

ARTICLE OPEN



Novel time-domain NMR-based traits for rapid, label-free Olive oils profiling

Vasco Rafael dos Santos^{1,2,3}, Victor Goncalves², Peishan Deng³, Ana Cristina Ribeiro^{1,2}, Mariana Maia Teigao², Bárbara Dias², Inês Mendes Pinto^{2,4}, Juan Gallo² and Weng Kung Peng³

Olive oil is one of the oldest and essential edible oils in the market. The classification of olive oils (e.g. extra virgin, virgin, refined) is often influenced by factors ranging from its complex inherent physiochemical properties (e.g. fatty acid profiles) to the undisclosed manufacturing processes. Therefore, olive oils have been the target of adulteration due to its profitable margin. In this work, we demonstrate that multi-parametric time-domain NMR relaxometry can be used to rapidly (in minutes) identify and classify olive oils in label-free and non-destructive manner. The subtle differences in molecular microenvironment of the olive oils induce substantial changes in the relaxation mechanism in the time-domain NMR regime. We demonstrated that the proposed NMR-relaxation based detection (AUC = 0.95) is far more sensitive and specific than the current gold-standards in the field i.e. near-infrared spectroscopy (AUC = 0.84) and Ultraviolet-visible spectroscopy (AUC = 0.73), respectively. We further show that, albeit the inherent complexity of olive plant natural phenotypic variations, the proposed NMR-relaxation based traits may be a viable mean (AUC = 0.71) in tracing the regions of origin for olive trees, in agreement with their geographical orientation.

npj Science of Food (2022)6:59; <https://doi.org/10.1038/s41538-022-00173-z>

INTRODUCTION

Olive oil (OO) is one of the oldest and essential edible oils commercially traded in the history of mankind. Olive oils are commonly classified into extra-virgin olive oils (EVOOs), virgin olive oils (VOOs) or mixed with refined olive oils (refined OOs) (Fig. 1A), depending on among other factors, its fatty acids (FA) profiles and the trace compounds (e.g. concentration of free fatty acids (FFA) or acid value (AV)^{1–4}, phenolic compounds⁵). FAs are predominantly defined by its saturation levels (e.g. saturated fatty acids (SAFA), monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA)) (Fig. 1B). The FFA content is influenced by a number of phytosanitary factors and extraction processes^{6–8}. As a consequent of variation in processing (e.g. poor olive quality or inadequate extraction process), triacylglycerols structural breakdown may occurs (due to for example high temperature and moisture induced hydrolysis⁹), resulted in an increment in the final acidity of the oils^{3,4} (Fig. 1C, D).

The high demand of OO comes from its multiple nutritional benefits and its irreplaceable organoleptic properties^{10,11}. Olive oil is by far one of the most frequently adulterated food products due to its high customer appeal and large profitable margin^{12,13}. The highly desired and expensive EVOO is frequently diluted with cheaper adulterated oils, leading to indirect economic consequences and health concerns. Hence, olive oil has been the subject of rigorous quality regulations, with its standardization characteristics set amid tight legislation.

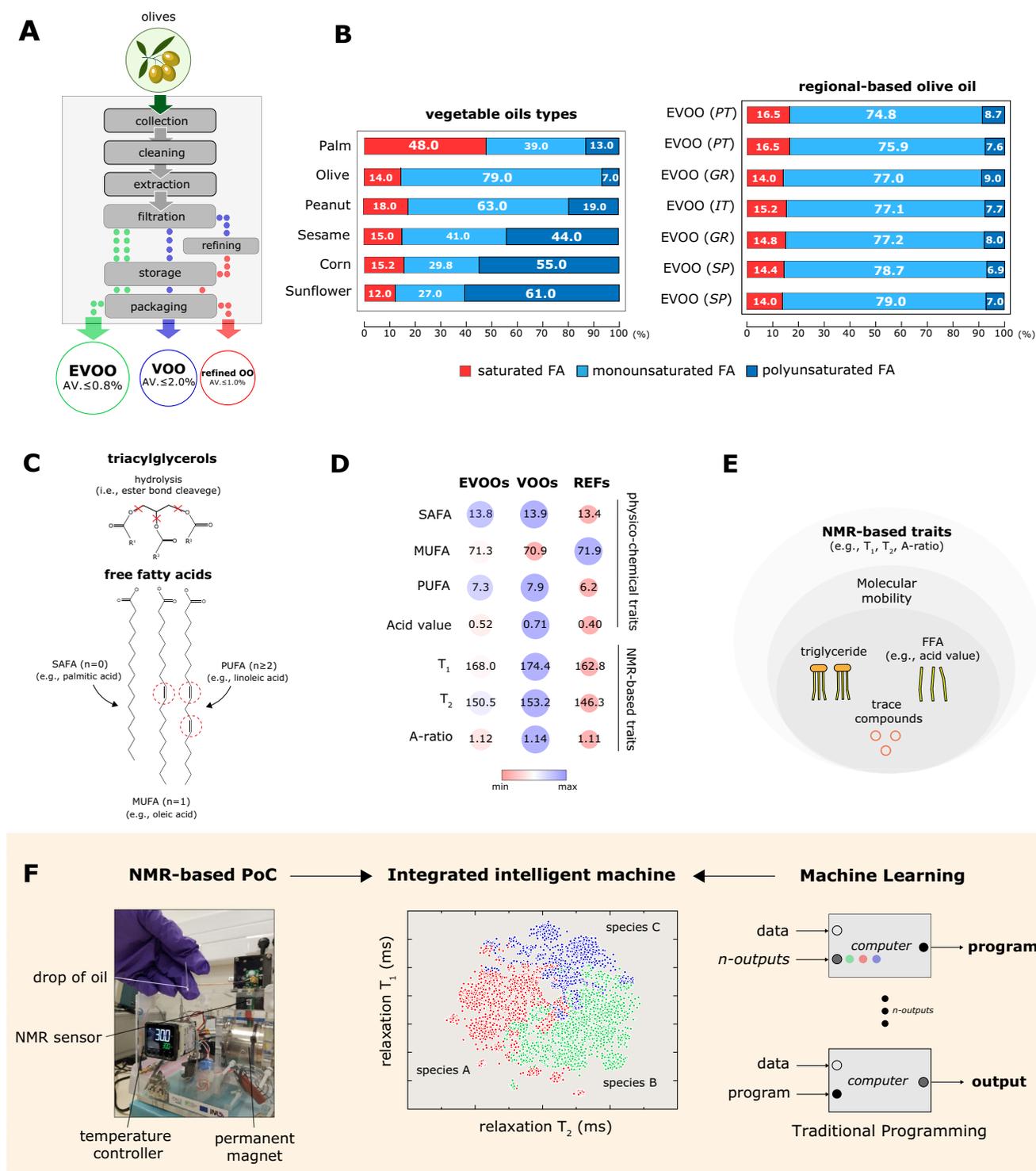
Laboratory-based methods, such as chromatography^{13–16}, spectroscopy^{12,17–21}, or DNA analysis^{13,22,23}, have been extensively developed to reduce the cases of adulteration. Nuclear magnetic resonance (NMR) spectroscopy in the high-field frequency domain has also been proposed to be an effective method on the detection of authentication, quality control, and adulteration of the oils. High-field NMR, however, has a number of drawbacks, such as the requirement of large, dedicated laboratory facilities

with costly cryogenic cooling gases, complicated pre-analysis steps, and the need of a highly specialized workforce^{17,24,25}. None of the above-mentioned detection methods are simple to use, require minimal sample preparation, nor present short turn-around time.

