Booksh Karl (Orcid ID: 0000-0003-1803-376X)
Lavine Barry (Orcid ID: 0000-0001-8463-5254)

Authentication of Edible Oils Using an Infrared Spectral Library and Digital Sample Sets – A Feasibility Study

Isio Sota Uba[1], Collin G. White[1], Karl Booksh[2], and Barry K. Lavine[1]

[1]Department of Chemistry, Oklahoma State University, Stillwater, OK 74078

[2]Department of Chemistry & Biochemistry, University of Delaware, Newark, DE 19716

Abstract

A potential method to determine whether two varieties of edible oils can be differentiated by Fourier transform infrared (FTIR) spectroscopy is proposed using digitally generated data of adulterated edible oils from an infrared (IR) spectral library. The first step is the evaluation of digitally blended data sets. Specifically, IR spectra of adulterated edible oils are computed from digitally blending experimental data of the IR spectra of an edible oil and the corresponding adulterant using the appropriate mixing coefficients to achieve the desired level of adulteration. To determine whether two edible oils can be differentiated by FTIR spectroscopy, pure IR spectra of the two edible oils are compared to IR spectra of two edible oils digitally mixed using a genetic algorithm for pattern recognition to solve a ternary classification problem. If the IR spectra of the two edible oils and their binary mixtures are differentiable from principal component plots of the spectral data, then differences between the IR spectra of these two edible oils are of sufficient magnitude to ensure that a reliable classification by FTIR spectroscopy can be obtained. Using this approach, the feasibility of authenticating edible oils such as extra virgin olive oil (EVOO) directly from library spectra is demonstrated. For this study, both digital and experimental data are combined to generate training and validation data sets to assess detection limits in FTIR spectroscopy for the adulterants.

Keywords: edible oils, adulteration, genetic algorithms, variable selection, classification, mixture analysis

Introduction

Edible oils are an important component of the human diet due to their high nutritional value serving as a major source of fatty acids and fat-soluble vitamins in many diets.[1,2] These oils are primarily composed of triglycerides which contain saturated, monounsaturated and polyunsaturated fatty acids.[3] The relative quantity of each fatty acid is related to the specific variety of the edible oil.[4] For example, safflower oil has more polyunsaturated fatty acids than extra virgin olive oil. Edible oils are used in cooking and are also ingredients in many preprocessed foods because of their sensory characteristics.

Adulteration of edible oils is an important chemical analysis problem as the most frequently adulterated food is extra virgin olive oil (EVOO).[5] Adulteration of a more expensive edible oil by either substitution or blending with less expensive cooking oils is of concern to government and regulatory officials. Adulterated EVOO, which does not meet the International Olive Council's standards for the composition of monounsaturated fatty acids, free fatty acids, trans fatty acids, peroxides, and esterified fatty acids[6], cannot be detected by either the consumer or retailer because the adulterated cooking oil is often comparable in appearance and flavor to EVOO. In addition, adulteration of EVOO by less expensive edible oils such as peanut oil (which contains allergens) poses a serious health risk. The successful classification of edible oils by their variety (e.g., discrimination of EVOO from peanut oil) is a crucial first step in solving this problem.

Analysis of edible oils for purposes of discrimination or authentication is often performed using separation techniques such as gas chromatography/mass spectrometry (GC/MS) or liquid chromatography/mass spectrometry (LC/MS)[7, 8]. However, GC/MS and LC/MS are time consuming, expensive, and labor intensive. For this reason, there is interest in using Fourier transform infrared (FTIR) spectroscopy to authenticate edible oils. FTIR spectroscopy is fast, sample preparation is simple, and infrared (IR) analysis (unlike GC/MS

and LC/MS) can be performed in the field. Although an edible oil contains hundreds of constituents, an IR spectrum of an edible oil sample can serve as a chemical fingerprint[9]. However, analysis of these fingerprints by pattern recognition methods is necessary to extract information from the IR spectrum about the variety of the edible oil[10-13]. In these studies, principal component analysis was used to discriminate known edible oils and to classify unknown edible oils as well as to detect the presence of adulterants in edible oils.

In a previous study[14], pattern recognition methods were applied to the infrared (IR) spectra of ninety-seven edible oil samples from twenty plant-based varieties collected over a three-year period. The ninety-seven edible oil samples selected for this study encompassed multiple brands and manufacturers representing supplier to supplier variation and seasonal and batch variation within a supplier. Using a hierarchical classification scheme, the twenty plant-based varieties of edible oils were divided into four distinct groups. Edible oils from different oil groups could be reliably discriminated, whereas the discrimination of edible oils within the same group was problematic. Adulteration of the plant-based edible oils by other oils in the same group (e.g., EVOO by almond oil) could not be reliably detected using Fourier transform infrared (FTIR) spectroscopy, whereas adulteration of edible oils by other oils that were not part of the same oil group (e.g., EVOO adulterated by corn or canola oil) could be detected at concentration levels as low as 10% (v/v) which was consistent with previously published studies using partial least squares regression. A unique aspect of this study was the incorporation of edible oils collected systematically over three years, which introduced a heretofore unseen variability in the chemical composition of the edible oils. This work also demonstrated that previously published studies[15-18] (which relied on a single sample or brand to represent each variety of edible oil) provide an overly optimistic estimate of the capability of FTIR spectroscopy to discriminate plant based edible oils by variety as well as detect the presence of adulterants in edible oils.

In this study, a potential method to determine whether two varieties of edible oils can be differentiated is proposed using digitally generated data of adulterated edible oils from an IR spectral library. The first step is the evaluation of digitally blended data sets. Specifically, IR spectra of adulterated edible oils are computed from digitally blended experimental data of the IR spectra of an edible oil and the corresponding adulterant using the appropriate mixing coefficients for the computed spectra to achieve the desired level of adulteration. To determine whether two edible oils can be differentiated by FTIR spectroscopy, pure IR spectra of the two edible oils are compared to IR spectra of the two edible oils digitally mixed using pattern recognition techniques to solve a ternary classification problem. If the IR spectra of the two edible oils and their binary mixtures are differentiable, then differences between the IR spectra of these two edible oils are of sufficient magnitude to ensure that discrimination of these two edible oils can be achieved by FTIR spectroscopy. Using this approach, the feasibility of authenticating edible oils such as EVOO directly from IR library spectra has been demonstrated. For this study, both digital and experimental data were combined to generate training and validation data sets to assess detection limits for the adulterant

Experimental

A spectral database of 3720 IR spectra of both pure and adulterated edible oils was collected using an iS50 Thermo-Nicolet FTIR spectrometer equipped with a diamond ATR accessory and a DTGS detector. The pure edible oil samples (99 in total) comprising the library spanned 20 distinct plant-based edible oil varieties (see Table 1). This sample cohort was obtained from supermarkets in the greater metropolitan Newark, DE area over a three-year period to account for brand, lot, year, storage and seasonal variability for a particular manufacturer. To further characterize these edible oils, the peroxide value of each edible oil sample was measured using a Milwaukee Lab Mi490 Photometer (Milwaukee Instruments,

Rocky Mount, NC). Table 1 lists the twenty oil types that comprise the 99 samples collected, the number of samples collected for each edible oil and the number of spectra per edible oil. For each resolution, a total of 377 IR spectra were collected with 1508 IR spectra collected in total for the pure edible oils. A representative IR spectrum of an edible oil from the library is shown in Figure 1.

