Miniature spectrometer data analytics for food fraud

Fayas Asharindavida[a*], Omar Nibouche [a], James Uhomoibhi [b], Jun Liu [a], Jordan Vincent [a], and Hui Wang [c]

[a] School of Computing, Ulster University, UK; asharindavida-f@ulster.ac.uk; o.nibouche@ulster.ac.uk ; j.liu@ulster.ac.uk; vincent-j1@ulster.ac.uk

[b] School of Engineering, Ulster University, UK; j.uhomoibhi@ulster.ac.uk

[c] School of Electronics, Electrical Engineering and Computer Science, Queen's University of Belfast, UK; h.wang@qub.ac.uk

* Correspondence: asharindavida-f@ulster.ac.uk

S1.Related Works

LDA-KNN model has been applied and found effective in discriminating extra virgin olive oil from the adulterated one with lower quality vegetable oils (Garrido-Delgado, Eugenia Muñoz-Pérez, & Arce, 2018). In addition, regression techniques have also been used; the Partial Least Square (PLS) regression technique has been applied on FTIR spectral data for the prediction of the quantity of chemical components present in olive oil (Uncu, Ozen, & Tokatli, 2019). Olive oil adulterated with low quality vegetable oil is efficiently discriminated using Principal Component Analysis (PCA) and PLS methods (Lizhi, Toyoda, & Ihara, 2010). Classification results can be further improved if PLS is combined with dimension reduction techniques, PCA, LDA, probabilistic neural networks or by employing Partial Least Square Discriminant Analysis (PLS-DA) (Borràs et al., 2016). In (Dankowska & Kowalewski, 2019), a set of six classification algorithms, namely, LDA, the Quadratic Discriminant Analysis (QDA), Regularized Discriminant Analysis (RDA), KNN, SVM and RF, have been used for establishing the freshness of olive oil and its classification into its various types. Out of the six methods used, KNN and SVM have shown better performance in terms of misclassification error rates, attaining 5.7% and 3.5%, respectively. PLS regression has been successfully applied to predict the quantity of extra chemical components present in olive oil (Uncu & Ozen, 2015) and beverages (Wang et al., 2020). A variant regression model termed Support Vector Machine Regression (SVMR) was used on spectral data to find the density of rice roots (Xu, Zhao, Wang, & Shi, 2017). Further, a novel L2-regularized regression method has been introduced in (Song, Wang, Maguire, & Nibouche, 2018) for the classification of low quality spectral data acquired from portable devices. In addition, PLS regression models have been used to analyze spectral data from portable and miniature devices for checking the quality of meat (Grimmler, Kuhfuß, Heiden, & Schmidt, 2018), vegetables (Jantra, Slaughter, Liang, & Pathaveerat, 2017) and olive oil (Yildiz Tiryaki & Ayvaz, 2017). Furthermore, the Dempster-Shafer (DS) theory was used to improve the accuracy of olive oil classification (Saha & Saha, 2017). The resulting classifier out from the fusion of single classifiers, namely PCA, SVM and Artificial Neural Network (ANN) has been shown to improve classification accuracy compared with such standalone classifiers when used on NIR data. A deep learning based model has been applied on spectral data from a portable VIS-NIR spectrometer for the classification of different food powders (You et al., 2019). ML combined with spectral data analysis has been an emerging research area in food quality check for spice powder authentication (Zhou et al., 2022)(Oliveira, Cruz-Tirado, Roque, Teófilo, & Barbin, 2020).

Table S. 1 Notations and Symbols used in the manuscript.

SYMBOL	DESCRIPTION	SYMBOL	DESCRIPTION
$\boldsymbol{\Sigma}$	Summation operator	$\boldsymbol{\mu}$	Mean of all samples
$\boldsymbol{\Pi}$	Product operator	$\boldsymbol{\mu_i}$	mean of samples with class i
$'$	Transpose of a matrix, e, g. $\boldsymbol{v}'$, $\boldsymbol{W}'$	$\boldsymbol{v}$	Weight vector
$\boldsymbol{X}$	Data samples	$\boldsymbol{\phi}(\boldsymbol{v})$	Mapping function for mapping input space to feature space
$\boldsymbol{X_{corrected}}$	Corrected set of data samples	$\boldsymbol{\delta}$	Bias
$\boldsymbol{Y}$	Response value	$\boldsymbol{k}$	Number of nearest neighbours
$\boldsymbol{N}$	Number of datapoint instances	$\boldsymbol{B_{PLS}}$	Regression coefficient matrix
$\boldsymbol{x}$	Test sample	$\boldsymbol{R}, \boldsymbol{U}$	Latent representation of X and Y
$\boldsymbol{c_x}$	Class label for x	$\boldsymbol{P}, \boldsymbol{Q}$	Loading matrix of $X\ and\ Y$
$\boldsymbol{W}$	Projection matrix	$\boldsymbol{E}, \boldsymbol{F}$	Residual matrices in PLS decomposition
$\boldsymbol{J}(\boldsymbol{W})$	Fisher criterion	$\boldsymbol{P}(\boldsymbol{c}\|\boldsymbol{x})$	Probability of sample x classified to class c