We have recently demonstrated that two-dimensional time-domain NMR can be used to classify edible oils based on their physiochemical composition (e.g. saturation levels) with much higher accuracy than the conventional methods²⁶. The low-field NMR-based point-of-care (PoC)^{27–32} analysis is based on pairing the longitudinal (T_1) and transversal (T_2) relaxation times, which improves the sensitivity and specificity of the detection significantly. It works on the rationale that accumulative characteristics of each dimensionality form a specific and unique signature, in a way similar to the radiomics technique developed in the field of radiology.

In this work, we demonstrate that NMR-based phenotypic traits in the time-domain (at molecular level) can be used for classifying the OOs. Using just a single droplet, we demonstrated that using benchtop sized NMR^{33–35}, olive oils can be rapidly classified (into EVOOs, VOOs or refined OOs) in non-destructive manner (i.e. label-free or without sample pre-treatment). The subtle differences in physiochemical composition and molecular microenvironment of the olive oils induce substantial changes in the relaxation mechanism in the time-domain NMR regime (Fig. 1E, F). With the aid of machine learning, the sensitivity and specificity of the detection were shown to have AUC = 0.95 using T_1 relaxation and T_2 relaxation, much higher than current gold-standards, the near-infrared spectroscopy (NIRS, AUC = 0.84) and Ultraviolet-Visible spectroscopy (UV-Vis, AUC = 0.73) (Table 1), and much better performance in the identification of regions of origin (Table 2). In addition, the proposed NMR-based detection methods were much cheaper per assay, user-friendly, and can be used at point-of-detection (Table 3). This work demonstrated the spirit of

¹University of Minho, Braga 4704-553, Portugal. ²International Iberian Nanotechnology Laboratory, Braga 4715-330, Portugal. ³Songshan Lake Materials Laboratory, 523-808 Dongguan, China. ⁴Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal. ✉email: pengwengkung@sslab.org.cn



combining the (old-fashioned) machine with the (new-wave) of machine learning, to produce an 'intelligent machine'^{30,36,37}, an attractive scientific solution for the food science community.

RESULTS

Rapid identification and characterization of olive oils with NMR-based PoC

In order to demonstrate the industrial applications, we use the proposed technique to validate the authenticity of EVOO from

VOOs and refined OO (Fig. 2). The relaxometry measurements and acidity determination (details in Methods) were performed on thirty-six types of OOs (i.e. 21 EVOOs, 8 VOOs, and 7 refined OOs,) without disclosing the manufacturers label and country of origin. For each sample, the relaxation measurements were carried out using five different samplings, with the refined OO was performed as control experiment.

A two-dimensional map T_1 - T_2 magnetic state diagram was used to enumerate the object clustering based on the composite intrinsic relaxation properties of the oils, thereof, forming a

Fig. 1 Identification and classification of Olive oils using home-built NMR-based PoC. **A** Olive oil production process. The mechanical processes (e.g. cold press) in the extraction of olive oils, in particularly in the separation phase (e.g. filtration, refining) play a major role in preserving the amount of free fatty acids in OOs, and thus in the final grading of the product (e.g. EVOOs, VOOs, and refined OO). EVOO must be obtained using exclusively mechanical extraction procedures and preserving an acid value (AV) of less than 0.8%. On the other hand, acid value of VOO and refined OOs must be below 2.0% and 1.0%, respectively. **B** In OOs, fatty acids are predominantly in the form of triacylglycerols and are defined by their saturation level (e.g. saturated, monounsaturated, polyunsaturated). FA concentration varies greatly based on the type of vegetable oil and the regions of origin of the starting products. OOs are dominated by oleic acid (monounsaturated FA), similarly to peanut oil. Palm oils are saturated FA dominant, in contrast to sunflower and corn oils which are polyunsaturated FA dominant. **C** Olive oil composition is dominated by triacylglycerols (e.g. triolein) with free fatty acids and 'trace' compounds (e.g. vitamin E, and other bioactive molecules) in smaller amounts. Fatty acids are characterized by their saturation level (n) and by the hydrocarbon chain properties (mainly length). For $n = 0$, FAs are known as SAFAs (saturated FAs). For $n = 1$ and $n > 1$, FAs are known as MUFAs (monounsaturated) and PUFAs (polyunsaturated), respectively. An increase in free fatty acid pool (e.g. acidity) occurs due to the hydrolysis of triacylglycerols esters. **D** A bird-eye view of the relationship between the NMR-based traits, physicochemical properties, and OOs grading (EVOO, VOO, and refined OOs). The classification of OOs is influenced by the FA and FFA profiles, among other factors. **E** NMR-based traits originate on the relaxation dynamics of nuclear spins of proton nuclei due to the composite effect of triacylglycerols, free fatty acids, trace compounds, and the overall environment of the protons. **F** The concept of using integrated intelligent machine as proposed in this work. The developed NMR-based PoC consists of a portable commercial console, home-built detection circuit coil, and a palm-sized permanent magnet ($B = 0.5$ T). For high-throughput analysis, a microcapillary tube designed to be slotted into the NMR detection coil is used to store minute sample (e.g. a single drop of oil). The entire assay completes in less than 5 min. The NMR measurements were carried out in single blinded manner on each oil.

calibration standard for the (EVOOs, VOOs, refined OOs), and (150.5 ms, 168.0 ms), (153.2 ms, 174.4 ms), (146.3 ms, 162.8 ms), respectively (Fig. 2A and details in Supplementary Fig. 1). The oil types were significantly clustered ($P < 0.005$) indicate that the intra-variation samplings were much smaller than the inter-variation of the OOs (Fig. 2B). The details breakdown for each commercial brand is shown in heatmap (Fig. 2C). In addition, the Receiver Operating Characteristics (ROC) analysis (Fig. 2D and Supplementary Fig. 2) indicated that relaxometry measures have excellent detection sensitivity and specificity with Area Under the Curve (AUC) of 0.95 as compared to its counterparts NIRS (0.84) and UV-Vis (0.73), respectively (Table 1).