Each pure and adulterated edible oil sample was analyzed at 4 cm^{-1}, 6 cm^{-1}, 8 cm^{-1}, and 16 cm^{-1} resolution. The adulterated samples in the library were prepared by mixing EVOO, extra light olive oil (ELOO) or sesame oil with less expensive edible oils (corn, canola, almond, peanut, sunflower, hazelnut, grapeseed, safflower, and vegetable) using a digital pipette to prepare adulterated mixtures by v/v in known amounts from 5% to 90%. For example, a 10% adulterated mixture of EVOO with corn oil as the adulterant was prepared by mixing 900 μL of extra virgin olive oil and 100 μL of corn oil in a 15 mL sterile falcon tube using a Thermolyne MaxiMixPlus vortex mixer. 416 IR spectra of EVOO, ELOO, or sesame oil adulterated by corn oil, canola oil, almond, peanut, sunflower, hazelnut, grapeseed, safflower, and vegetable oils were also collected at 4 cm^{1}, 6 cm^{-1}, 8 cm^{-1}, and 16 cm^{-1} resolution for a total of 1664 IR spectra (see Table 2). Ternary mixtures (which consist of two adulterants added to EVOO, ELOO, or sesame oil) were also prepared. The adulterants used in the ternary mixtures included corn, canola, almond, hazelnut, vegetable, grapeseed and safflower oils. 128 FTIR spectra of the ternary mixtures were also collected at 4 cm^{-1}, 6 cm^{-1}, 8 cm^{-1}, and 16 cm^{-1} resolution for a total of 512 spectra (see Table S1). The FTIR IR spectra (4000 cm^{-1} to 400 cm^{-1}) of each pure and adulterated edible oil sample were collected in triplicate, quadruplicate, or quintuplicate, each at 64 scans. All FTIR spectra in the database were baseline corrected using OMNIC and normalized to unit length with MATLAB (MathWorks, Natick, MA). Apodization of the spectra was performed using OMNIC (Thermo-Nicolet) and the Happ-Genzel function.

Preparation of Digitally Blended Spectra from IR Spectra of Edible Oils

Digital blending refers to the proportion of each edible oil that comprises the digitally simulated adulterated oil mixtures. Digital blending was performed on unprocessed IR spectra. To obtain a digital blend representing an 80% EVOO and 20% corn oil mixture, the IR spectrum of an EVOO sample is multiplied by 0.8 and added to an IR spectrum of a corn oil sample that is multiplied by 0.2. Gaussian distributed noise is then added to the IR spectrum of each digital blend to homogenize the spectral data. For each spectrum, noise is only added to the regions which contain IR bands (402 cm^{-1} to 1525 cm^{-1}, 1600 cm^{-1} to 1850 cm^{-1} and 2750 cm^{-1} to 3150 cm^{-1}). For a training set of digitally blended IR spectra, the largest absorbance value at each wavelength is identified, and one thousandth of this value is multiplied by Gaussian distributed random noise which has a mean of zero and standard deviation of one. If the largest absorbance value is less than or equal to zero, noise is not added to the blended spectrum at that particular wavelength. For the pattern recognition studies that were undertaken to demonstrate equivalency between real data and digitally blended data, the full spectral range (4000 cm^{-1} to 400 cm^{-1}) was employed.

Genetic Algorithm for Pattern Recognition Analysis

Wavelengths characteristic of the variety of the edible oil were identified using a genetic algorithm for pattern recognition, which takes advantage of both supervised and unsupervised learning to identify spectral features that optimize the separation of the spectra by edible oil type in a plot of the two or three largest principal components of the data[19-24]. Because principal components (PCs) maximize variance, the bulk of the information encoded by the wavelengths selected by the pattern recognition GA was about the differences between the assembly plants. A principal component (PC) plot that shows separation of the data by assembly plot can only be generated using wavelengths whose variance or information is primarily about the differences between these assembly plants. Thus, the fitness function of

the pattern recognition dramatically reduces the size of the search space as it limits the search to these types of wavelengths. In addition, the pattern recognition GA was able to focus on those classes and/or samples that were difficult to classify by boosting the weights of the samples or classes that were consistently misclassified. Over time, the algorithm learns its optimal parameters in a manner similar to a neural network. The pattern recognition GA integrates aspects of artificial intelligence and evolutionary computations to yield a "smart" one-pass procedure for wavelength selection and pattern classification. Further details about the configuration of the pattern recognition GA including the reproduction and mutation operators can be found elsewhere[25-28].

For this study, the fitness function of the pattern recognition GA was modified to allow for incorporation of model inference into the variable selection process. The goal was to identify variables that minimize the error across the entire model (PC plot of the data). This was accomplished by assessing the uncertainty of the sample scores in the principal component plot using the jack-knife[29] to generate estimates of dispersion. During each generation, the fitness function of the pattern recognition GA evaluates thousands of principal component plots, one for each feature subset (i.e., chromosome) in the population of solutions. For each principal component score plot, the corresponding training set samples are removed one at a time, and the score matrix and loading matrix for the resampled (i.e., jackknifed) training set is recomputed. Due to the rotational ambiguities of PCA, the loading matrix for each resampled training set must be rotated using a Procrustean rotation[30] to match the loading matrix associated with the score plot containing all the samples. For each training set sample, scores across all leave-one-out score plots are projected onto the original principal component plot of the feature subset which is then scored using the fitness function. Thus, information about the level of confidence in the classification of each training set sample is directly incorporated into the variable selection process with the jack-knifed scores for each

sample effectively comprising an error cloud to depict the uncertainty associated with each training set sample.

Results and Discussion

To determine whether two edible oils can be differentiated by FTIR spectroscopy, the IR spectra of the pure edible oils was compared to the IR spectra of their mixtures (which simulate an adulterated edible oil). The focus of these studies was EVOO (which is frequently a target of adulteration), with each comparison formulated as a three-way classification problem: EVOO, adulterated EVOO and adulterant (corn oil, canola oil or almond oil). The EVOO-adulterant mixtures used for the training and validation set for both the experimental and digitally blended data span a large concentration range. For each comparison, it is assumed that the IR spectra of the adulterated mixture can be represented by the IR spectra of EVOO and the other edible oil (adulterant), with the weights of the constituents defining the mixing proportion of each edible oil comprising the mixture. If the IR spectra of EVOO and the adulterant adhere to a linear mixture model then the results of the three-way classification study for both the experimental and blended data would be similar.

To identify the wavelengths in each three-way classification problem that convey information about the degree of adulteration for both the experimental and digitally blended data, the pattern recognition GA was applied to each of these data sets. Each data set (EVOO/Corn, EVOO/Canola, and EVOO/Almond) was autoscaled and analyzed by the pattern recognition GA using the same set of parameters (number of chromosomes, selection pressure, configuration of initial population, and K_c). Figures 2 and 3 show the plots of the two largest principal components of the 118 FTIR spectra (see Table 3) and the 8 and 17 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, EVOO-corn oil mixtures, and corn oil. EVOO, corn oil, and the binary

mixtures of EVOO-corn oil cluster in separate regions of the PC plot for both the experimental data (see Figure 2) and digitally blended data (see Figure 3). The first principal component appears to be correlated to the amount of adulterant (i.e., corn oil) in each sample.

The predictive ability of the 8 and 17 spectral features identified by the pattern recognition GA was assessed using an external prediction set of 12 EVOO-corn oil spectral mixtures whose composition varied from 0% to 40% corn oil. The plant-based edible oil samples used to prepare the EVOO-corn oil mixtures comprising the prediction set were excluded from the training set. Figures 4 and 5 show the plots of the 12 prediction set spectra projected onto the principal component score plot of the 118 IR spectra comprising the training set and the 8 and 17 spectral features identified by the pattern recognition GA. All 12 FTIR spectra in the prediction set for both the experimental data (see Figure 4) and digitally blended data (see Figure 5) were correctly classified as each spectrum is located in a region of the principal component score plot that contain samples tagged with the same class label. Clearly, EVOO can be differentiated from corn oil. The detection limit for corn oil in EVOO from the principal component score plot is approximately 10% for both the experimental data and digitally blended data which agrees with the detection limits previously reported for corn oil using PLS[31]. Furthermore, the agreement between the results obtained for the experimental data and digitally blended data suggests that digitally blended data can be used to assess whether two varieties of edible oils (e.g., EVOO versus corn oil) can be differentiated by FTIR spectroscopy through a ternary classification study.

