Computational and Structural Biotechnology Journal 27 (2025) 1359–1369

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

NMR-based metabolomic approach to estimate chemical and sensorial profiles of olive oil

Gaia Meoni [a,*], Leonardo Tenori [a], Francesca Di Cesare [a], Stefano Brizzolara [b], Pietro Tonutti [b], Chiara Cherubini [c], Laura Mazzanti [c], Claudio Luchinat [d]

[a] Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Florence 50019, Italy
[b] Institue of Crop Sciences, Scuola Superiore Sant'Anna, Pisa 56127, Italy
[c] ANALYTICAL Food, Scandicci, Florence 50018, Italy
[d] Giotto Biotech S.r.l., Sesto Fiorentino, Florence 50019, Italy

ARTICLE INFO

Keywords:
NMR
EVOO
Olive oil quality
Chemometrics
Cultivar
Harvesting
Machine Learning

ABSTRACT

This study investigates the potential of ^{1}H NMR spectroscopy for predicting key chemical and sensory attributes in olive oil. By integrating NMR data with traditional chemical analyses and sensory evaluation, we developed multivariate models to evaluate the predictive power of NMR spectra coupled with machine learning algorithms for 50 distinct olive oil quality parameters, including physicochemical properties, fatty acid composition, total polyphenols, tocopherols, and sensory attributes. We applied Random Forest regression models to correlate NMR spectra with these parameters, achieving promising results, particularly for predicting major fatty acids, total polyphenols, and tocopherols. We have also found the collected data to be highly effective in classifying olive cultivars and the years of harvest. Our findings highlight the potential of NMR spectroscopy as a rapid, non-destructive, and environmentally friendly tool for olive oil quality assessment. This study introduces a novel approach that combines machine learning with ^{1}H NMR spectral analysis to correlate analytical data for predicting essential qualitative parameters in olive oil. By leveraging ^{1}H NMR spectra as predictive proxies, this methodology offers a promising alternative to traditional assessment techniques, enabling rapid determination of several parameters related to chemical composition, sensory attributes, and geographical origin of olive oil samples.

1. Introduction

Olive oil (*Olea europaea L.)* is a cornerstone of the Mediterranean diet since antiquity, and has seen a notable increase in global interest and consumption, including countries without domestic production such as Canada, Japan and China [1]. This growth is largely attributed to its recognized sensory appeal, nutritional and health benefits, which derive from its balanced fatty acid composition - saturated, monounsaturated (MUFA), and polyunsaturated (PUFA) - and its minor components such as polyphenols and tocopherols [2–5]. The unique quality traits and composition of extra virgin olive oil (EVOO) have made it increasingly desirable, fueling a notable surge in global demand in recent years. In the European Union, production is expected to reach 2.5 million tons by 2031, representing a 22 % increase compared to 2020 [6]. This trend highlights the growing global request for high-quality olive oil products that align with environmentally sustainable practices [7]. To safeguard its grade and authenticity, international regulations have been in place since 1991, outlining stringent analytical criteria for olive oil quality assessment (European Communities 1991; International Olive Oil Council 1995). Typically, these assessments rely on traditional analytical methods, which, although relatively accurate, are often labour-intensive, costly, and require extensive sample preparation, chemical-demanding procedures, expensive reference standards, and laborious experimental settings [8–10]. These limitations have spurred interest in novel, rapid, and cost-effective analytical techniques capable of determining multiple quality parameters in a single measurement [11–13]. Methods like Near-Infrared (NIR) spectroscopy, Mid-Infrared (MIR) spectroscopy, fluorescence spectroscopy, and Nuclear Magnetic Resonance (NMR) spectroscopy provide important advantages, including non-destructive sample analysis, minimal preparation, and fast data acquisition [14]. NIR and MIR are praised for their non-destructive, rapid analyses of parameters like fatty acids and

* Corresponding author.
E-mail address: meoni@cerm.unifi.it (G. Meoni).

https://doi.org/10.1016/j.csbj.2025.03.045
Received 12 November 2024; Received in revised form 24 March 2025; Accepted 26 March 2025
Available online 29 March 2025