S_b	Between class matrix	T	Number of decision trees
S_w	Within class matrix	$D(x)$	Decision of classifier with test sample
a, m	Additive effect and multiplicative effect of the spectrum	L	Number of classifiers
N_i	Number of samples in class i	$b_{ji}(x)$	Belief degree for class j and classifier i for classifying x
x_{ij}	j^{th} sample in class i	$\Psi_{j,i}(x)$	Mapping function to find relationship between class j and the classifier i
C_N	Number of classes	$\check{R}$	Number of random subsets

Table S. 2 Datasets summary [number of samples (n), number of features (p) and the number of classes (k)]

Dataset	n	p	k	Classes description	Device used
OO1	3800	1024	6	Six levels of adulteration (Raw data)	STS-UV
OO2-DHUV	1200	1024	6	Six levels of adulteration	STS-UV+DH Light source
OO2-DRNIR	1200	1024	6	Six levels of adulteration	STS-NIR + DR Probe
OO2-DRUV	1200	1024	6	Six levels of adulteration	STS-UV +DR Probe
OO2-DHNIR	1200	1024	6	Six levels of adulteration	STS-NIR +HL 2000
Apple	2400	1024	10	Ten different types of Apples	STS-NIR + HL 2000
Dates	1600	1024	16	Sixteen types of Dates	STS-NIR + HL 2000
Powder-VIS	900	1024	18	Flours, Spice, Chocolate, Salt, and Sugar	STS-UV + HL 2000
Powder-NIR	900	1024	18	Flours, Spice, Chocolate, Salt, and Sugar	STS-NIR + HL 2000

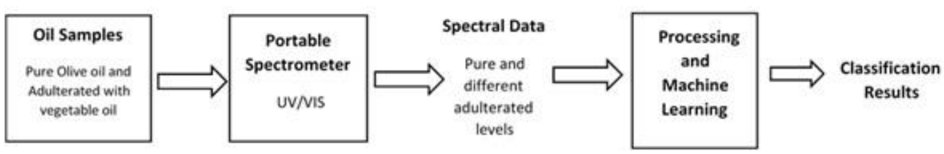

Figure S. 1 Olive oil adulteration check process.

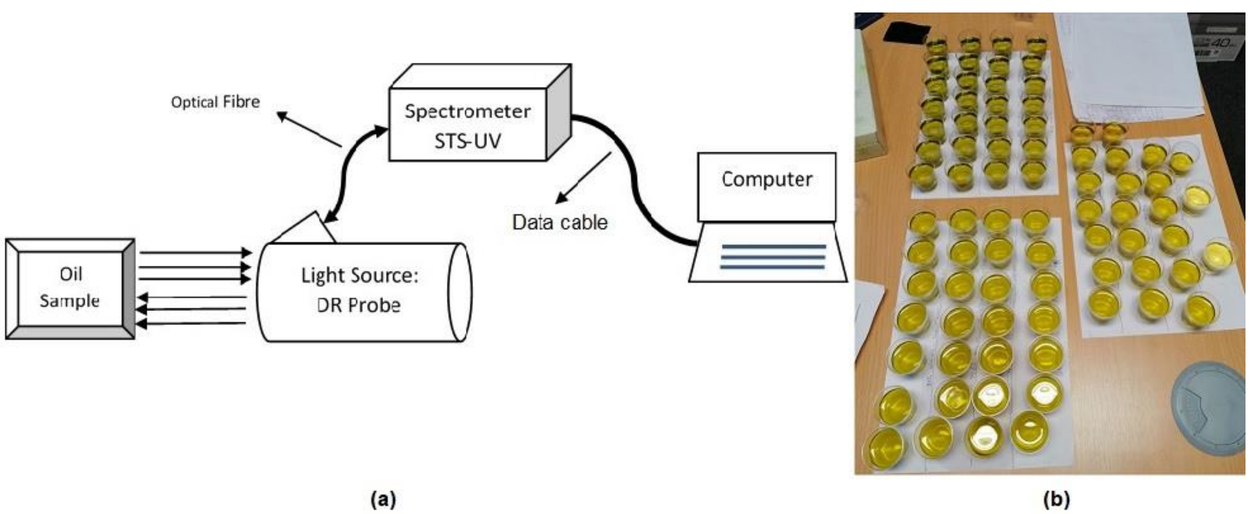

Figure S. 2 (a) Full view of the experimental setup with all components connected (b) Prepared oil samples.