Figures 6 and 7 show the plots of the two largest principal components of the 115 IR spectra comprising the training set (see Table 4) and the 9 and 11 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, EVOO-canola oil mixtures, and canola oil. EVOO, canola oil, and the binary mixtures of EVOO-canola oil cluster in separate regions of the PC plot for both the experimental data (Figure 6)

and digitally blended data (see Figure 7). Again, the first principal component appears to be correlated to the amount of adulterant (i.e., canola oil) in each sample. The discriminating relationship developed from these 9 and 11 spectral features was successfully validated using the 15 FTIR spectra comprising the prediction set for both the experimental and digitally blended data (see Figures 6 and 7). The plant-based edible oil samples used to prepare the adulterated EVOO mixtures comprising the prediction set were again excluded from the training set. The detection limit for canola oil in EVOO from the principal component score plot of the FTIR spectra is again 10% in agreement with PLS[32].

Figures 8 and 9 show principal component score plots of the 100 IR spectra comprising the training set (see Table 5) and the 5 and 25 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, EVOO-almond oil, and almond oil. The first principal component does not appear to be well correlated to the amount of almond oil in the mixtures. Furthermore, several EVOO-almond oil samples in the training and prediction set for both the experimental data (see Figure 8) and digitally blended data (see Figure 9) are not correctly classified. The absence of spectral features in the experimental and digitally blended data that can differentiate EVOO from EVOO adulterated with almond oil (as well as the first principal component being only weakly correlated to the amount of almond oil in the samples) would indicate that EVOO and almond oil have similar IR spectra and (thus) would be difficult to discriminate by FTIR. Furthermore, detecting adulteration of EVOO by almond oil would be problematic. The agreement between the results for the experimental data (see Figure 8) and digitally blended data (see Figure 9) for EVOO-corn, EVOO-canola, and EVOO-almond oil indicates that one can differentiate two edible oils by sample type (i.e., variety) using FTIR spectroscopy if the pure IR spectra of the two edible oils and their digitally blended spectra can be discriminated in a ternary classification problem. For this comparison, it is crucial that samples representing each

edible oil account for seasonal and batch variations within each supplier as well as variations between suppliers to avoid obtaining overly optimistic results.

Conclusions

In this study, a basic methodology for assessing the suitability of discriminating two edible oils from their FTIR spectra is described. The FTIR spectra of the pure edible oils are compared to the FTIR spectra of their mixtures using library spectra and digitally blended data. Each comparison is formulated as a three-way classification problem. For each comparison, it is assumed that the IR spectra of the mixture can be represented by the IR spectra of the two edible oils in question, with the weights of the constituents defining the mixing proportion of each edible oil that comprises the mixture. If the IR spectra of the two edible oils and their digitally blended mixtures are differentiable, then differences between the IR spectra of these two edible oils are of sufficient magnitude to ensure that a reliable classification of these two edible oils by FTIR spectroscopy can be obtained. Using this approach, the feasibility of authenticating edible oils such as EVOO directly from library spectra has been demonstrated.

The FTIR spectral library described in this study is a flexible platform as it allows chemists to test new data analysis methodologies. Experimental designs can be constructed using similar edible oils (e.g., EVOO and sunflower oil) or oils with relatively distinct spectra (e.g., EVOO and corn oil). One can progress from simple classifications of mixtures (e.g., extra virgin olive oil that is adulterated with corn oil), quantitative mixture analysis (relative concentrations of adulterants in edible oils) to quantitative determinations of intrinsic properties of edible oils (e.g., peroxide number to assess rancidity). The effect of spectral resolution on the outcome of a classification or calibration can also be investigated using this spectral library.

Acknowledgements

BKL, ISU, and CGW acknowledge the financial support of the National Science Foundation under awards CHE-1506915 and CHE-2003867. KB acknowledges the financial support of the National Science Foundation under award CHE-1506853.

References

1. Patel A, Rova U, Christakopoulos P, Matsakas L, Introduction to Essential Fatty Acids, John Wiley & Sons, Hoboken, NJ, 2022, pp. 1-22.
2. Orsavova J, Misurcova L, Ambrozova JV, Vicha R, Mlcek J. Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids. *Int. J. Mol. Sci*. 2015: 16 (6):12871–12890.
3. Lichtenstein AH, Fats and Oils in Encyclopedia of Human Nutrition, 3rd edition, B. Caballero (Ed.) Academic Press, Waltham, MA 2013: pp. 201-208.
4. Chowdhury K, Banu LA, Khan SA, Latif A. Studies on the fatty acid composition of edible oil, *Bangladesh J. Sci. Ind. Res.* 2007; 42: 311-316.
5. Spink J, Moyer DC. Defining the public health threat of food fraud. *J. Food Sci.* 2011; 76 (9): R157-R163.
6. Moore JC, Spink J, Lipp M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *J. Food Sci.* 2012; 77(4): R118-R126.
7. Andrikopoulos NK, Giannakis IG, Tzamtzis V. Analysis of olive oil and seed oil triglycerides by capillary gas chromatography as a tool for the detection of the adulteration of olive oil. *J. Chromat. Sci.* 2001**;** *39* (4): 137-145.
8. Aparicio R., Aparicio-Ruíz R. Authentication of vegetable oils by chromatographic techniques. *J. Chromat. A.* 2000; *881* (1): 93-104.
9. Karoui R, Downey G, Blecker C. Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships-a review. *Chem. Rev.*2010**;** 110 (10): 6144-6168.
10. Safwan M, Obeidat MS, Obeidat KW. Classification of edible oils and uncovering adulteration of virgin olive oil using FTIR with the aid of chemometrics. *AJBAS.* 2009**;** *3* (3), 2048-2053.
11. Moore JC, Lipp M, Griffiths JC Preventing the adulteration of food protein with better analytical methods: Avoiding the "next melamine". *INFORM - International News on Fats, Oils and Related Materials* 2011**;** *22* (6): 373-375.
12. de la Mata P, Dominguez-Vidal A, Bosque-Sendra JM, Ruiz-Medina A, Cuadros-Rodríguez L, Ayora-Cañada MJ. Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. *Food Control.* 2012; *23* (2): 449-455.
13. Zhang Q, Liu C, Sun Z, Hu X, Shen Q, Wu J. Authentication of edible vegetable oils adulterated with used frying oil by Fourier Transform Infrared Spectroscopy. *Food Chem.* 2012; *132* (3):1607-1613.
14. Sota-Uba I, Bamidele M, Moulton JT, Booksh K, Lavine BK. Authentication of edible oils using Fourier transform infrared spectroscopy and pattern recognition methods. *Chemom. Intellig. Lab. Syst.* 2021; 210: 104251.