Identification of OO based on the regions of origin

We demonstrated the feasibility of using the proposed NMR analyses in identification of production based on their countries (or regions) of origin. Apart from the genotypic variation, the variation in phenotypic traits is governed by number of factors, such as migration drift (e.g. diversification and domestication events)³⁸, and abiotic stress (e.g. local climate, soil conditions)^{39,40}. For the identification of the regions of origin for OO, a matrix of data subsets, encompasses four different regions taken from the European regions (i.e. 3 Greece, 4 Italy, 9 Portugal, and 5 Spain) were enumerated using two-dimensional T_1 - T_2 magnetic state diagram (Fig. 3) and the details of each oil variations (details in Supplementary Fig. 3, and Supplementary Table 1).

The mean T_1 relaxation times of (166.3, 166.7, 168.9, and 168.9) ms, and for T_2 relaxation times of (147.7, 150.1, 150.2, and 151.0) ms for (Greece, Spain, Italy, Portugal), respectively (Fig. 3, and details in Supplementary Fig. 3). The regional-based identification for NMR technique is $AUC = 0.71$, much higher or comparable to NIRS ($AUC = 0.70$) and UV-Vis ($AUC = 0.69$) (details in Table 2). Interestingly, when a pair-wise comparison matrix (i.e. pair-wise ROC-AUC evaluation) is employed using NMR-based traits (e.g. T_1 , T_2 , A-ratio) it resembles the geographical orientation (Fig. 4A). For example, Greece-Italy ($AUC = 0.74$), Greece-Spain ($AUC = 0.84$), and Greece-Portugal ($AUC = 0.89$) shown as a heatmap (Fig. 4B). The Iberian region (i.e. Spain-Portugal) and Italy-Greece displayed stronger similarities with AUCs of (0.69, 0.74), respectively. This is to be expected as neighbouring countries are expected to have much higher of species exchange due to its proximity in geographical location. The details of each oils purchased displayed a unique information on their location (Fig. 4C and D).

Limit of detection of NMR-based traits technique

We evaluated the limit-of-detection of NMR-based traits by mixing sunflower oil into a selected EVOO, to mimic the cases of

adulteration. For each sample, the relaxation measurements were conducted in double using five different samplings, covering from 0% (as control) to 100% of OO in the mixed edible oil (Supplementary Fig. 4). As clearly indicated in the T_1 - T_2 magnetic state diagram, a linear relation ($r^2 = 0.93$) between NMR-based traits and the concentration of sunflower oil (PUFA-rich) reduced into EVOOs (MUFA-rich) relaxation effect becomes clearer (due to a decrease in saturation level). Therefore, the (T_2 , T_1) coordinates were (188.3, 202.9) and (155.3, 174.6) for sunflower oils and EVOO (controls), respectively. The limit of detection for NMR-based traits were approximately (1%), were either comparable to NIRS (1%) or much better than UV-Vis spectroscopy (5%) (details in Supplementary Fig. 4).

DISCUSSION

We report NMR-based point-of-care technology for fast, label-free, and distinctive OO profiling and to assure its high quality, which can be used to reduce the attempts in adulteration. The NMR-based phenotypic traits represent the intrinsic molecular relaxation dynamics (or molecular mobility) due to the composite effect of the FA profiles (e.g. saturation level) and concentration of FFA (e.g. acid value). Nevertheless, despite OOs consists of predominantly the monounsaturated fat (more than 70%), we found in this work that the overall saturation levels (e.g. increasing PUFA/MUFA ratio, lower SAFA content) has profound impact on the NMR traits (details in Supplementary Fig. 5). Secondly, we observed that FFA concentration has direct effect on the NMR-based phenotypic traits. We hypothesized that, with similar mechanism i.e. the saturation levels and FFA concentration disrupts the packing^{41,42} 'efficiency' (i.e. weakening of Van der Waals forces) leading to a disruption in the molecular mobility and hence introducing much longer pathways for relaxations (i.e. longer T_1 and T_2). This is in agreement with the recent work reported by Cistola⁴³.

Conventionally, chromatographic-based techniques, are extremely slow, time-consuming, require complicated multiple sample preparation steps with expensive laboratory equipment, while complicated chemometric analysis (e.g. vibrational, RAMAN spectroscopy) is required for in depth data interpretation, in comparison to the proposed NMR-based detection methods and other state-of-the-art technologies (refer to SWOT-like Table 3). The information derived from the analytical instrument represents one of the major challenges faced by food scientist during the identification and classification of pure and adulterated food samples. With the introduction of EU Protected Designation of Origin registration and equivalent in other geographical locations, rapid classification (preferably in non-destructive manner) of EVOOs will be invaluable to industry and regulatory agencies alike.

Table 1. Classification of olive oils using the Receiver Operating Characteristic analysis.

techniques	Model	AUC	CA	F1	Precision	Recall
UV-VIS ($\lambda = 415$ nm)	kNN	0.898	0.833	0.833	0.835	0.833
	Logistic Regression	0.444	0.417	0.37	0.333	0.417
	Naive Bayes	0.615	0.472	0.479	0.495	0.472
	Neural Network	0.762	0.667	0.663	0.672	0.667
	Random Forest	0.937	0.778	0.781	0.79	0.778
	Average	0.731	0.633	0.625	0.625	0.633
NIRS ($\lambda = 670$ nm)	kNN	1.000	1.000	1.000	1.000	1.000
	Logistic Regression	0.542	0.417	0.362	0.333	0.417
	Naive Bayes	0.771	0.75	0.743	0.778	0.75
	Neural Network	0.875	0.833	0.822	0.889	0.833
	Random Forest	1.000	1.000	1.000	1.000	1.000
	Average	0.838	0.800	0.785	0.800	0.800
NMR-based traits	kNN	0.974	0.889	0.889	0.889	0.889
	Logistic Regression	0.984	0.889	0.889	0.889	0.889
	Naive Bayes	0.95	0.861	0.864	0.878	0.861
	Neural Network	0.918	0.889	0.889	0.889	0.889
	Random Forest	0.919	0.833	0.831	0.834	0.833
	Average	0.949	0.872	0.872	0.876	0.872

Area Under the Curve (range between 0 and 1) of the various supervised models evaluated to predict the oil types. Models were validated using Leave-one-out method with averaged NMR-based traits (e.g. T_1 relaxation, T_2 relaxation, and A-ratio). The wavelength (λ) used for UV-Vis spectroscopy (at 415 nm) and NIRS (at 670 nm) was chosen upon the region of highest peaks resolution. Confusion matrix of individual models is displayed in Supplementary Fig. 7.