peroxide value, while fluorescence spectroscopy excels in detecting early oxidation and compositional changes. However, these techniques face challenges such as sample heterogeneity and calibration demands [15–19]. ^{1}H NMR spectroscopy allows the acquisition of the whole sample spectrum, significantly reducing the preparation time and the need for large amounts of organic solvents [20,21]. Importantly, NMR does not alter the sample composition. Another advantage of NMR in this context is its ability to extract valuable information about the oil by also analysing minor components. This analysis can reveal details about adulteration, freshness or oxidation level based on signals related to β-sitosterol, α-linolenic acid, squalene, water, monoglycerides (MG), diglycerides (DG), aldehydes, or hydroperoxides [21–24]. NMR coupled with chemometrics allows for a more comprehensive assessment of olive oil composition and has the potential to improve the speed and accuracy of quality control processes in the olive oil industry [24,25]. While other spectroscopic techniques, such as NIR, MIR, and Raman spectroscopy, have been widely applied in food quality control due to their speed, non-destructive nature, and minimal sample preparation, their effectiveness is often limited to broad compositional trends. These methods are particularly effective for screening bulk properties, such as moisture content, fatty acid ratios, and peroxide value [10,15,19,26–30]. However, when it comes to differentiating minor components, detecting subtle quality markers, or identifying complex molecular interactions, ^{1}H NMR offers a significantly higher chemical resolution. Indeed, NMR directly detects and quantifies molecular structures within the sample, making it a more precise and robust technique. Additionally, while NMR instrumentation may have a higher initial cost than optical spectroscopic methods, the actual cost per analysis is relatively low. The technique requires no expensive reagents, minimal sample preparation, and allows for high-throughput analysis, making it a cost-effective solution for large-scale olive oil quality control and authentication. The ability to extract multiple quality parameters from a single rapid measurement enhances both efficiency and accuracy, positioning ^{1}H NMR as a powerful alternative to both traditional and spectroscopic methods in the evaluation of olive oil authenticity, quality, and traceability. What sets ^{1}H NMR apart is its capacity to offer a truly holistic approach to olive oil analysis [31]. Unlike traditional methods that focus on single parameters or require separate analyses, NMR can simultaneously capture a wide spectrum of chemical and sensory characteristics [20,32]. This approach enhances efficiency while offering deeper insights into the interplay between chemical composition and sensory attributes, which are critical for determining quality and authenticity [33,34]. Furthermore, the non-destructive nature of NMR ensures sample integrity, making it highly suitable for repeated testing and traceability studies, including geographical and botanical origin verification[35].

In this study, we aimed to explore the potential of ^{1}H NMR spectroscopy for the prediction of key chemical and sensory attributes in olive oil. For the first time, we combined NMR data with traditional chemical analyses and sensory evaluations from olive oil samples to develop a multivariate Random Forest regression model capable of simultaneously predicting a wide range of olive oil attributes. These include UV spectrophotometric indices (K232, K268), free fatty acid content (acidity), peroxide value, fatty acid composition, total and specific polyphenols, total tocopherols, as well as sensory qualities such as perceived bitterness, sweetness, pungency, and overall pleasantness.

By integrating NMR data with traditional analytical and sensory evaluation methods, we developed a comprehensive predictive model capable of evaluating 50 distinct olive oil quality parameters in a single measurement. Working with monovarietal EVOOs, we successfully classified olive oils based on the specific olive cultivar used for their production and the year of harvest, leveraging all available analytical data. The results of this study underscore the strong potential of NMR spectroscopy as a powerful tool for olive oil characterization, offering unprecedented insights into its chemical profile, sensory attributes, and authenticity markers. Moving forward, further refinements, such as expanding the dataset to include a broader range of cultivars and environmental conditions, will further enhance the robustness and predictive power of the model. Nonetheless, the integration of ^{1}H NMR with machine learning algorithms, as demonstrated in this study, represents a significant advancement in the field of olive oil authentication and quality control, paving the way for more efficient, accurate, and scalable analytical solutions in the industry.

2. Materials and methods

2.1. Olive oil samples

A total of 140 monovarietal olive oil bottles were collected between 2017 and 2022, stored in their original dark green glass bottles at room temperature (~20 °C) under standard household lighting conditions. Of these, 42 samples were from the Moraiolo cultivar (cv.), 14 from the Frantoio cv., 51 from the Leccino cv., and 4 from the Leccio del Corno cv., while 29 samples had unspecified cultivars. These samples were obtained as part of different research projects, and the olive oils labelled as "monovarietal" were specifically produced within a project (TIMONE, [36]) focused on optimizing olive oil processing. Only olives from distinct, well-identified cultivars were used to ensure the authenticity of these monovarietal samples, which had a certain and controlled origin. Regarding the harvest year, 2 samples were declared from the 2017 harvest, 10 from 2018, 38 from 2019, 28 from 2020, 48 from 2021, and 6 from 2022, while the year of harvest was not indicated for 8 samples. All the 140 samples underwent standard chemical analyses (free acidity, peroxide value, K232, K268, fatty acid composition, tocopherols, and polyphenols), while 80 samples were subjected to sensory evaluation. A total of 340 1D ^{1}H NMR spectra were collected, including triplicate measurements for 100 samples. To ensure the integrity of the analyses, all samples were analyzed either immediately after production or upon receipt or purchase from the market.

2.2. Chemical analyses

The free acidity, peroxide value, UV spectrophotometric indices (K232, K268), and fatty acid composition of the olive oil samples were determined following the standard analytical procedures outlined in EEC Regulation 2568/1991. Tocopherol content was measured according to the ISO 9936:2006/Corr.1:2008 method. All analyses were conducted by Analytical s.r.l., ensuring compliance with established protocols for olive oil quality assessment. UV spectrophotometric indices were determined using a UV-1601 spectrophotometer (Shimadzu, Kyoto, Japan). The fatty acid profile was determined using a GC-2010 (Shimadzu, Kyoto, Japan). Tocopherol quantification was carried out with an LC-20AT/RF-20A Liquid chromatograph/Prominence fluorescence detector (Shimadzu, Kyoto, Japan).