15. Rohman A, Che Man YB. The use of Fourier transform mid infrared (FT-MIR) spectroscopy for detection and quantification of adulteration in virgin coconut oil. *Food Chem.* 2011; 129 (2): 583-588.
16. Rohman A, Man YBC. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Int. Food Res. J.* 2010; *43* (3), 886-892.
17. Lerma-García MJ, Ramis-Ramos G, Herrero-Martínez JM, Simó-Alfonso EF. Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chem.* 2010; *118* (1): 78-83.
18. Tay A, Singh RK, S, Krishnan S, Gore JP, Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. *LWT - J. Food Sci. Technol.* 2002; *35* (1): 99-103.
19. Lavine BK, White CG, Ding T, Gaye MM, Clemmer DE. Wavelet based classification of MALDI-IMS-MS spectra of serum N-linked glycans from normal controls and patients diagnosed with Barrett's esophagus, high Grade dysplasia, and esophageal adenocarcinoma. *Chemometr Intell Lab Syst*. 2018; 176: 74 - 81**.**
20. Perera UDN, Nishikida K, Lavine BK. Development of infrared library search prefilters for automotive clear coats from simulated ATR Spectra". *Appl. Spectrosc*., 2018; 186: 662-669.
21. Lavine, B. K.; Mirjankar, N; Delwiche, S. Classification of the waxy condition of durum wheat by near infrared reflectance spectroscopy using wavelets and a genetic algorithm. *Microchem. J*. 2014; 117, 178-182.
22. Lavine BK, Davidson CE, Moores AJ. Genetic algorithms for spectral pattern recognition. *Vib. Spec*. 2002; 28(1): 83-95.
23. Lavine BK, Moores AJ. Genetic algorithms for pattern recognition analysis and fusion of sensor data in Pattern recognition, chemometrics and imaging of optical environmental monitoring. Siddiqui A, Eastwood D (Editors), Proceedings of SPIES. 1999; 103-112.
24. Lavine BK, Davidson CE, Moores AJ, Griffiths PR. Raman spectroscopy and genetic algorithms for the classification of wood types. *Appl. Spectros*. 2001; 55(8): 960 - 966.
25. Lavine BK. Davidson CE, Rayens WT. Machine learning based pattern recognition applied to microarray data. *Comb. Chem. High T. Screen*. 2004; 7: 115-131.
26. Lavine BK, Nuguru K, Mirjankar N. One stop shopping - feature selection, classification, and prediction in a single step. *J. Chemomem*. 2011; 25: 116-129
27. Lavine BK, White CG, Davidson CE. Genetic algorithms for variable selection and pattern recognition in S. D. Brown, R. Tauler, and B. Walczak (Editors) Comprehensive Chemometrics, Volume 3, Elsevier, 2nd edition, 2020.
28. Lavine BK, Davidson CE, Moores AJ, Griffiths PR. Raman spectroscopy and genetic algorithms for the classification of wood types. *Appl. Spec*. 2001; 55(8): 960 - 966.
29. Meloun M, Militky J, Forina M. Chemometrics for analytical chemistry. Volume 1: PC Aided Statistical Data Analysis, Ellis Horwood, NY, 1992, pp. 138-139.
30. Vandeginste BGM, Massart DL, Buydens LMC, DeJong S, Lewi PJ, Smeyers-Verbeke J, Handbook of Chemometrics and Qualimetrics. Elsevier, Amsterdam, Netherlands, 1998, pp. 310.
31. Vlachos N, Skopelitis Y, Psaroudaki M, Konstantinidou V, Chatzilazarou A, Tegou E. Applications of Fourier transform-infrared spectroscopy to edible oils. *Anal. Chim. Acta***.** 2006**;** *573-574*: 459-465.
32. Gurdeniz G, Ozen B. Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem.* 2009; 116, 519 -525

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

Table 1. Pure Edible Oils in the Spectral Library

Pure edible oil	Oil Type ID	Number of samples	Number of spectra
Extra virgin olive oil	1	26	83
Extra light olive oil	2	8	27
Olive oil	3	8	26
Avocado oil	5	2	9
Peanut oil	6	4	19
Corn oil	7	9	42
Grapeseed oil	8	9	36
Safflower oil	9	2	9
Hazelnut oil	10	2	9
Canola oil	13	9	36
Canola-vegetable blend	16	1	3
Vegetable oil	17	4	14
Canola-Sunflower-soybean blend	18	1	9
Sunflower oil	19	1	3
Sweet almond oil	23	2	6
Almond oil	27	4	15
Extra virgin sesame oil	28	3	15
Toasted sesame oil	32	1	3
Walnut oil	33	2	10
Avocado-olive-flaxseed blend	34	1	3
		99	377

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

Table 2. Adulterated Edible Oils in the Spectral Library

Binary Mixtures	Oil Type ID	Number of spectra
ELOO-corn mixture	40	45
EVOO-corn mixture	44	60
EVOO-peanut mixture	45	24
Sesame-sunflower mixture	47	24
Sesame-canola mixture	48	21
Sesame-corn mixture	49	26
EVOO-almond mixture	51	36
Sesame-grapeseed mixture	55	18
ELOO-hazelnut mixture	54	18
Sesame-vegetable mixture	58	18
EVOO-canola mixture	60	78
ELOO-canola mixture	61	24
ELOO-safflower mixture	62	24
		416

Table 3. Training and Prediction Sets for Experimental and Blended Data

	Number of spectra in training set/prediction set
EVOO	73/0
Corn	33/0
EVOO-corn	12/12
Total	118/12

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

Table 4. Training and Prediction Sets for Experimental and Blended Data

	Number of spectra in training set/prediction set
EVOO	73/0
Canola	27/0
EVOO-canola	15/15
Total	115/15

Table 5. Training and Prediction Set for Experimental and Blended Data

	Number of spectra in training set/prediction set
EVOO	73/0
Almond	12/0
EVOO-almond	15/15
Total	100/15

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

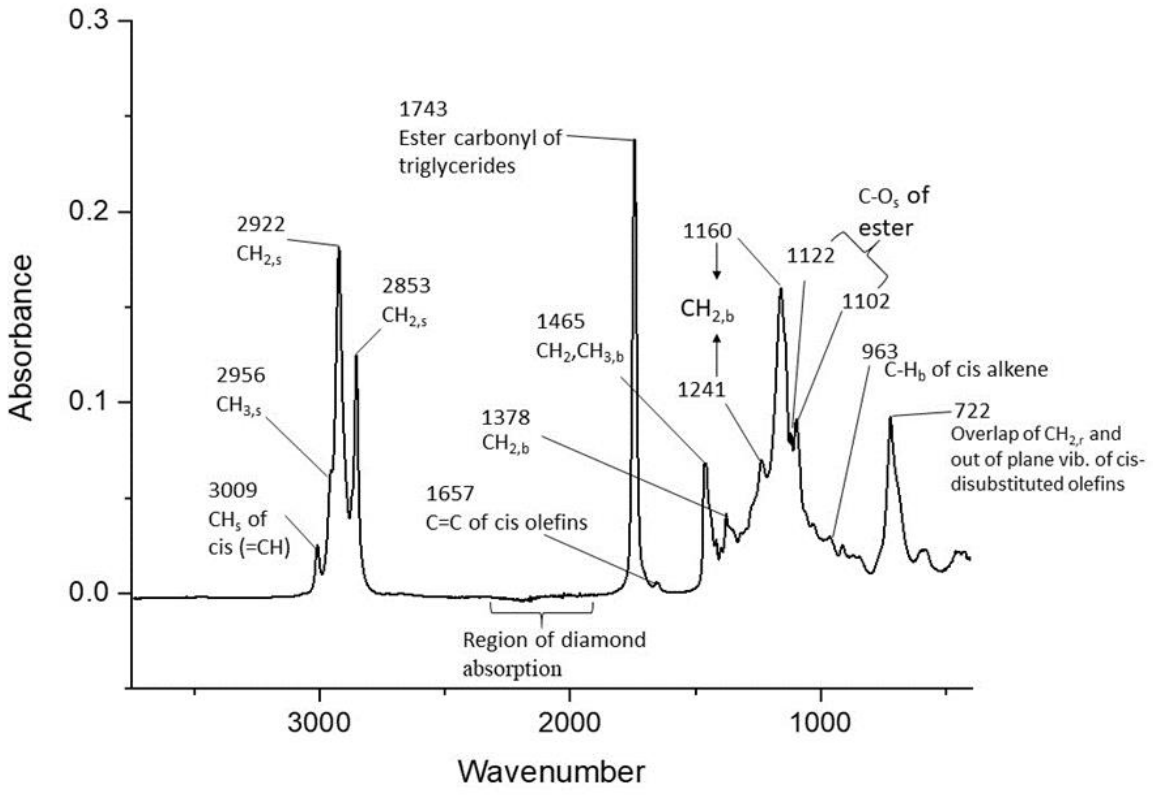

Figure 1. A representative IR spectrum of an edible oil (corn oil) from the IR spectral library is shown. The IR spectrum was collected using an iS50 Thermo-Nicolet FTIR spectrometer equipped with a diamond ATR accessory and a DTGS detector.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