Table 2. Receiver Operating Characteristic analysis for regions of origin.

techniques	Model	AUC	CA	F1	Precision	Recall
UV-VIS ($\lambda = 670$ nm)	kNN	0.856	0.667	0.646	0.679	0.667
	Logistic Regression	0.403	0.179	0.203	0.274	0.179
	Naive Bayes	0.674	0.385	0.387	0.422	0.385
	Neural Network	0.751	0.641	0.62	0.684	0.641
	Random Forest	0.786	0.694	0.685	0.704	0.694
	Average	0.694	0.513	0.508	0.553	0.513
NIRS ($\lambda = 670$ nm)	kNN	0.856	0.667	0.646	0.679	0.667
	Logistic Regression	0.43	0.179	0.131	0.127	0.179
	Naive Bayes	0.687	0.385	0.383	0.382	0.385
	Neural Network	0.753	0.436	0.433	0.456	0.436
	Random Forest	0.752	0.641	0.625	0.635	0.641
	Average	0.696	0.462	0.444	0.456	0.462
NMR-based traits	kNN	0.718	0.538	0.539	0.541	0.538
	Logistic Regression	0.658	0.433	0.365	0.331	0.433
	Naive Bayes	0.667	0.433	0.383	0.348	0.433
	Neural Network	0.788	0.576	0.561	0.571	0.576
	Random Forest	0.699	0.500	0.497	0.497	0.500
	Average	0.706	0.496	0.469	0.458	0.496

Area Under the Curve (range between 0 and 1) of the various supervised models evaluated to predict OO origin. Models were validated using Leave-one-out method with NMR-based traits (e.g. T_1 relaxation, T_2 relaxation, and A-ratio). The wavelength (λ) used for UV-Vis spectroscopy and NIRS was 670 nm (chosen under the region with highest resolution). Confusion matrixes of individual models are displayed in Supplementary Fig. 8.

On the other hand, the proposed NMR-based technology provides rapid, precise, low-cost, label-free, and accurate analysis for grading the olive oils quality using the NMR-based phenotypic traits in the time-domain NMR. In this framework, the central hypothesis of radiomics is that it is possible to decode tissue characteristics and pathology by examining the textural features

in medical images. Similarly, clustering NMR techniques work on the rationale that accumulative characteristics of each dimensionality form a specific and unique signature ('molecular fingerprint') is extremely powerful for rapid and accurate classification of OOs based on the NMR-based phenotypic traits. In addition, with the introduction of machine learning, it is now inexpensive to process

Table 3. Qualitative performance of the NMR-based PoC against conventional methods (UV-Vis, NIR).

features	Integrated NMR-based PoC	UV-VIS, NIR Spectrometry
Sensitivity	Very high	Medium/high
Specificity	Very high	Medium/high
ROC (AUC)	(0.95)	(0.73, 0.84)
Extensive experience	Not required	Not required
Time to results	Minutes	Minutes
Sample processing	Nil (no solvents needed)	Need specific solvents, nil
Price per assay	Ultra-cheap	Expensive (cuvettes, solvents)
Equipment size	Point-of-care testing	Benchtop

SWOT-like analysis between the state-of-the-art technologies (e.g. Near-Infrared spectroscopy and UV-Vis (Supplementary Figs. 9–10)) versus the integrated intelligent machine proposed in this work (machine learning assisted NMR-based PoC).

large datasets running in almost real-time setting, opening door to intelligent machine which can make interpretation with much higher sensitivity and specificity.

METHODS

Details and sample preparation of the OOs

OOs analyzed were cooking oils bought locally in Braga, Portugal or purchased online (e.g. international brands). The commercial brands names were disclosed (in details in Supplementary Table 2). No further processing was made before the NMR measurements and all other measurements.

NMR measurements and parameters

The ^1H magnetic resonance measurements of olive oils were acquired at the resonance frequency of 21.7579 MHz polarized using a portable permanent magnet (Metrolab Instruments, Switzerland), $B_0 = 0.5\text{ T}$, using a benchtop-type console (Kea Magritek, New Zealand). A temperature controller was set to maintain the measurement chamber at 30°C . The T_1 relaxation and T_2 relaxation times were acquired using standard inversion recovery (IR) and Carr-Purcell-Meiboom-Gill (CPMG) train pulse sequences, respectively. The experimental parameters used were