2.2.1. Determination of polyphenols in olive oils by High-performance liquid chromatography (HPLC)

Two grams of olive oil were transferred into a plastic test tube, followed by the addition of 1 mL of an internal standard solution (syringic acid, 1.504 mg mL^{-1} in a MeOH/H_2O 80:20 solution) and 5 mL of a MeOH/H_2O 80:20 solution [37]. The mixture was homogenized by vortexing for 1 minute and then sonicated (YJ5120–1 Ultrasonic Cleaner, Citizen Scale, India) at room temperature for 15 minutes in an ultrasonic bath. Subsequently, the mixture was centrifuged at 5000 rpm for 25 min. An aliquot of the resulting supernatant was drawn through a 5 mL plastic syringe and filtered using a 0.45 μm PVDF filter into a 1.5 mL vial. The filtered solution was immediately subjected to chromatographic analysis. Chromatographic separation was performed using an HP 1100 Liquid Chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a 1100 Autosampler, column heater module, quaternary pump, diode array detector (DAD), and mass spectrometry (MS) detector, with an HP 1100 MSD API-electrospray interface. The column used was a Hypersil Gold QRP-18 (4.6 mm internal

diameter, 250 mm length, 3 μm particle size; Thermo Electron Corporation, Austin, TX, USA), coupled with a 10 × 4 mm pre-column of the same phase, and maintained at 30 °C. Elution was carried out at a flow rate of 0.8 mL min^{-1} using a mobile phase consisting of water at pH 3.2 (adjusted with formic acid) as solvent A, acetonitrile (B), and methanol (C). A three-step linear gradient was applied: 2.5–27.5 % of B and C over 45 min, followed by an increase to 50 % over 10 min. The gradient was then maintained isocratically at 50 % for 5 min. All solvents used were of HPLC grade. Total phenolic compounds were quantified following the COI/T.20/Doc. No. 29 analytical method (COI/T.20/Doc.29, 2009).

2.2.2. Sensory evaluation

The sensory evaluation of 80 olive oil bottles was conducted following a quantitative sensory characterization approach, based on the methodology previously described by the Department of Agriculture, Food and Environment (DAFE) at the University of Pisa [38]. The evaluation was carried out by a panel of trained assessors, selected and trained according to the internal procedure for assessor selection and training, following the principles established by the International Olive Oil Council (IOC). While the evaluation incorporated key elements of the official European Union method for olive oil sensory analysis [39], it was not considered an official panel test for regulatory classification because the panel size and specific methodological adaptations differed from the official EU requirements. Each sample was assessed individually by the panelists, and the final scores were obtained by averaging their evaluations. The eight key sensory attributes considered in this study were: olfactory intensity, fruity aroma, taste intensity, bitterness, pungency, sweetness, developmental stage, and overall pleasantness. Additional descriptors (rancid, fruity, heated, winey, metallic, and frostbitten olive) were excluded from the analysis due to incomplete data.

2.3. ^{1}H NMR analyses

For sample preparation, approximately 60 mg of olive oil (accurately weighted and recorded) was dissolved in 600 μL of deuterated chloroform ($CDCl_3$). $CDCl_3$ provides an effective medium to dissolve lipophilic components of olive oil, ensuring that the oil molecules, including TGs, fatty acids and unsaponifiable fractions, dissolve effectively. $CDCl_3$ also provides a deuterium lock and chemical shift reference (7.24 ppm), ensuring accurate spectral acquisition and analysis [20,40,41]. A 600 μL aliquot of this solution was transferred into a 5 mm NMR tube. NMR spectra were acquired using an AVANCE III 400 MHz Bruker spectrometer, operating at 300 K. For each olive oil sample, the following ^{1}H NMR experiments were conducted [23]: (i) ZG1H, a standard single-pulse experiment consisting of RD-P(90°)-acquisition of the free induction decay (FID), and (ii) NOESYGPPS, a one-dimensional ^{1}H NMR pulse sequence with strong suppression of saturating glycerides signals. For detailed experimental parameters, we refer the reader to Dourou et al. [42]. Phase and baseline corrections were automatically applied using Topspin software (version 3.5 pl 7, Bruker BioSpin Srl). The spectra were aligned to the glycerol signal at 4.30 ppm (-CH_2OCOR). Prior to statistical analysis, each ^{1}D NMR spectrum was divided into 0.02 ppm chemical shift intervals (bins), and the integration of the corresponding spectral areas was performed using an in-house developed R script. This binning or bucketing procedure is a common NMR pre-processing method used to reduce the number of variables and to compensate for small shifts in signals, thereby enhancing data robustness and reproducibility. The binned matrix for the ZG1H spectra covered the full spectral range from 0.2 to 11.9 ppm, excluding the residual chloroform signal at 7.24 ppm, resulting in an array of 580 bins. For the NOESYGPPS spectra, the same spectral range was binned (0.2–11.9 ppm) after removing suppressed saturating signals, reducing the array to 515 bins. To generate a comprehensive spectrum, the bins from 11.90 to 5.62 ppm of the NOESYGPPS experiment were combined with those from 6.20 to 0.30 ppm of the ZG1H experiment. Each resulting binned spectrum was scaled based on the oil weight measured during sample preparation, and total area normalization was applied to complete the 1D spectral processing phase. A total of 39 molecular features in olive oil samples were correctly assigned in all spectra (Fig. 1) using a library of NMR spectra of pure organic compounds, public databases [43] (*e.g.* FooDB, PhytoHub, PhenolExplorer, etc.) storing reference, and literature data [21,44–48].

