authors are concerned, the best results were obtained in the SIMCA model. This soft classification model permitted not only the classification of samples according to the oil raw material (seed or fruit), but also the classification of those samples belonging to HOSFO as inconclusive due to their high oleic acid content similar to olive oil.

Finally, these results allow for the proposed combination of SORS and chemometric tools to be considered not only for use in quality control for routine analysis in the food industry, but also in other fields including the authentication of different types of edible oils, such as sunflower oil. Furthermore, instrumental fingerprints can be easily obtained through short-time analysis without any sample processing. This, combined with simple chemometric analysis, allows for the quality of genetically modified sunflower seeds to be highlighted for use in frying processes and baking.

**Author Contributions:** Conceptualization, M.G.B.-G. and A.G.-C.; formal analysis, F.O.-G. and M.G.B.-G.; funding acquisition, A.G.-C.; investigation, G.J.-H. and F.O.-G.; methodology, G.J.-H., F.O.-G. and M.G.B.-G.; project administration, M.G.B.-G. and A.G.-C.; resources, A.G.-C.; software, G.J.-H., F.O.-G. and M.G.B.-G.; supervision, F.O.-G., M.G.B.-G. and A.G.-C.; validation, F.O.-G. and M.G.B.-G.; writing—original draft, G.J.-H., F.O.-G. and M.G.B.-G.; writing—review & editing, F.O.-G., M.G.B.-G. and A.G.-C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The corresponding authors can provide the data presented in this study upon request. The data is not publicly available at this time but will be uploaded to the institutional repository (UGR Repository) after the article is published.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Long, R.; Gulya, T.; Light, S.; Bali, K.; Mathesius, K.; Meyer, R.D. Sunflower Hybrid Seed Production in California. Available online: <https://escholarship.org/uc/item/14k450p6> (accessed on 11 November 2023).
2. European Commission, Directorate-General for Agriculture and Rural Development, EU Agricultural Outlook for Markets, Income and Environment 2021–2031, Publications Office of the European Union. 2021. Available online: <https://data.europa.eu/doi/10.2762/753688> (accessed on 11 November 2023).
3. Praveen, H.G.; Nagarathna, T.K.; Gayithri, M.; Patil, M.I. Genetic Variability for Seed Yield, Oil Content and Fatty Acid Composition in Germplasm Accessions of Sunflower (*Helianthus annuus* L.) and their Response to Different Seasons. *Int. J. Curr. Microbiol. Appl. Sci.* **2018**, *7*, 2120–2129. [[CrossRef](#)]
4. Salas, J.J.; Martínez-Force, E.; Harwood, J.L.; Venegas-Calderón, M.; Aznar-Moreno, J.A.; Moreno-Pérez, A.J.; Ruíz-López, N.; Serrano-Vega, M.J.; Graham, I.A.; Mullen, R.T.; et al. Biochemistry of high stearic sunflower, a new source of saturated fats. *Prog. Lipid Res.* **2014**, *55*, 30–42. [[CrossRef](#)] [[PubMed](#)]
5. Salas, J.J.; Bootello, M.A.; Garcés, R. Food Uses of Sunflower Oils. In *Sunflower Chemistry, Production, Processing, and Utilization*; Salas, J.J., Enrique, M.F., Dunford, N.T., Eds.; AOCS Press: Champaign, IL, USA, 2015; pp. 441–464. [[CrossRef](#)]
6. Harún, M. Fatty Acid Composition of Sunflower in 31 Inbred and 28 Hybrid. *Biomed. J. Sci. Technol. Res.* **2019**, *16*, 12032–12038. [[CrossRef](#)]
7. Zambelli, A.; León, A.; Garcés, R. Mutagenesis in sunflower. In *Sunflower*; American Oil Chemist's Society Press: New York, NY, USA, 2015; pp. 27–52.
8. Anushree, S.; André, M.; Guillaume, D.; Frédéric, F. Stearic sunflower oil as a sustainable and healthy alternative to palm oil. A review. *Agron. Sustain. Dev.* **2017**, *37*, 18. [[CrossRef](#)]
9. Esteki, M.; Simal-Gandara, J.; Shahsavari, Z.; Zandbaaf, S.; Dashtaki, E.; Vander Heyden, Y. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control* **2018**, *93*, 165–182. [[CrossRef](#)]
10. Jiménez-Carvelo, A.M.; Pérez-Castaño, E.; González-Casado, A.; Cuadros-Rodríguez, L. One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl trans esterified fraction. *Food Chem.* **2017**, *221*, 1784–1791. [[CrossRef](#)] [[PubMed](#)]

11. Sherma, J.; Rabel, F. A review of thin layer chromatography methods for determination of authenticity of foods and dietary supplements. *J. Liq. Chromatogr. Relat. Technol.* **2018**, *41*, 645–657. [[CrossRef](#)]
12. COI/T.20/Doc. No 26/Rev. 4; Determination of the Sterol Composition and Content and Alcoholic Compounds by Capillary Gas Chromatography. International Olive Council: Madrid, Spain, 2018.