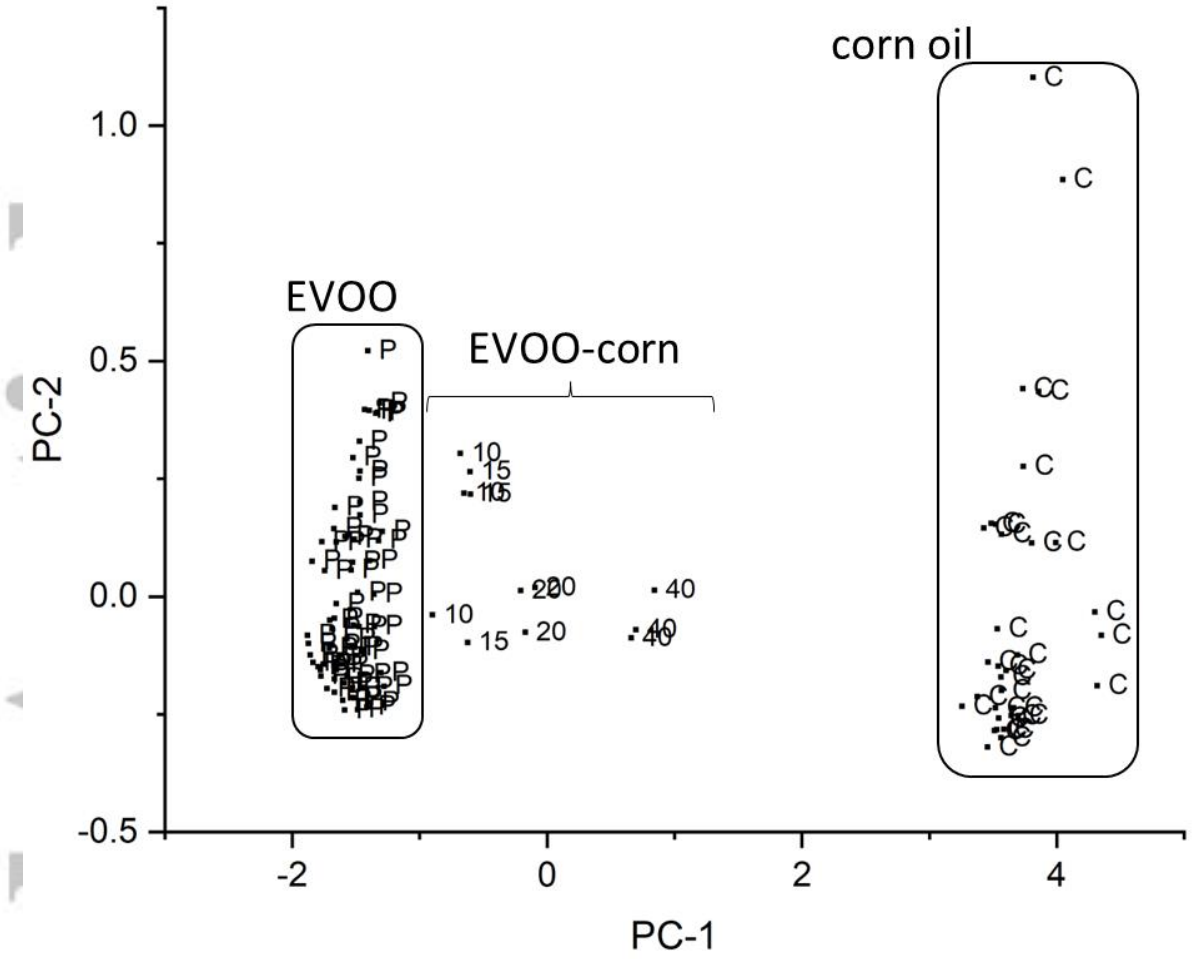

Figure 2. Plot of the two largest principal components of the 118 IR training set spectra (black) and the 8 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, corn oil and EVOO-corn oil mixtures (10% corn oil to 40% oil). The total cumulative variance explained by the two largest principal components for the experimental data is 98.57%. P = EVOO, C = corn oil, 10 = 10% corn oil, 15 = 15% corn oil, 20 = 20% corn oil, and 40 = 40% corn oil.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

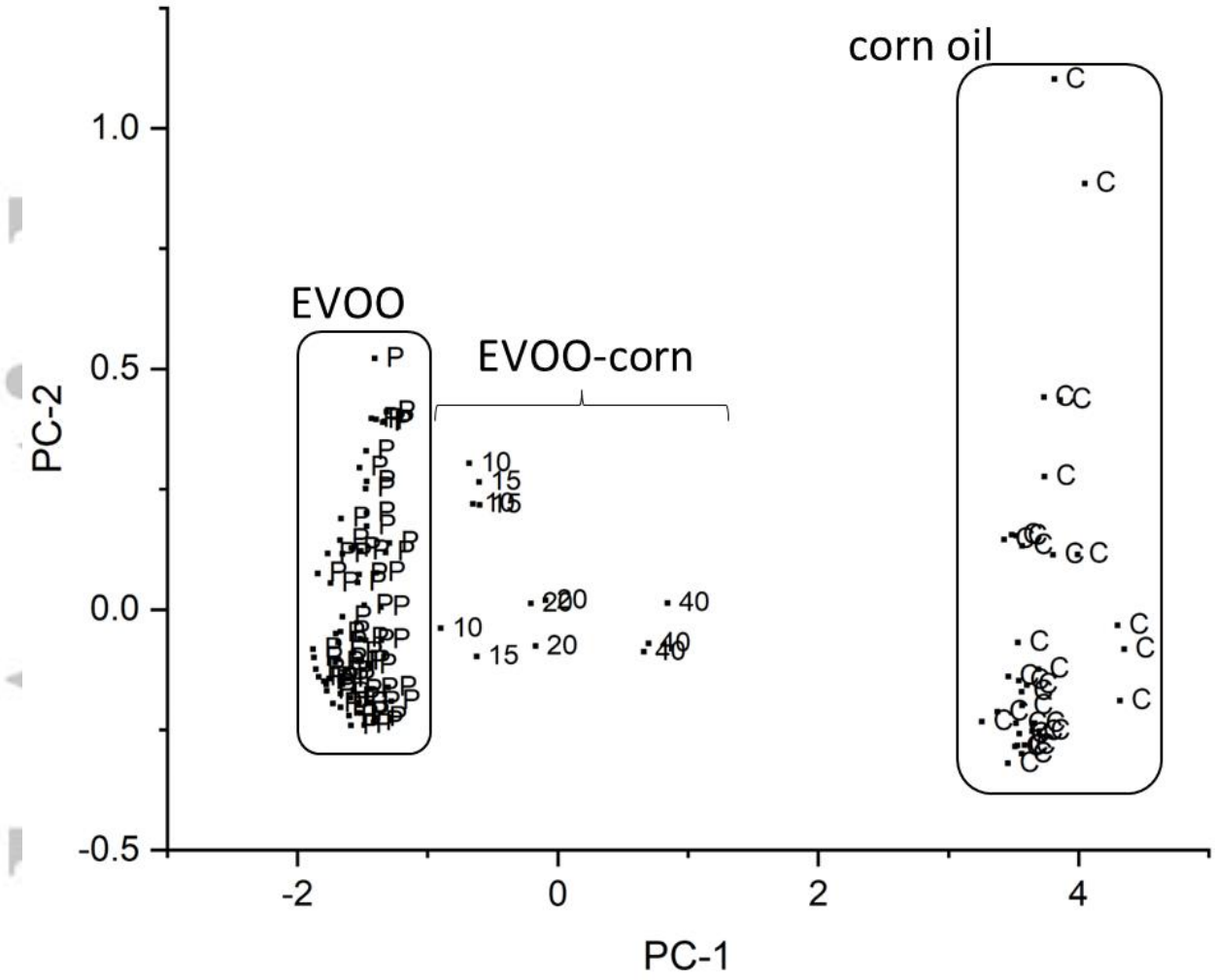

Figure 3. Plot of the two largest principal components of the 118 IR training set spectra (black) and the 17 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, corn oil and EVOO-corn oil mixtures. The total cumulative variance explained by the two largest principal components for the digitally blended data is 97.18%. P = EVOO, C = corn oil, 10 = 10% corn oil, 15 = 15% corn oil, 20 = 20% corn oil, and 40 = 40% corn oil.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