The quantification of the assigned molecules was performed directly by integrating the signals in the spectra in a defined spectral range, using a house-developed tool. For completeness, the molecules assigned and quantified are listed in Supplementary Table 1, reporting the spectral region of the assignment. The concentration of cycloartenol, Δ7 sterols, β-Sitosterol, squalene, sn-1,2 MG + DG, elenolicacid, oleacein, oleocalnthal, formaldehyde, terpene1, terpene2, terpene3, 1-(S)-(*E*)-elenolide, (*E*)-2alkenals, oleuropein, alkenals, oleokoronal were calculated as a molar percentage of the total TGs, obtained integrating the area of the αCH_2 signal (F, 2.36 ppm), assuming that all fatty acids are esterified and that TGs are the main constituent of EVOO[50,51]. In this way, it is possible to approximately quantify the percentage content (mol/mol_{TGs}) of any metabolite (x) according to the **Equation 1**.

$$\frac{mol}{mol_{TGs}}\% = \frac{Ix \bullet 6}{I\alpha CH_2 \bullet Hx} \bullet 100$$

Equation 1. I=Area of the peak; x = metabolite of interest; H=n° of protons.

Fatty acids values and the calculated NMR concentrations of cycloartenol, Δ7 sterols, β-Sitosterol, squalene, sn-1,2 MG + DG, elenolicacid, oleacein, oleocalnthal, formaldehyde, terpene1, terpene2, terpene3, 1-(S)-(*E*)-elenolide, (*E*)-2alkenals, oleuropein, alkenals, oleokoronal have been used together to perform the multivariate analysis to classify cultivars and years of harvest. Further, univariate analysis was applied to characterize single molecular variations.

2.4. Statistical analysis

To robustly predict n = 42 analytical variables and n = 8 sensory characteristics, a total of 340 ^{1}H NMR spectra were used for the analytical variables, and 214 ^{1}H NMR spectra were used for the sensory characteristics. These represent subsets of the total spectra acquired, selected based on complete analytical and sensory characterization. Random Forest (RF) regression models were constructed for each variable using 2000 decision trees per model, following established methods for predictive modelling in chemometrics [52–55]. RF regression is widely employed in chemometrics due to its robustness in handling complex datasets and reducing overfitting [56]. To ensure the validity of the models and prevent overfitting, an iterative leave-one-group-out cross-validation (LOGO-CV) strategy was implemented. In this scheme, technical replicates of samples from the same bottle were excluded all together from the training set, ensuring that the models were evaluated without influence from similar replicated samples [57]. In each step of this cross-validation strategy, the replicates of one bottle are leave aside as validation set, while all the other samples are maintained in the training set. In this way the overestimation of predictive accuracy is avoided and the model obtained is more reliable because the model does not learn any information from the three replicates excluded during each iteration, in fact when the validation set is tested, in the model there are not samples from that excluded bottle. The predictive accuracy of the RF models was assessed by calculating Pearson's correlation coefficient (R), comparing the actual and predicted values of each variable [58]. Pearson's R provides a measure of the linear relationship between predicted and directly measured values, offering a robust indicator of model performance. The percentage error for each variable was calculated using the following formula (**Equation 2**), where the measured value represents the experimentally measured data, and the predicted value represents the value obtained from the prediction model:

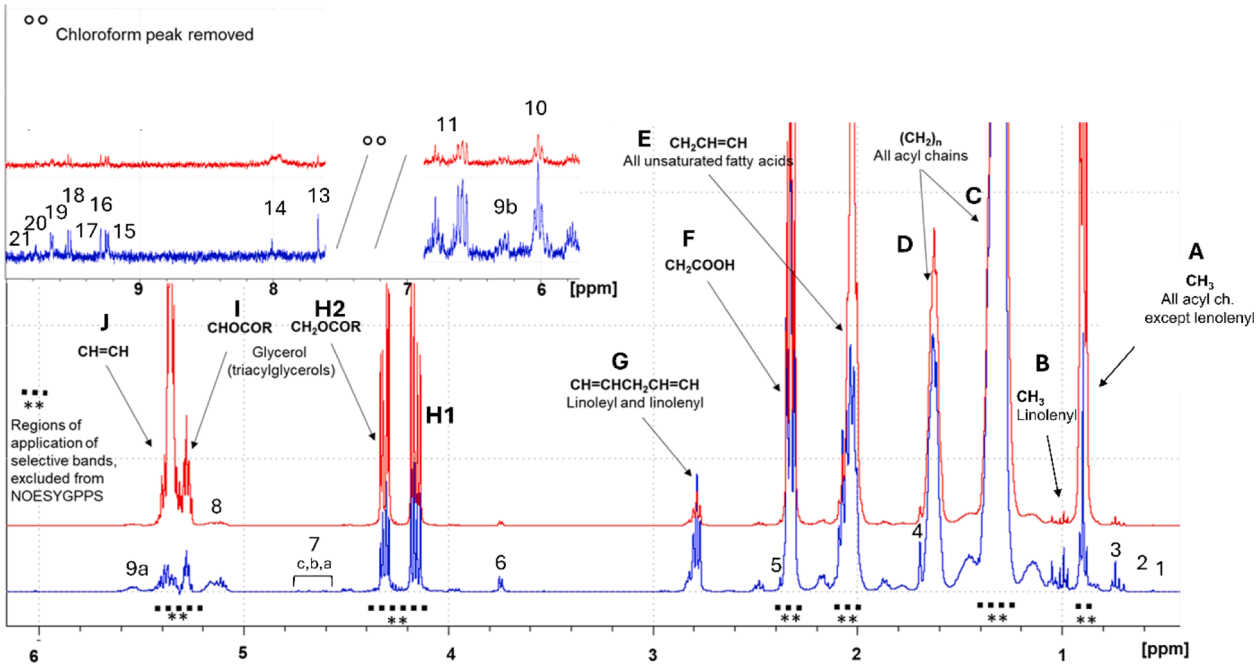

Fig. 1. One-dimensional ^{1}H NMR spectra of olive oil in $CDCl_3$ (adapted from [49]). ZG (red spectrum); NOESYGPPS (blue spectrum). 1) Cycloartenol(*exo*); 2) Cycloartenol (*endo)*; 3) β-Sitosterol/Δ5sterols; 4) Squalene; 5) hydrogenation (saturation); 6) sn-1,2 MG + DG; 7a)Terpene1; 7b) Terpene2 (24-methylenecycloartanol); 7c) Terpene3; 8) sn-1,2DG; 9a) Hydroperoxides ((*Z*,*E*)-conjugated double bonds associated with hydroperoxides (OOH)); 10) Hydroperoxides ((E,E)-conjugated double bonds associated with hydroperoxides (OOH)); 9b) Hydroperoxides; 11) Hydroxytyrosol & derivatives; 12) Tyrosol & derivatives; 13) elenolic acid; 14) Formaldehyde; 15) Oleacein (3,4-DHPEAEDA); 16) Oleocalnthal (pHPEAEDA); 17) 1-(S)-(*E*)-elenolide; 18) (*E*)-2alkenals; 19) oleuropein; 20) oleacein+oleocanthal+elenolicacid; 21) alkanals. A) All acyl chains; B) LN-ω3; C) All acyl chains; D) All acyl chains; E) Unsaturated FA; F) α-CH2; G) Polyunsaturated FA; H1) triacylglycerols; H2) triacylglycerols; I) TGs; J) Polyunsaturated chains.

$$error\% = \frac{Measured\ value - Predicted\ value}{Measured\ value} \bullet 100$$

Equation 2. Error %

Furthermore, the Robust Coefficient of Determination (R robust) was calculated using the "rlm" function of the MASS package (version 7.3–64, R software) [59] to provide a more reliable assessment of model performance in the presence of outliers [60]. Unlike the standard R, which can be overly sensitive to extreme values, R robust reduces the influence of such data points by incorporating robust statistical techniques, such as robust multivariate linear regression. This approach ensures a more stable and realistic evaluation of predictive power. Additionally, the predictive performance of the models was evaluated using the Mean Absolute Error (MAE), which measures the average absolute difference between predicted and actual values [61]. MAE provides an easily interpretable metric of error magnitude, with lower MAE values indicating smaller average errors and, consequently, better model performance. For example, an MAE of 5 would mean that, on average, the model's predictions deviate from the actual values by 5 units. Therefore, minimizing MAE is a key goal when assessing the accuracy of prediction models. The Ratio of Performance to Deviation (RPD) was also used to assess the predictive capability of the models. RPD is calculated as the ratio between the standard deviation of the reference data and the RMSEP. It provides an indication of the model's ability to predict unknown samples. Higher RPD values indicate better predictive performance, with the following interpretation thresholds: RPD < 1.5 indicates poor model performance, not recommended for prediction; $1.5 \leq$ RPD < 2.0 indicates acceptable performance; RPD ≥ 2.0 indicates excellent predictive ability [62]. To evaluate the performance of RF models built for cultivars and year of harvest classifications, accuracy (%) and F-score (%) were calculated as a measure of model performance. Accuracy (%) is defined by the following **Equation 3**:

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Equation 3. FP = false positives, TP = true positives, FN = false negatives, and TN = true negatives.