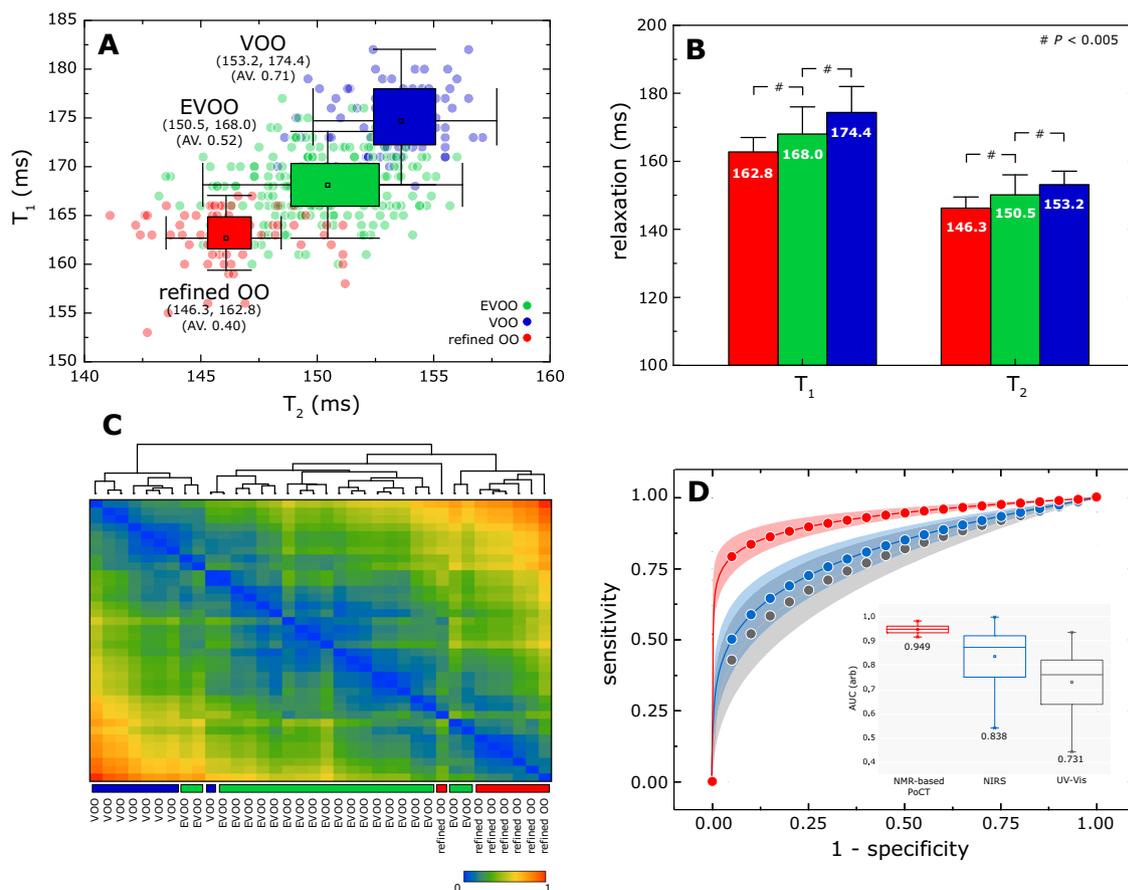


Fig. 2 Rapid identification and characterization of olive oil using NMR-based traits. **A** Two-dimensional mapping of EVOOs (green), VOOs (blue), and refined OO (red) in the T_1 - T_2 magnetic phase diagram. A wide variety of Olive oils (i.e. 21 EVOOs, 8 VOOs, and 7 refined OOs) commercially available from different manufacturers were purchased off-the-shelf in Braga, Portugal or otherwise through online platforms. Each data point represents one sampling with a total 360 samplings collected. The mean T_1 and T_2 values and acid value were denoted below. **B** Average T_1 and T_2 relaxation times (ms) for the different types of OOs. The statistical analysis of the data was calculated using unpaired two-tailed Student t -tests ($P < 0.005$) (details in Supplementary Fig. 1). **C** Rapid classification of OOs using the NMR-based traits in the form of clustering analysis. This hierarchical clustering was constructed based on the Euclidean distance between the averaged measures per sample (details in Supplementary Table 1). Their quantitative linkages (e.g. inter- and intra-cluster similarity) are shown as a heatmap. **D** The ROC curves for NMR-based traits (red), NIRS (blue), and UV-Vis (grey) calculated from a number of supervised models (details in Supplementary Fig. 2). The error zone were 99% of the confidence band.

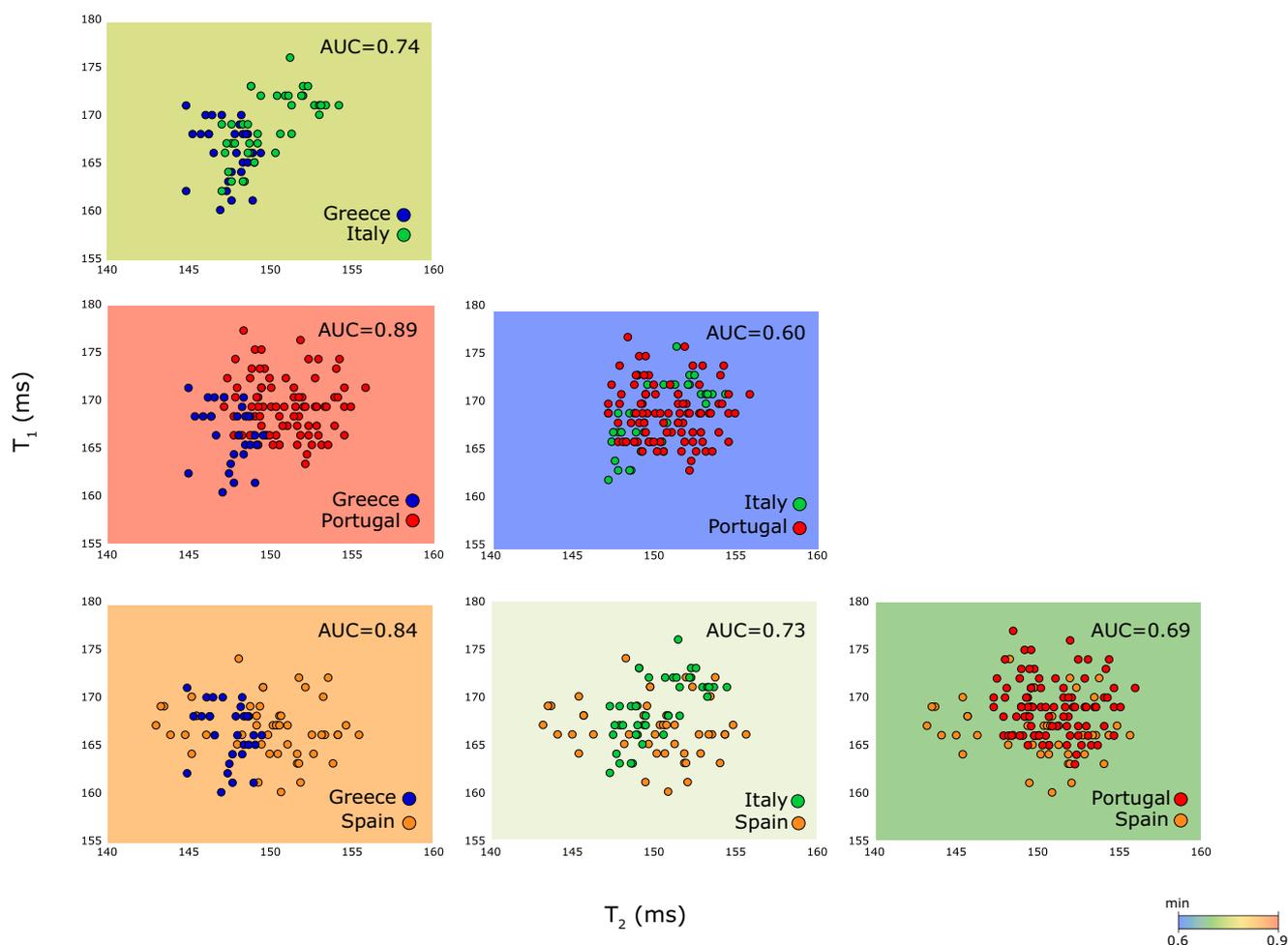


Fig. 3 Identification of the regions of origin of OO. EVOO samples were studied based on their regions of origin. Off-the-shelf EVOO samples originated from different European regions (i.e. 9 Portugal, 5 Spain, 4 Italy, 3 Greece) according to their labelling. Pair-wise two-dimensional mapping of EVOOs origin in the T_1 - T_2 magnetic state diagram, according to unpaired two-tailed Student t -tests (details in Supplementary Fig. 3). The sensitivity and specificity of each region were calculated using receiver operating characteristic (ROC). The substantially high AUC, ranging from 0.6 to 0.9 of each pair/wise region were evaluated.

echo time = 200 μ s, number of echoes = 10,000, and signal averaging = 32. A recycle delay of 2 s was set between each experiment to provide sufficiently long time to allow all molecular spins to return to thermal equilibrium. (T_1 relaxation, T_2 relaxation) measurements were carried out on commercial EVOOs, VOOs and refined OOs. NMR measurements were performed blindly on each oil ten repeated times, with a total of 360 points for olive type classification, and 210 points for origin assessment. Clustering NMR methodology uses a pair of relaxation times (T_2 , T_1) for each object (oils in this case) to construct a (pseudo) two-dimensional map (Figs. 2A and 3).