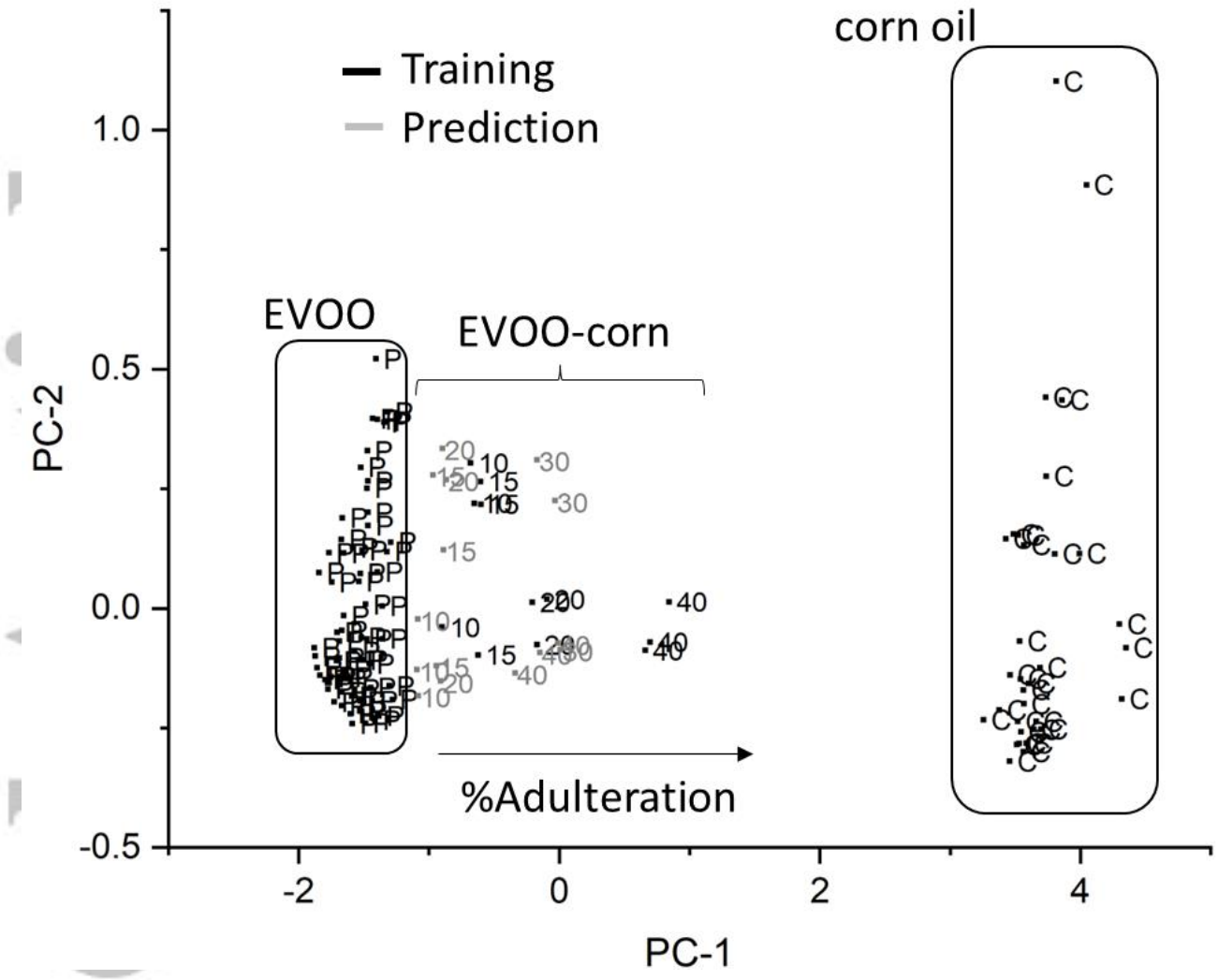

Figure 4. Projection of the 12 prediction set spectra (grey) onto the PC-plot developed from the 118 training set spectra and 8 features identified by the pattern recognition GA for the experimental data. P = EVOO, C = corn oil, 10 = 10% corn oil, 15 = 15% corn oil, 20 = 20% corn oil, and 40 = 40% corn oil.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

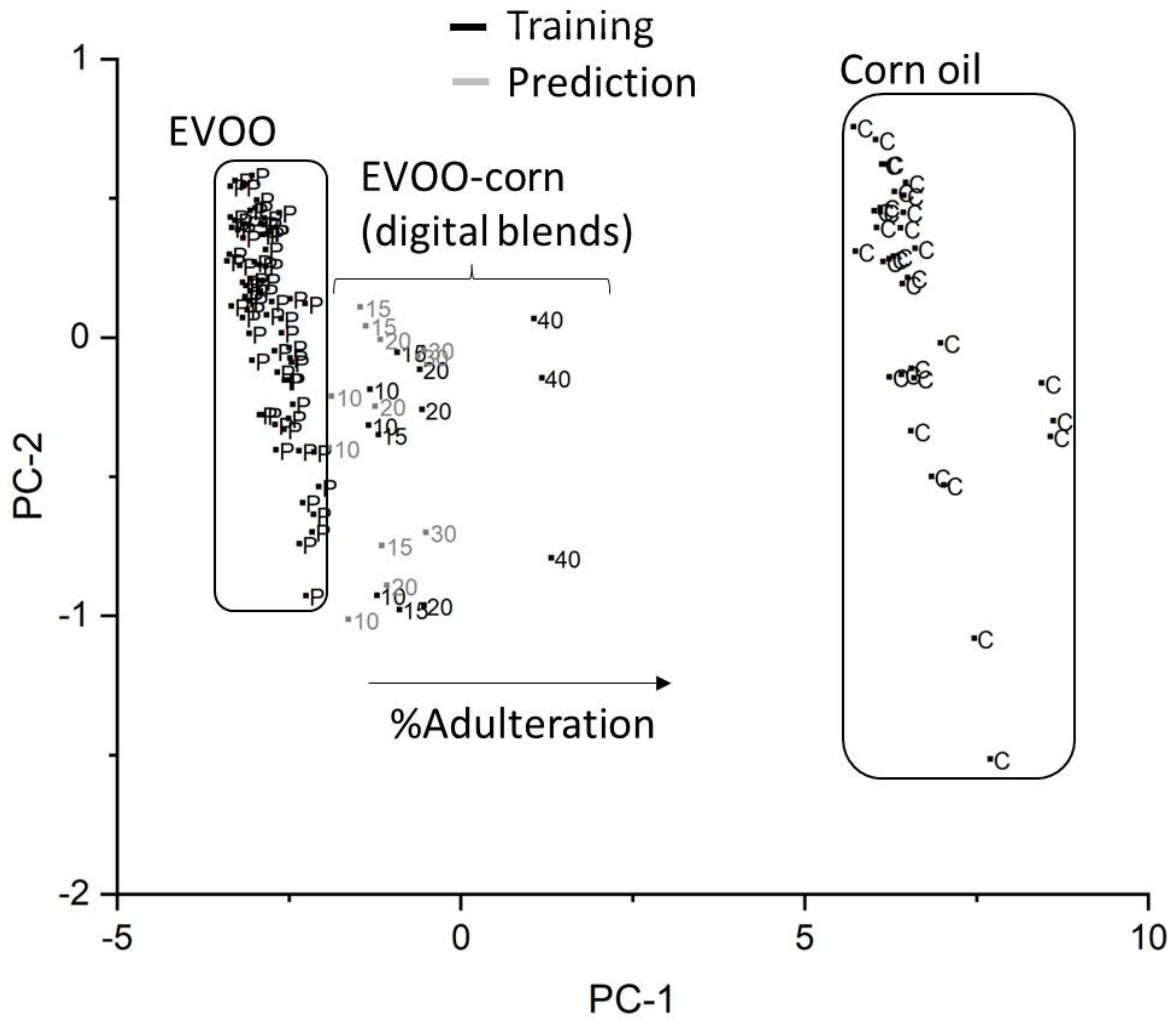

Figure 5. Projection of the 12 prediction set spectra (grey) onto the PC-plot developed from the 118 training set spectra and 17 features identified by the pattern recognition GA for the digitally blended data. P = EVOO, C = corn oil, 10 = 10% corn oil, 15 = 15% corn oil, 20 = 20% corn oil, and 40 = 40% corn oil.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