13. Carranco, N.; Farrés-Cebrián, M.; Saurina, J.; Núñez, O. Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration. *Foods* **2018**, *7*, 44. [[CrossRef](#)]
14. COI/T.20/Doc. No 20/Rev. 4; Method of Analysis. Difference between Actual and Theoretical Content of Triacylglycerols with ECN 42. International Olive Council: Madrid, Spain, 2017.
15. COI/T.20/Doc. No 33/Rev. 4; Method of Analysis. Determination of Fatty Acid Methyl Esters by Gas Chromatography. International Olive Council: Madrid, Spain, 2017.
16. Ruiz-Samblás, C.; González-Casado, A.; Cuadros-Rodríguez, L. Triacylglycerols determination by high-temperature gas chromatography in the analysis of vegetable oils and foods: A review of the past 10 years. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 1618–1631. [[CrossRef](#)]
17. Gómez-Caravaca, A.M.; Maggio, R.M.; Cerretani, L. Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. *Anal. Chim. Acta* **2016**, *913*, 1–21. [[CrossRef](#)]
18. Casale, M.; Simonetti, R. Review: Near infrared spectroscopy for analysing olive oils. *J. Near Infrared Spectrosc.* **2014**, *22*, 59–80. [[CrossRef](#)]
19. Jiménez-Sanchidrián, C.; Ruiz, J.R. Use of Raman spectroscopy for analyzing edible vegetable oils. *Appl. Spectrosc. Rev.* **2016**, *51*, 417–430. [[CrossRef](#)]
20. Chen, J.; Zhao, Y.; Wu, R.; Yin, T.; You, J.; Hu, B.; Jia, C.; Rong, J.; Liu, R.; Zhang, B.; et al. Changes in the Quality of High-Oleic Sunflower Oil during the Frying of Shrimp (*Litopenaeus vannamei*). *Foods* **2023**, *12*, 1332. [[CrossRef](#)] [[PubMed](#)]
21. Arroyo-Cerezo, A.; Jiménez-Carvelo, A.M.; González-Casado, A.; Koidis, A.; Cuadros-Rodríguez, L. Deep (offset) non-invasive Raman spectroscopy for the evaluation of food and beverages—A review. *LWT* **2021**, *149*, 111822. [[CrossRef](#)]
22. Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic fingerprinting: An innovative approach for food ‘identification’ and food authentication—A tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. [[CrossRef](#)] [[PubMed](#)]
23. Cuadros-Rodríguez, L.; Ortega-Gavilán, F.; Martín-Torres, S.; Arroyo-Cerezo, A.; Jiménez-Carvelo, A.M. Chromatographic Fingerprinting and Food Identity/Quality: Potentials and Challenges. *J. Agric. Food Chem.* **2021**, *69*, 14428–14434. [[CrossRef](#)]
24. Szymańska, E. Modern data science for analytical chemical data—A comprehensive review. *Anal. Chim. Acta* **2018**, *1028*, 1–10. [[CrossRef](#)] [[PubMed](#)]
25. Mialon, N.; Roig, B.; Capodanno, E.; Cadiere, A. Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers. *Food Chem.* **2023**, *398*, 133856. [[CrossRef](#)]
26. Chemometrics vs Machine Learning. Available online: <https://ondalys.fr/en/scientific-resources/chemometrics-vs-machine-learning/> (accessed on 11 October 2023).
27. Bikrani, S.; Jiménez-Carvelo, A.M.; Nechar, M.; Bagur-González, M.G.; Souhail, B.; Cuadros-Rodríguez, L. Authentication of the geographical origin of margarines and fat-spread products from liquid chromatographic UV-absorption fingerprints and chemometrics. *Foods* **2019**, *8*, 588. [[CrossRef](#)]
28. Pérez-Castaño, E.; Medina-Rodríguez, S.; Bagur-González, M.G. Discrimination and classification of extra virgin olive oil using a chemometric approach based on TMS-4,4'-desmethylsterols GC(FID) fingerprints of edible vegetable oils. *Food Chem.* **2019**, *274*, 518–525. [[CrossRef](#)]
29. Ortega-Gavilán, F.; Jiménez-Carvelo, A.M.; Cuadros-Rodríguez, L.; Bagur-González, M.G. The chromatographic similarity profile—An innovative methodology to detect fraudulent blends of virgin olive oils. *J. Chromatogr. A* **2022**, *1679*, 463378. [[CrossRef](#)] [[PubMed](#)]
30. Guerrero-Chanivet, M.; Ortega-Gavilán, F.; Bagur-González, M.G.; Valcárcel-Muñoz, M.J.; García-Moreno, M.V.; Guillén-Sánchez, D.A. Pattern Recognition of GC-FID Profiles of Volatile Compounds in Brandy de Jerez Using a Chemometric Approach Based on Their Instrumental Fingerprint. *Food Bioprocess Technol.* **2023**, *16*, 1963–1975. [[CrossRef](#)]