UV-VIS and NIR measurements and detection

UV-Vis measurements were performed in a SHIMADZU UV-2550 spectrophotometer (Kyoto, Kyoto, Japan), while for NIR measurements a PerkinElmer LAMBDA 950 instrument was used. All samples were measured in matched 1 cm path length quartz or optical glass cells, running an empty cell as a reference. UV-Vis spectra were measured within 200–800 nm spectral range at 1 nm spectral resolution, while NIR, spectra were obtained within 500–2200 nm with 5 nm steps. NIR spectra spike removal algorithms⁴⁴ were applied (cut-off = 6, threshold = 10). Every

sample was measured three times and the mean values were taken as representation.

Acid value measurements

The acid value determination was performed under the EN ISO 660:2009⁴⁵ protocol for oleic acid quantification. Simply, 10 mL of edible oil were weighted and diluted in 20 mL of ethanol ($\phi = 99\%$), with small amounts of phenolphthalein. Titrations with 0.1 mol/L of potassium hydroxide (KOH) were done under magnetic stirring until slight colour changes appear (and persisted for +10 s). Measures were executed twice per sample. The acid value was extrapolated from the amount of KOH required for each sample, defined as the amount of KOH required to neutralize 1 g of chemical substance, with the following formula:

$$w_{AV} = \frac{56.1 \times cV}{m} \quad (1)$$

where, c is the exact concentration of the standard KOH solution (mol/L), V the volume of KOH added (mL), and m the mass (g) of the test portion. Acidity, or the free fatty acid content, can be estimated by:

$$w_{FFA} = \frac{VcM}{10 \times m} \approx 0.5 \times w_{AV} \quad (2)$$

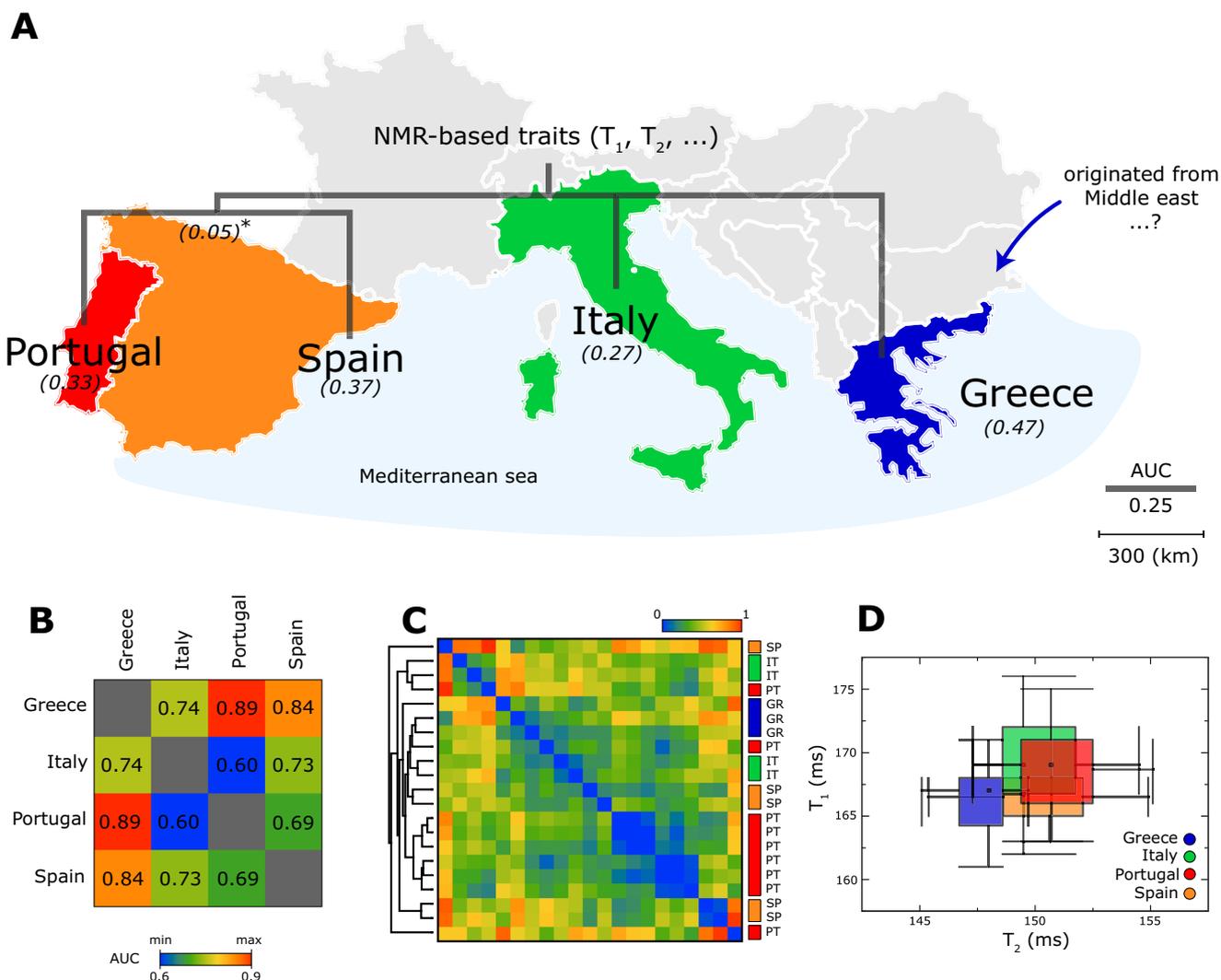


Fig. 4 NMR-based traits in identification of the regions of origin. **A** The NMR-based phylogenetic tree was built using the AUC distance matrix (details in Fig. 3) using neighbour joining algorithm (Supplementary Fig. 6) which splits the NMR-based traits into three main regions (i.e. Iberian, Italy, Greece). Iberian are Portugal and Spain which shared the same land border. The proposed NMR-relaxation based traits (legends of AUC is 0.25 in vertical) in agreement with their geographical orientation (shown in legend of 300 km per cm). Higher similarities are expected to be found species that are closely related. Neighbouring countries are expected to have higher species exchange and genes flow due to their geographical proximity. Similarities fade away with, for example, geographical distance. **B** A summary of the AUCs between countries evaluated using ROC analysis (in details in Fig. 3). **C** The detailed analysis of each commercial brands represented in the form of heatmap. Hierarchical clustering was constructed based on the Euclidean distance (between T_1 relaxation, T_2 relaxation, and A-ratio) of the averaged measurement per sample (Supplementary Table 1). **D** T_1 - T_2 magnetic state diagram in the forms of box plots for the regions of origin.

wherein, M is the molar mass (g/mol) of the predominant fatty acid in the edible oil, in this case oleic acid (282.47 g/mol).