The F-score is the harmonic mean of precision and sensitivity, where precision is defined as the ratio of true positives to the sum of true positives and false positives, and sensitivity is the ratio of true positives to the sum of the true positives and false negatives (**Equation 4**):

$$F\% = 2 \bullet \frac{Precision \bullet Sensitivity}{Precision + Sensitivity} \bullet 100$$

Equation 4. F-score %

Perfect performance yields an F-score of 100 %.

This metric was used to assess the models' ability to accurately predict categorical sensory data, ensuring balanced performance across classes, especially when dealing with imbalanced datasets or small sample sizes. The combined use of Pearson's R for regression accuracy and the F-score for classification provides a comprehensive evaluation of the predictive model's robustness and effectiveness in forecasting olive oil quality parameters. All analyses were performed with R software (version 4.1.0) [63]. The randomForest package [52,53] was used to build RF models.

3. Results

3.1. Prediction of chemical analyses

The nutritional benefits of olive oil are primarily attributed to its high content of monounsaturated and polyunsaturated fatty acids [64], and minor polar compounds with strong antioxidant properties [65,66]. In this study, we evaluated whether the full ^{1}H NMR spectra of olive oil could be used to accurately predict the concentration of these key

compounds measured with standard analytical protocols described in the Section 2.2, which are used to define the chemical profile of olive oil. RF regression models were applied to correlate whole bucketed ^{1}H NMR spectra with the analytical values of 42 chemical parameters measured in olive oil. Separate models were constructed for each parameter. Table 1 reports the directly quantified and NMR-predicted values, the standard deviation (SD), Pearson's correlation coefficient (R) and R^2, P-values (*P*), and the median error percentage (error %), MAE and RPD. Further performance parameters are reported in Supplementary Table 2. The regulated physicochemical quality parameters, including free fatty acids, peroxide value, and UV absorption characteristics (K268 and K232), were predicted from the ^{1}H NMR bucketed spectra, yielding R values of 0.71, 0.78, 0.87, and 0.76, respectively, with corresponding error percentages of 14.1 %, 14.6 %, 5.8 %, and 9.4 % (Table 1, Supplementary Figure 1). Additionally, individual fatty acid components were analysed. Olive oil primarily consists of oleic acid (73 % m/m, percentages calculated from the analytical data collected on the olive oil samples analysed in this study and reported as mean value in Table 1), palmitic acid (14 % m/m), linoleic acid (8 % m/m), and stearic acid (2 % m/m), along with smaller proportions of linolenic and palmitoleic acids (1 % m/m each). The prediction models for these fatty acids showed R values ranging from 0.71 (palmitoleic acid) to 0.91 (oleic acid), with error percentages ranging from 0.54 % (oleic acid) to 8.5 % (palmitoleic acid) (Table 1, Supplementary Figure 2). Among the less abundant fatty acids, arachidic acid (0.33 % m/m) was predicted with an error of 4.4 %, while eicosanoic (0.28 % m/m) and behenic (0.09 % m/m) acids had errors of 5.9 % and 7.3 % and R values of 0.85, 0.75, and 0.71, respectively. Other quantified fatty acids, such as margaroleic acid (0.08 % m/m), lignoceric acid (0.04 % m/m), and margaric acid (0.03 % m/m), exhibited R values of 0.58, 0.74, and 0.8, with error percentages of 6.9 %, 16.6 %, and 10.1 %, respectively. RF models for total polyphenols and their individual components were also evaluated (Table 1, Supplementary Figure 3). Most detectable polyphenols in olive oil include simple phenolic acids (*e.g.,* gallic, vanillic, benzoic, cinnamic, caffeic, and coumaric acids), phenyl alcohols (*e.g.,* hydroxytyrosol, tyrosol), secoiridoids (*e.g.,* oleuropein, ligstroside), flavonoids (*e.g.,* apigenin, luteolin), and lignans (*e.g.,* pinoresinol, 1-acetoxypinoresinol). Among these, secoiridoids and lignans are the most abundant phenolic compounds in EVOO. The most abundant secoiridoids in olives are oleuropein and ligstroside, which degrade during olive oil production into various aglycone forms (*e.g.,* DHPEA-EA, p-HPEA-EA, DHPEA-EDA, p-HPEA-EDA) and their oxidized derivatives [67]. Phenolic concentrations in olive oil are influenced by several factors, primarily the olive growing conditions and harvesting time, and the oil extraction protocols

Table 1
RF regression models performed on chemical parameters measured on olive oil. For each analytical parameter the mean of the directly quantified value (analytic.) and the spectrum-specific predicted values (pred.) ± the standard deviation (sd), the Pearson's R and R^2, the P-value (*P* * *<0.001), median of the errors (real measured-predicted/real measured*100) expressed in % (e %), Mean Absolute Error (MAE) and Ratio of Performance to Deviation (RPD) are reported.