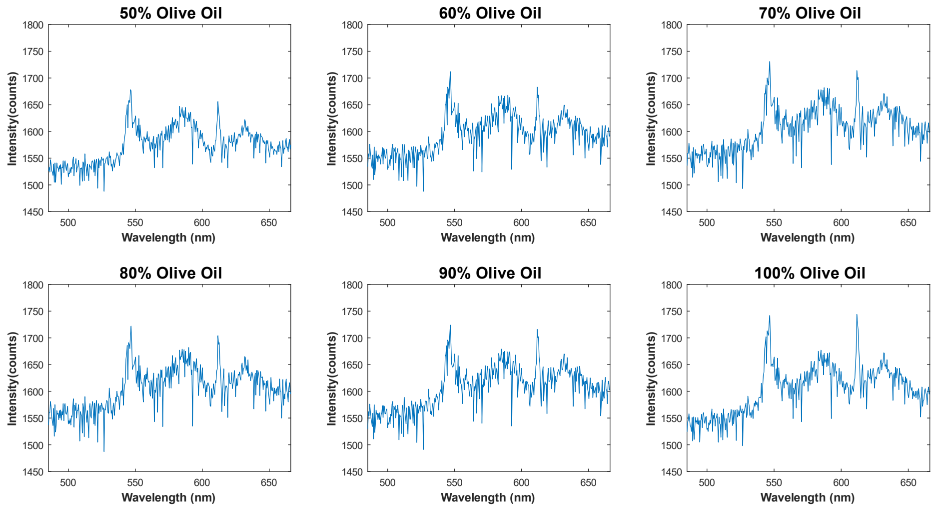

Figure S. 3 Plot of all classes of oil (showing the wavelength response in VIS region)

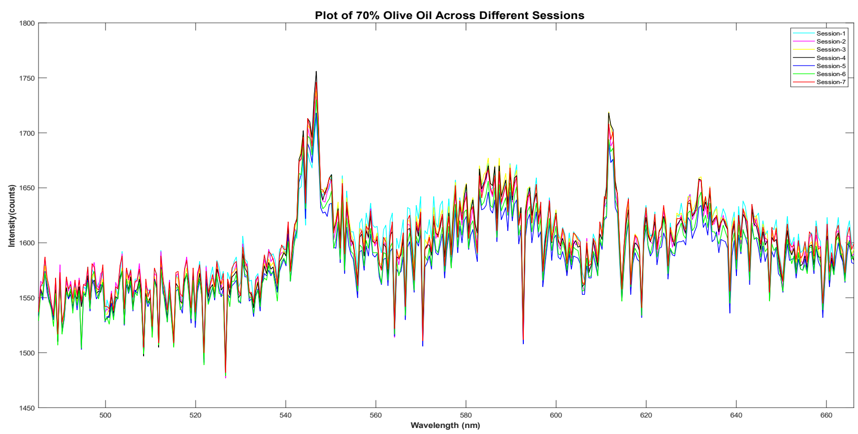

Figure S. 4 Variation of Olive oil spectral data across different sessions.

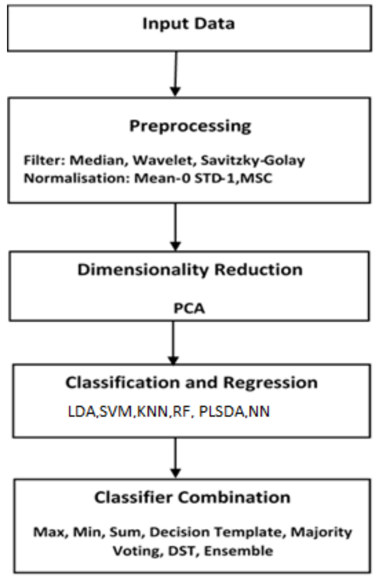

Figure S. 5 Pipeline of the methods used.

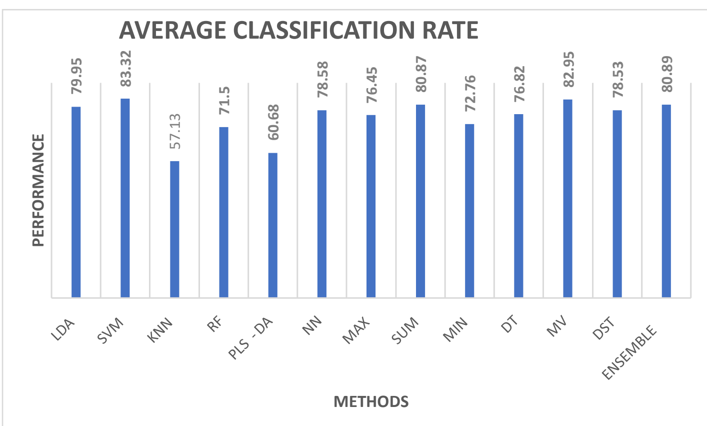

Figure S. 6 Average Classification results using all methods.