Machine learning algorithm and workflows

Using statistical programming languages (e.g. Orange 3.1.2⁴⁶ or R), the raw datasets were processed using supervised and unsupervised learning techniques. The machine learning algorithms were written and run on a personal laptop (Intel Core Pentium i7 CPU @ 2.70 GHz, 8.00GB RAM). Once the model in machine learning was built, all the tasks run simultaneously and completed typically in less than 1 min. Using unsupervised learning, the relationship between each object was rapidly constructed using clustering analysis (e.g. hierarchical clustering) and its quantitative linkages (e.g. inter-/intra-cluster similarity) were shown on a dendrogram and a heatmap. Supervised learning models (i.e. Neural Network, kNN, Logistic Regression, Naive Bayes, and Random Forest) were

used to train the datasets and the best model with the highest accuracy was chosen to predict the object classification (e.g. oil classification) using pre-trained datasets.

Statistical analysis

For any two groups of separation, it is considered as statistically significant when this criterion ($P < 0.5$) is achieved or otherwise denote as non-significant (n.s). The student's unpaired t -test was used throughout this study. One-tailed and two-tailed were used as mentioned in the figure captions. OriginLab-Pro 8 was used to handle all the graphs plotting.

Receiving operating characteristic

The analyses were used to evaluate the specificity and sensitivity of the diagnostic techniques. Various supervised models were used for the ROC tests. These were namely the kNN, Logistic

Regression, Naïve Bayes, Neural Network, and Random Forest models. A fitting of power function $y = ax^b$ were used through the study. Iterations were run with the Levenberg–Marquardt algorithm until a chi-squared tolerance of 10^{-9} was achieved. Final function AUC was compared to the real averaged AUC from all supervised models (details in Supplementary Fig. 2).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

All of the datasets used in these analyses were shared in Supplementary Information or available from the corresponding authors upon request. All the raw data are shared at <https://github.com/VascoRafaelSantos/OliveOil-profiling>.

CODE AVAILABILITY

The code to reproduce the analysis and figures is available from the corresponding authors on reasonable request.

Received: 27 January 2022; Accepted: 30 November 2022;