Parameters (Units)	mean±sd (analytic.)	mean±sd (pred.)	R	R^2	e %	*MAE*	*RPD*	*P*
Free acidity (%oleic acid)	0.31 ± 0.13	0.3 ± 0.07	0.71	0.50	14	0.06	1.37	* *
Peroxide index (meq/kg)	10.9 ± 4.7	10.7 ± 3.22	0.78	0.61	15	2.04	1.59	* *
UV (nm)								
K 232	2.1 ± 0.52	2.1 ± 0.38	0.87	0.75	5.8	0.18	1.92	* *
K 268	0.14 ± 0.04	0.14 ± 0.02	0.76	0.58	9.4	0.02	1.50	* *
Methyl esters of fatty acids (%m/m)								
Palmitic acid (C16:0)	13.82 ± 1.21	13.83 ± 0.89	0.87	0.76	2.4	0.46	1.96	* *
Palmitoleic acid (C16:1)	1.06 ± 0.22	1.05 ± 0.12	0.71	0.50	8.5	0.12	1.39	* *
Margaric acid (C17:0)	0.03 ± 0.01	0.03 ± 0.012	0.8	0.64	10	0.01	1.63	* *
Margaroleic acid (C17:1)	0.08 ± 0.02	0.08 ± 0.01	0.58	0.34	6.9	0.01	1.21	* *
Stearic acid (C18:0)	2.19 ± 0.25	2.19 ± 0.16	0.8	0.63	3.5	0.11	1.61	* *
Oleic acid (C18:1)	72.97 ± 1.65	72.97 ± 1.24	0.9	0.81	0.5	0.55	2.18	* *
Linoleic acid (C18:2)	8.29 ± 1.0	8.30 ± 0.76	0.91	0.83	2.9	0.32	2.28	* *
Linolenic acid (C18:3)	0.79 ± 0.14	0.79 ± 0.10	0.85	0.72	4.7	0.05	1.85	* *
Arachidic acid (C20:0)	0.33 ± 0.06	0.33 ± 0.04	0.85	0.73	4.4	0.02	1.86	* *
Eicosanoic acid (C20:1)	0.28 ± 0.06	0.28 ± 0.03	0.75	0.56	5.9	0.02	1.45	* *
Behenic acid (C22:0)	0.09 ± 0.02	0.09 ± 0.01	0.71	0.51	7.3	0.01	1.40	* *
Lignoceric acid (C24:0)	0.04 ± 0.02	0.04 ± 0.01	0.74	0.54	17	0.01	1.46	* *
Profile of polyphenols (mg/kg)								
DHPEA	10.1 ± 13.72	8.73 ± 4.81	0.75	0.56	37	4.82	1.28	* *
pHPEA	9.43 ± 12.97	8.18 ± 4.56	0.67	0.45	39	4.86	1.23	* *
Vanillic acid+Caffeic acid	1.42 ± 0.80	1.33 ± 0.32	0.42	0.17	30	0.51	1.09	* *
Vanilline	2.52 ± 1.23	2.46 ± 0.53	0.47	0.22	22	0.80	1.13	* *
p-coumaric acid	1.58 ± 0.90	1.48 ± 0.38	0.42	0.18	30	0.57	1.10	* *
3,4-DHPEA-AC	1.55 ± 1.16	1.45 ± 0.5	0.48	0.23	37	0.69	1.14	* *
Ferulic acid	2.05 ± 1.30	1.94 ± 0.67	0.54	0.29	34	0.79	1.18	* *
o-coumaric acid	1.88 ± 1.09	1.78 ± 0.45	0.38	0.15	32	0.68	1.08	* *
DHPEA-EDAox	5.25 ± 6.32	4.90 ± 3.35	0.75	0.57	46	2.51	1.44	* *
DHPEA-EDA	57.53 ± 60.1	57.41 ± 47.7	0.93	0.87	34	17.47	2.58	* *
DHPEA-EAglu	19.38 ± 34.46	17.91 ± 18.13	0.82	0.67	61	10.94	1.56	* *
DHPEA-EA	43.62 ± 46.90	42.35 ± 30.11	0.75	0.55	55	4.46	1.47	* *
pHPEA-EDAox	15.28 ± 17.44	15.08 ± 11.74	0.87	0.76	34	5.89	1.90	* *
pHPEA-EDA	42.11 ± 26.89	42.15 ± 18.9	0.78	0.61	24	12.30	1.59	* *
Pinoresinol, 1-acetoxypinoresinol	21.31 ± 17.09	20.51 ± 11.15	0.78	0.62	25	6.73	1.58	* *
Cinnamic acid	7.50 ± 8.62	6.75 ± 4.22	0.73	0.53	40	3.50	1.37	* *
p-HPEA-EA	8.80 ± 13.08	7.71 ± 6.83	0.78	0.61	53	4.70	1.47	* *
DHPEA-EA-OH	10.63 ± 12.62	9.94 ± 7.67	0.75	0.56	42	22.41	1.48	* *
p-HPEA-EA-OH	12.58 ± 13.62	11.4 ± 6.03	0.73	0.53	35	5.65	1.34	* *
DHPEA-EA-OHox	15.08 ± 8.17	14.64 ± 5.22	0.74	0.55	26	4.19	1.47	* *
p-HPEA-EA-OHox	49.73 ± 45.55	46.91 ± 26.10	0.71	0.50	32	20.04	1.39	* *
Luteolin	5.48 ± 4.39	4.98 ± 2.03	0.55	0.30	45	2.50	1.18	* *
methyl-luteolin	7.99 ± 5.29	7.81 ± 3.01	0.66	0.43	34	3.08	1.32	* *
apigenin	5.89 ± 4.13	5.47 ± 1.80	0.51	0.26	34	2.50	1.15	* *
Total polyphenols (mg/kg)	359.9 ± 161.8	360.8 ± 137	0.93	0.86	11	48.09	2.62	* *
Tocopherols (mg/kg)	303.4 ± 105.8	303.8 ± 82.4	0.89	0.78	10	38.32	2.10	* *