31. Guerrero-Chanivet, M.; Ortega-Gavilán, F.; Bagur-González, M.G.; Valcárcel-Muñoz, M.J.; García-Moreno, M.V.; Guillén-Sánchez, D.A. Influence of Oak Species, Toasting Degree, and Aging Time on the Differentiation of Brandy Using a Chemometrics Approach Based on Phenolic Compound UHPLC Fingerprints. *J. Agric. Food Chem.* [[CrossRef](#)] [[PubMed](#)]
32. Ellis, D.I.; Eccles, R.; Xu, Y.; Griffen, J.; Muhamadali, H.; Matousek, P.; Goodall, I.; Goodacre, R. Through-Container, Extremely Low Concentration Detection of Multiple Chemical Markers of Counterfeit Alcohol Using a Handheld SORS Device. *Sci. Rep.* **2017**, *7*, 12082. [[CrossRef](#)] [[PubMed](#)]
33. Schorn-García, D.; Ezenarro, J.; Aceña, L.; Busto, O.; Boqué, R.; Giussani, B.; Mestres, M. Spatially Offset Raman Spectroscopic (SORS) Analysis of Wine Alcoholic Fermentation: A Preliminary Study. *Fermentation* **2023**, *9*, 115. [[CrossRef](#)]
34. Ostovar Pour, S.; Afshari, R.; Landry, J.; Pillidge, C.; Gill, H.; Blanch, E. Spatially offset Raman spectroscopy: A convenient and rapid tool to distinguish cheese made with milks from different animal species. *J. Raman Spectrosc.* **2021**, *52*, 1705–1711. [[CrossRef](#)]
35. Arroyo-Cerezo, A.; Jiménez-Carvelo, A.M.; González-Casado, A.; Ruisánchez, I.; Cuadros-Rodríguez, L. The potential of the spatially offset Raman spectroscopy (SORS) for implementing rapid and non-invasive in-situ authentication methods of plastic-packaged commodity foods—Application to sliced cheeses. *Food Control* **2023**, *146*, 109522. [[CrossRef](#)]



36. Jiménez-Carvelo, A.M.; Arroyo-Cerezo, A.; Bikrani, S.; Jia, W.; Koidis, A.; Cuadros-Rodríguez, L. Rapid and non-destructive spatially offset Raman spectroscopic analysis of packaged margarines and fat-spread products. *Microchem. J.* **2022**, *178*, 107378. [[CrossRef](#)]
37. Varnasseri, M.; Muhamadali, H.; Xu, Y.; Richardson, P.I.C.; Byrd, N.; Ellis, D.I.; Matousek, P.; Goodacre, R. Portable through Bottle SORS for the Authentication of Extra Virgin Olive Oil. *Appl. Sci.* **2021**, *11*, 8347. [[CrossRef](#)]
38. Araújo, G.A.; Azcarate, S.M.; Špánik, I.; Khvalbota, L.; Goicoechea, H.C. Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis. *TrAC Trends Anal. Chem.* **2023**, *164*, 117105. [[CrossRef](#)]
39. Durakli Velioglu, S.; Ercioglu, E.; Tumay Temiz, H.; Murat Velioglu, H.; Topcu, A.; Boyaci, I.H. Raman Spectroscopic Barcode Use for Differentiation of Vegetable Oils and Determination of Their Major Fatty Acid Composition. *J. Am. Oil Chem. Soc.* **2016**, *93*, 627–635. [[CrossRef](#)]
40. Vali Zade, S.; Forooghi, E.; Jannat, B.; Hashempour-Baltork, F.; Abdollahi, H. A combined classification modeling strategy for detection and identification of extra virgin olive oil adulteration using Raman spectroscopy. *Chemom. Intell. Lab. Syst.* **2023**, *240*, 104903. [[CrossRef](#)]
41. Zhang, L.; Li, P.; Sun, X.; Wang, X.; Xu, B.; Wang, X.; Ma, F.; Zhang, Q.; Ding, X. Classification and Adulteration Detection of Vegetable Oils Based on Fatty Acid Profiles. *J. Agric. Food Chem.* **2014**, *62*, 8745–8751. [[CrossRef](#)] [[PubMed](#)]
42. Philippidis, A.; Poulakis, E.; Papadaki, A.; Velegarakis, M. Comparative Study using Raman and Visible Spectroscopy of Cretan Extra Virgin Olive Oil Adulteration with Sunflower Oil. *Anal. Lett.* **2017**, *50*, 1182–1195. [[CrossRef](#)]
43. Ali, H.; Nawaz, H.; Saleem, M.; Nurjis, F.; Ahmed, M. Qualitative analysis of desi ghee, edible oils, and spreads using Raman spectroscopy. *J. Raman Spectrosc.* **2016**, *47*, 706–711. [[CrossRef](#)]
44. Duraipandian, S.; Petersen, J.C.; Lassen, M. Authenticity and Concentration Analysis of Extra Virgin Olive Oil Using Spontaneous Raman Spectroscopy and Multivariate Data Analysis. *Appl. Sci.* **2019**, *9*, 2433. [[CrossRef](#)]
45. Berghian-Grosan, C.; Magdas, D.A. Raman spectroscopy and machine-learning for edible oils evaluation. *Talanta* **2020**, *218*, 121176. [[CrossRef](#)]
46. Kennard, R.; Stone, L. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137–148. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.