Published online: 13 December 2022

REFERENCES

- Gimeno, E., Castellote, A., Lamuela-Raventós, R. M., Torre, M. & López-Sabater, M. The effects of harvest and extraction methods on the antioxidant content (phenolics, tocopherol, and carotene) in virgin olive oil. *Food Chem.* **78**, 207–211 (2002).
- Conte, P., Fadda, C., Del Caro, A., Urgeghe, P. P. & Piga, A. Table olives: an overview on effects of processing on nutritional and sensory quality. *Foods* **9**, 514 (2020).
- Silva, M., Freitas, A. M., Cabrita, M. & Garcia, R. Olive oil composition: volatile compounds. In *Olive oil-constituents, quality, health properties and bioconversions*. pp 17–46. <https://tinyurl.com/OliveOilVolatileCompound> (2012).
- Morchio, G., De Anreis, R. & Fedeli, E. Investigations on total sterols content in the olive oil and their variation during the refining process. *Riv. Ital. Sostanze Grasse* **64**, 185 (1987).
- Jimenez-Lopez, C. et al. Bioactive compounds and quality of extra virgin olive oil. *Foods* **9**, 1014 (2020).
- Yorulmaz, A., Erinc, H. & Tekin, A. Changes in olive and olive oil characteristics during maturation. *J. Am. Oil Chemists' Soc.* **90**, 647–658 (2013).
- Gutiérrez, F., Varona, I. & Albi, M. A. Relation of acidity and sensory quality with sterol content of olive oil from stored fruit. *J. Agric. Food Chem.* **48**, 1106–1110 (2000).
- Jabeur, H., Zribi, A., Abdelhedi, R. & Bouaziz, M. Effect of olive storage conditions on Chemlali olive oil quality and the effective role of fatty acids alkyl esters in checking olive oils authenticity. *Food Chem.* **169**, 289 (2015).
- Zhuang, Y. et al. Impact of heating temperature and fatty acid type on the formation of lipid oxidation products during thermal processing. *Front. Nutr.* **9**, 913297 (2022).
- Kiritakis, A. & Markakis, P. In *Advances in Food Research* (eds. Chichester, C. O., Mrak, E. M. & Schweigert, B. S.) vol. 31, 453–482 (Academic Press, 1988).
- Puchades, R. & Maquieira, Á. ELISA tools for food PDO authentication. in *Comprehensive Analytical Chemistry* (eds. de la Guardia, M., González, A., de la Guardia, M. & González, A.) ch. 7, vol. 60, 145–193 (Elsevier, 2013).
- Ok, S. Detection of olive oil adulteration by low-field NMR relaxometry and UV-Vis spectroscopy upon mixing olive oil with various edible oils. *Grasas Aceites* **68**, e173 (2017).
- Meenu, M., Cai, Q., & Xu, B. A critical review on analytical techniques to detect adulteration of extra virgin olive oil. *Trends Food Sci Technol.* **91**, 391–408 (2019).
- Jabeur, H., Drira, M., Rebai, A. & Bouaziz, M. Putative markers of adulteration of higher-grade olive oil with less expensive pomace olive oil identified by gas chromatography combined with chemometrics. *J. Agric. Food Chem.* **65**, 5375–5383 (2017).
- Jabeur, H. et al. Detection of Chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. *J. Agric. Food Chem.* **62**, 4893 (2014).
- Cert, A., Moreda, W. & Garcia-Moreno, J. Determination of sterols and triterpene dialcohols in olive oils using HPLC separation and GC analysis. Standardization of the analytical method. *Grasas Aceites* **48**, 207 (1997).
- Fragaki, G., Spyros, A., Siragakis, G., Salivaras, E. & Dais, P. Detection of extra virgin olive oil adulteration with lampante olive oil and refined olive oil using nuclear magnetic resonance spectroscopy and multivariate statistical analysis. *J. Agric. Food Chem.* **53**, 2810–2816 (2005).
- Li, Y. et al. Detection of olive oil adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and SiPLS. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **189**, 37–43 (2018).
- Baeten, V., Meurens, M., Morales, M. T. & Aparicio, R. Detection of virgin olive oil adulteration by fourier transform raman spectroscopy. *J. Agric. Food Chem.* **44**, 2225–2230 (1996).
- Guimet, F., Ferré, J. & Boqué, R. Rapid detection of olive–pomace oil adulteration in extra virgin olive oils from the protected denomination of origin “Siurana” using excitation–emission fluorescence spectroscopy and three-way methods of analysis. *Anal. Chim. Acta* **544**, 143 (2005).
- Laroussi-Mezghani, S. et al. Authentication of Tunisian virgin olive oils by chemometric analysis of fatty acid compositions and NIR spectra. Comparison with Maghrebian and French virgin olive oils. *Food Chem.* **173**, 122 (2015).
- Vietina, M., Agrimonti, C. & Marmiroli, N. Detection of plant oil DNA using high resolution melting (HRM) post PCR analysis: A tool for disclosure of olive oil adulteration. *Food Chem.* **141**, 3820–3826 (2013).
- Kumar, S., Kahlon, T. & Chaudhary, S. A rapid screening for adulterants in olive oil using DNA barcodes. *Food Chem.* **127**, 1335–1341 (2011).
- Zamora, R., Alba, V. & Hidalgo, F. J. Use of high-resolution 13C nuclear magnetic resonance spectroscopy for the screening of virgin olive oils. *J. Am. Oil Chemists' Soc.* **78**, 89–94 (2001).
- Ogrinc, N., Košir, I. J., Spangenberg, J. E. & Kidrič, J. The application of NMR and MS methods for detection of adulteration of wine, fruit juices, and olive oil. A review. *Anal. Bioanal. Chem.* **376**, 424–430 (2003).
- Peng, W. K. Clustering Nuclear Magnetic Resonance: Machine learning assistive rapid two-dimensional relaxometry mapping. *Eng. Rep.* <https://doi.org/10.1002/eng2.12383> (2021).
- Kong, T. F., Peng, W. K., Luong, T. D., Nguyen, N.-T. & Han, J. Adhesive-based liquid metal radio-frequency microcoil for magnetic resonance relaxometry measurement. *Lab Chip* **12**, 287–294 (2011).
- Peng, W. K., Ng, T.-T. & Loh, T. P. Machine learning assistive rapid, label-free molecular phenotyping of blood with two-dimensional NMR correlational spectroscopy. *Commun. Biol.* **3**, 1–10 (2020).
- Peng, W. K., Chen, L., Boehm, B. O., Han, J. & Loh, T. P. Molecular phenotyping of oxidative stress in diabetes mellitus with point-of-care NMR system. *npj Aging Mech. Dis.* **6**, 1–12 (2020).
- Peng, W. K., Chen, L. & Han, J. Development of miniaturized, portable magnetic resonance relaxometry system for point-of-care medical diagnosis. *Rev. Sci. Instrum.* **83**, 095115 (2012).
- van Beek, T. A. Low-field benchtop NMR spectroscopy: status and prospects in natural product analysis. *Phytochem. Anal.* **32**, 24–37 (2021).
- Curti, E., Carini, E., Cobo, M. F., Bocher, T. & Vittadini, E. The use of two-dimensional NMR relaxometry in bread staling: a valuable tool? *Food Chem.* **237**, 766–772 (2017).
- Dupré, A., Lei, K.-M., Mak, P.-I., Martins, R. P. & Peng, W. K. Micro- and nanofabrication NMR technologies for point-of-care medical applications—a review. *Microelectron. Eng.* **209**, 66–74 (2019).
- Peng, W. K., Han, J. & Loh, T. P. Micro magnetic resonance relaxometry. US Patent 10,393,684 <https://patentimages.storage.googleapis.com/4d/21/78/91c5fcdca2992cd/US10393684.pdf> (2019).
- Peng, W. K. & Paesani, D. Omics meeting omics: towards the next generation of spectroscopic-based technologies in personalized medicine. *J. Personal. Med.* **9**, 39 (2019).
- Francis, B. M. et al. Two-dimensional nanostructures based ‘-omics’ and ‘-omics’ in personalized medicine. *Nanophotonics* <https://doi.org/10.1515/nanoph-2022-0439> (2022).
- Peng, W. K. et al. Micromagnetic resonance relaxometry for rapid label-free malaria diagnosis. *Nat. Med.* **20**, 1069–1073 (2014).
- Unver, T. et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl Acad. Sci. USA* **114**, E9413–E9422 (2017).
- Fernández-Escobar, R. Olive nutritional status and tolerance to biotic and abiotic stresses. *Front. Plant Sci.* **10**, 1–7 <https://www.frontiersin.org/articles/10.3389/fpls.2019.01151/full> (2019).
- Tejada, M. & Benítez, C. Effects of different organic wastes on soil biochemical properties and yield in an olive grove. *Appl. Soil Ecol.* **146**, 103371 (2020).
- Quinn, B. et al. Aggregation in complex triacylglycerol oils: coarse-grained models, nanophase separation, and predicted x-ray intensities. *J. Phys. Condens. Matter* **26**, 464108 (2014).
- Cistola, D. P., Hamilton, J. A., Jackson, D. & Small, D. M. Ionization and phase behavior of fatty acids in water: application of the Gibbs phase rule. *ACS Publ.* <https://doi.org/10.1021/bi00406a013> (2002).

43. Robinson, M. D. & Cistola, D. P. Nanofluidity of fatty acid hydrocarbon chains as monitored by benchtop time-domain nuclear magnetic resonance. *Biochemistry* **53**, 7515–7522 (2014).
44. Whitaker, D. A. & Hayes, K. A simple algorithm for despiking Raman spectra. *Chemometr. Intell. Lab. Syst.* **179**, 82–84 (2018).
45. ISO, E. *Animal and vegetable fats and oils. Determination of acid value matter content.* <https://www.iso.org/standard/75594.html> (2009).
46. Demšar, J. et al. Orange: data mining toolbox in python. *J. Mach. Learn. Res.* **14**, 2349–2353 (2013).

ACKNOWLEDGEMENTS

This research was supported by the Frontier Research Centre of Songshan Lake Materials Laboratory and International Iberian Nanotechnology Laboratory (INL). V.G. would like to thank FCT-DOCTORATES 4 COVID-19 Fellowship Award 2021. We thank Agropecuaria Mabelipa, Pontevedra, Spain, for their grateful donation of their EVOOs for our analysis.

AUTHOR CONTRIBUTIONS

W.K.P. and V.R. conceptualize the idea, W.K.P. and V.R. wrote, revised, and finalized the draft, V.R., V.G., P.S.D., A.C.R., M.T., and B.D. run the entire experimental workload and experimental validation, W.K.P., J.G., and I.P. supervised the project and acquired the source of funding.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41538-022-00173-z>.

Correspondence and requests for materials should be addressed to Weng Kung Peng.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022