Supplementary Material References

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., Calvo, A., & Busto, O. (2016). Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA). *Food Chemistry*, *203*, 314–322. https://doi.org/10.1016/j.foodchem.2016.02.038

Dankowska, A., & Kowalewski, W. (2019). Comparison of different classification methods for analyzing fluorescence spectra to characterize type and freshness of olive oils. *European Food Research and*

Technology, *245*(3), 745–752. https://doi.org/10.1007/s00217-018-3196-z

Garrido-Delgado, R., Eugenia Muñoz-Pérez, M., & Arce, L. (2018). Detection of adulteration in extra virgin olive oils by using UV-IMS and chemometric analysis. *Food Control*, *85*, 292–299. https://doi.org/10.1016/j.foodcont.2017.10.012

Grimmler, C., Kuhfuß, T., Heiden, M., & Schmidt, H. (2018). Non-invasive assessment of the bioburden of minced pork using a hand-held fluorescence device. *Technisches Messen*, *85*(3), 177–183. https://doi.org/10.1515/teme-2017-0092

Jantra, C., Slaughter, D. C., Liang, P. S., & Pathaveerat, S. (2017). Nondestructive determination of dry matter and soluble solids content in dehydrator onions and garlic using a handheld visible and near infrared instrument. *Postharvest Biology and Technology*, *133*, 98–103. https://doi.org/10.1016/j.postharvbio.2017.07.007

Lizhi, H., Toyoda, K., & Ihara, I. (2010). Discrimination of olive oil adulterated with vegetable oils using dielectric spectroscopy. *Journal of Food Engineering*, *96*(2), 167–171. https://doi.org/10.1016/j.jfoodeng.2009.06.045

Oliveira, M. M., Cruz-Tirado, J. P., Roque, J. V., Teófilo, R. F., & Barbin, D. F. (2020). Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder. *Journal of Food Composition and Analysis*, *87*, 103403. https://doi.org/10.1016/J.JFCA.2019.103403

Saha, S., & Saha, S. (2017). A novel approach to classify edible oil using multiple classifier fusion based on spectral data. *2016 International Conference on Intelligent Control, Power and Instrumentation, ICICPI 2016*, 108–113. https://doi.org/10.1109/ICICPI.2016.7859684

Song, W., Wang, H., Maguire, P., & Nibouche, O. (2018). Collaborative representation based classifier with partial least squares regression for the classification of spectral data. *Chemometrics and Intelligent Laboratory Systems*, *182*, 79–86. https://doi.org/10.1016/j.chemolab.2018.08.011

Uncu, O., & Ozen, B. (2015). Prediction of various chemical parameters of olive oils with Fourier transform infrared spectroscopy. *LWT - Food Science and Technology*, *63*(2), 978–984. https://doi.org/10.1016/j.lwt.2015.05.002

Uncu, O., Ozen, B., & Tokatli, F. (2019). Use of FTIR and UV–visible spectroscopy in determination of chemical characteristics of olive oils. *Talanta*, *201*(April), 65–73. https://doi.org/10.1016/j.talanta.2019.03.116

Wang, Y. T., Li, B., Xu, X. J., Ren, H. Bin, Yin, J. Y., Zhu, H., & Zhang, Y. H. (2020). FTIR spectroscopy coupled with machine learning approaches as a rapid tool for identification and quantification of artificial sweeteners. *Food Chemistry*, *303*, 125404–125415. https://doi.org/10.1016/j.foodchem.2019.125404

Xu, S., Zhao, Y., Wang, M., & Shi, X. (2017). Determination of rice root density from Vis–NIR spectroscopy by support vector machine regression and spectral variable selection techniques. *Catena*, *157*(April), 12–23. https://doi.org/10.1016/j.catena.2017.05.008

Yildiz Tiryaki, G., & Ayvaz, H. (2017). Quantification of soybean oil adulteration in extra virgin olive oil using portable raman spectroscopy. *Journal of Food Measurement and Characterization*, *11*(2), 523–529. https://doi.org/10.1007/s11694-016-9419-8

You, H., Kim, H., Joo, D. K., Lee, S. M., Kim, J., & Choi, S. (2019). Classification of Food Powders with Open Set using Portable VIS-NIR Spectrometer. *1st International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2019*, 423–426. https://doi.org/10.1109/ICAIIC.2019.8668992

Zhou, L., Tan, L., Zhang, C., Zhao, N., He, Y., & Qiu, Z. (2022). A portable NIR-system for mixture powdery food analysis using deep learning. *LWT*, *153*, 112456. https://doi.org/10.1016/J.LWT.2021.112456