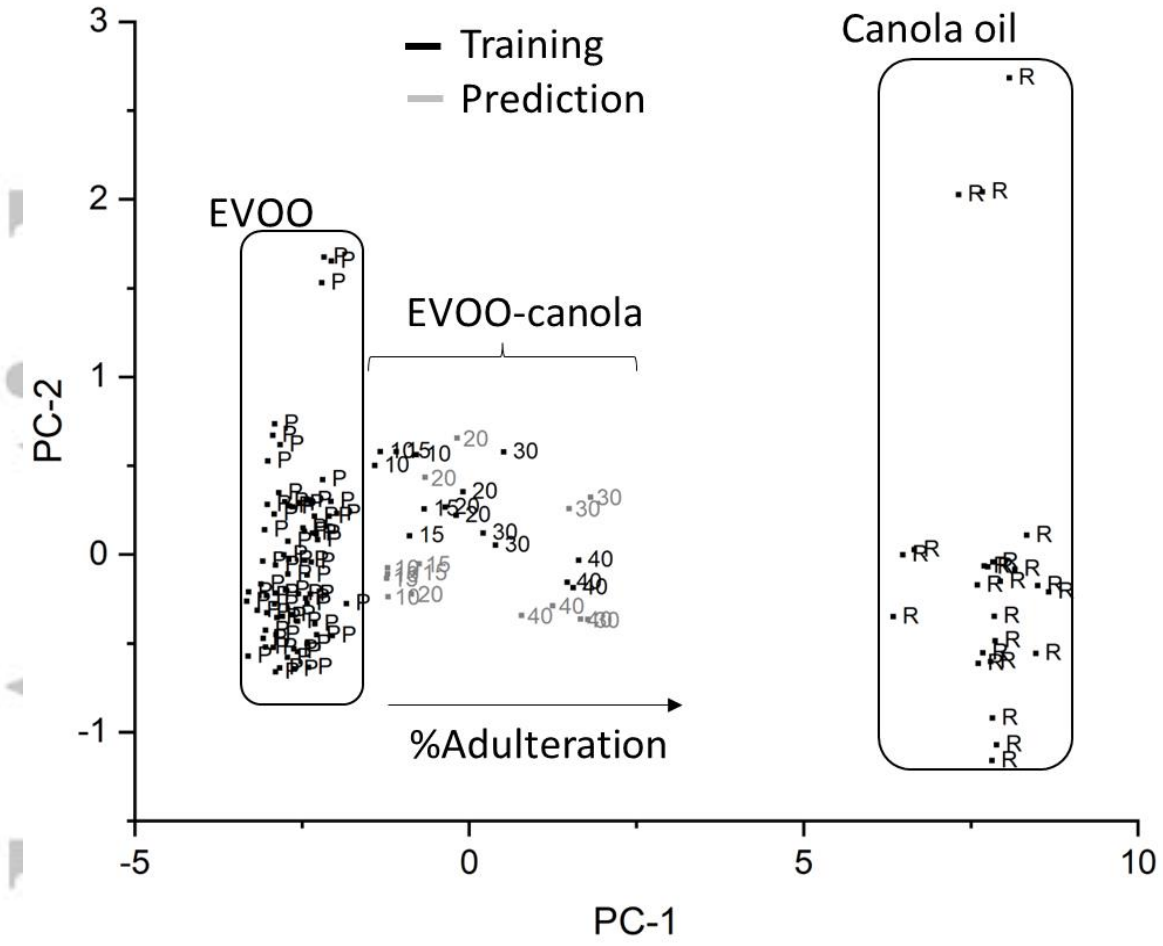

Figure 6. Plot of the two largest principal components of the 115 IR training set spectra (black) and the 9 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, canola oil and EVOO-canola oil mixtures (10% canola oil to 40% oil). The total cumulative variance explained by the two largest principal components for the experimental data is 96.82%. The prediction set spectra are represented in grey. P = EVOO, R = canola oil, 10 = 10% canola oil, 15 = 15% canola oil, 20 = 20% canola oil, 30 = 30% canola oil and 40 = 40% canola.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

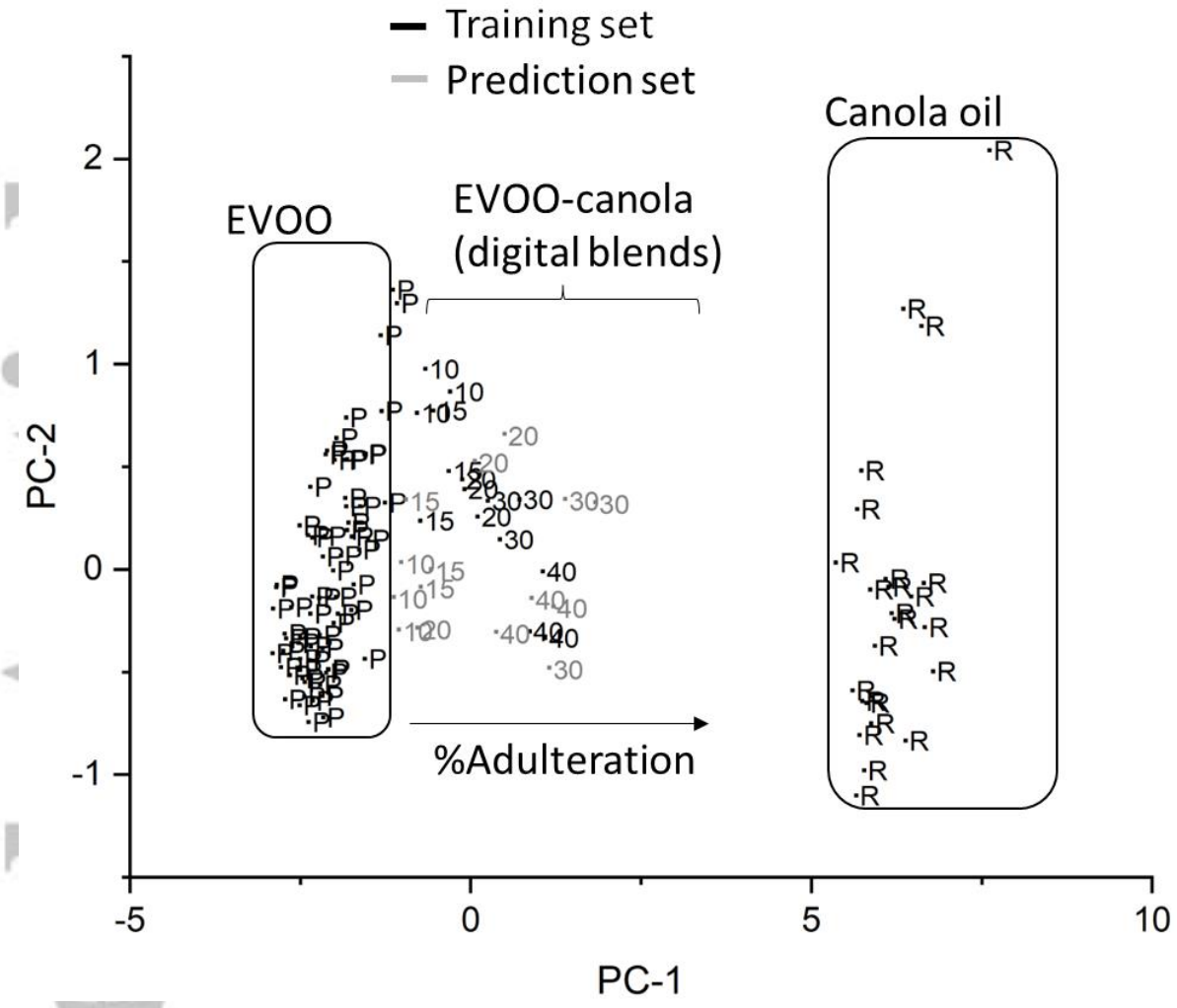

Figure 7. Plot of the two largest principal components of the 115 IR training set spectra (black) and the 11 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, canola oil and EVOO-canola oil mixtures (10% canola oil to 40% oil). The total cumulative variance explained by the two largest principal components for the blended data is 95.62%. The prediction set spectra are represented in grey. P = EVOO, R = canola oil, 10 = 10% canola oil, 15 = 15% canola oil, 20 = 20% canola oil, 30 = 30% canola oil and 40 = 40% canola.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

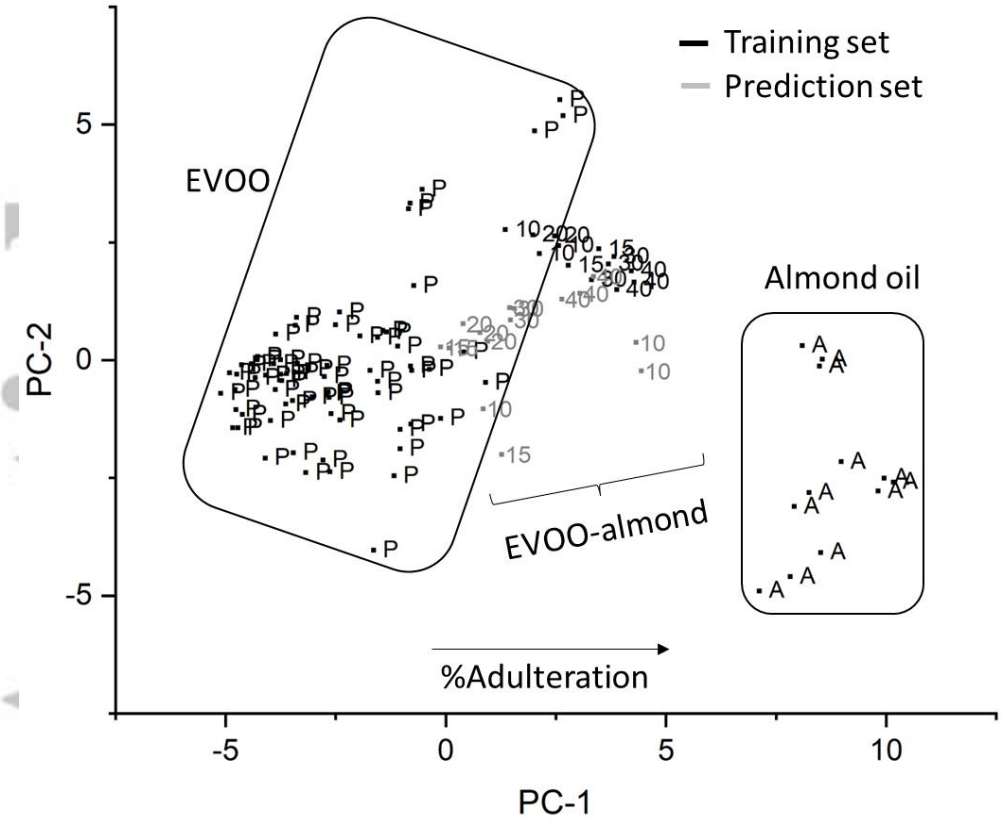

Figure 8. Plot of the two largest principal components of the 100 IR training set spectra (black) and the 5 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, almond oil and EVOO-almond oil mixtures (10% almond oil to 40% oil). The total cumulative variance explained by the two largest principal components for the experimental data is 85.29%. The prediction set spectra are represented in grey. P = EVOO, A = almond oil, 10 = 10% almond oil, 15 = 15% almond oil, 20 = 20% almond oil, 30 = 30% almond oil and 40 = 40% almond.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/cem.3469

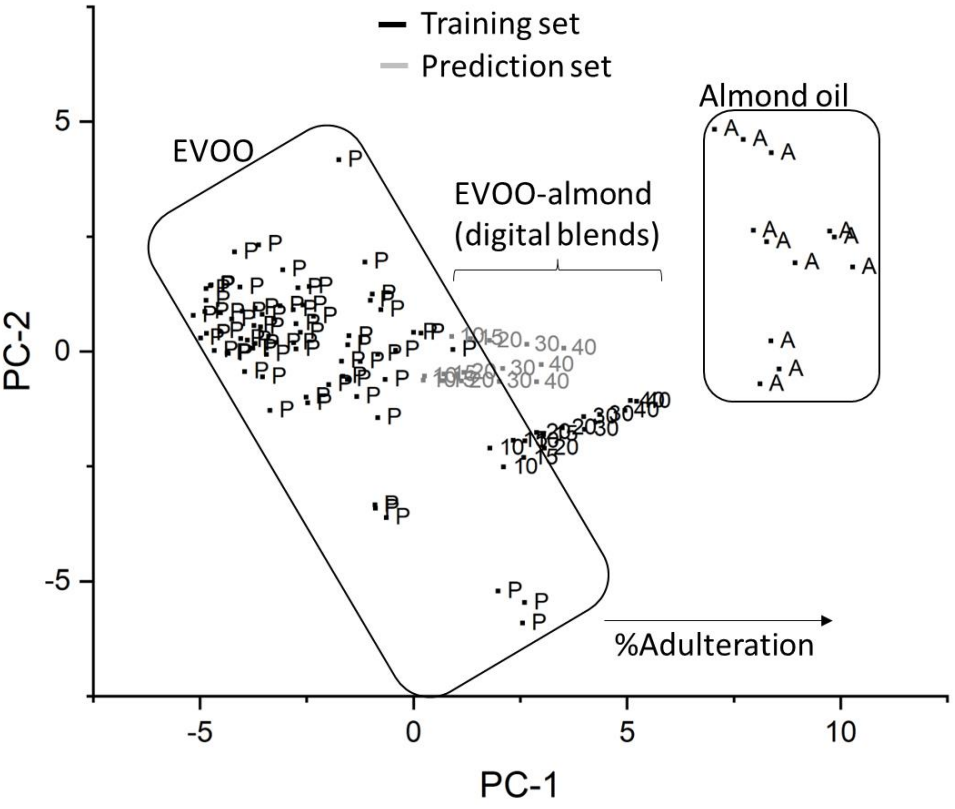

Figure 9. Plot of the two largest principal components of the 100 IR training set spectra (black) and the 25 spectral features identified by the pattern recognition GA for the three-way classification problem: EVOO, almond oil and EVOO-almond oil mixtures (10% almond oil to 40% oil). The total cumulative variance explained by the two largest principal components for the digitally blended data is 78.7%. The prediction set spectra are represented in grey. P = EVOO, A = almond oil, 10 = 10% almond oil, 15 = 15% almond oil, 20 = 20% almond oil, 30 = 30% almond oil and 40 = 40% almond.