and storage. EVOO generally has a higher total phenol content than refined virgin olive oil, with values typically ranging from 100 to 800 mg/kg [68]. The model for total polyphenols demonstrated an optimal R value of 0.93 with an error percentage of 11.2 %. Tocopherol concentrations in EVOO range from 97 to 785 mg/kg, contributing to the protection of lipids against peroxidation and enhancing the oxidative stability of olive oil in combination with polyphenols [69]. Our model for tocopherol content achieved an R value of 0.89 with an error percentage of 10.3 % (Table 1, Supplementary Figure 4). In evaluating model performance, the most reliable predictions were observed for total polyphenols ($R^2 = 0.86$, RPD = 2.62), peroxide index ($R^2 = 0.61$, RPD = 1.59), K232 ($R^2 = 0.75$, RPD = 1.92), oleic acid ($R^2 = 0.81$, RPD = 2.18), linoleic acid ($R^2 = 0.83$, RPD = 2.28), and DHPEA-EDA ($R^2 = 0.87$, RPD = 2.58), and tocopherols ($R^2 = 0.78$, RPD = 2.1). These compounds are key indicators of olive oil quality and stability, reinforcing the robustness of the predictive model for these parameters. Moderate predictions were obtained for compounds such as DHPEA-EAglu ($R^2 = 0.67$, RPD = 1.56), pHPEA-EDAox ($R^2 = 0.76$, RPD = 1.9), cinnamic acid ($R^2 = 0.53$, RPD = 1.37), suggesting need for improvement with refined models or expanded datasets. However, weak performances were noted for p-coumaric acid ($R^2 = 0.18$, RPD = 1.10), ferulic acid ($R^2 = 0.29$, RPD = 1.18), o-coumaric acid ($R^2 = 0.15$, RPD = 1.08), luteolin ($R^2 = 0.30$, RPD = 1.18), and methyl-luteolin ($R^2 = 0.43$, RPD = 1.32), likely due to higher variability in their concentrations or insufficient data coverage. Additionally, the R^2 robust coefficient was calculated and presented in Supplementary Figure 5 and Supplementary table 2 to assess cases where data points were clustered in specific regions with outliers at extremes. Notably, in some instances, such as margaroleic acid and DHPEA-EA, the robust R^2 exhibited better performance, implying that increasing the number of samples and covering a broader range of values could enhance model accuracy. These findings highlight the effectiveness of ^{1}H NMR-based prediction models for key olive oil compounds while underscoring areas where further refinement could optimize predictive capabilities.

3.2. Prediction of sensory attributes

The RF regression approach was also applied to relate the ^{1}H NMR bucketed spectra to the sensory profiles of olive oil samples. The results of the RF regression models for each sensory attribute are summarized in Table 2 and Supplementary Table 3, with corresponding plots provided in Supplementary Figure 6. Independent models were constructed for each sensory descriptor, including bitterness, sweetness, pungency, taste intensity, fruity aroma, olfactory intensity, overall pleasantness, and developmental stage, using a dataset of 84 distinct olive oil samples. The NMR-based models yielded promising results for specific sensory characteristics. Notably, the model for olfactory intensity achieved a Pearson's correlation coefficient R= 0.76 with a prediction error of 19.2 % and a MAE of 0.69 indicating good accuracy. Additionally, the RPD value of 1.45 suggests that the model provides a reliable and acceptable level of predictive performance for this variable. Similarly, the model for bitterness, an attribute that is generally linked to the level of polyphenols, demonstrated a strong correlation with R= 0.75 and an error of 14.9 %, a MAE of 0.97, and an RPD of 1.47, indicating reliable predictive performance. Likewise, the model for overall pleasantness achieved a correlation of R= 0.70, with a prediction error of 15.4 %, a MAE of 1.14, and an RPD of 1.33, suggesting a good level of accuracy in capturing this complex sensory attribute. Conversely, the predictions for pungency, gustatory intensity, and developmental stage were moderate, explaining less than 30 % of the variation, but still providing marginally useful results. Sweetness and olfactory fruity aroma had the weakest predictive performance, with R$\leq$ 0.27, high MAE, and RPD $<$ 1.5, indicating that the models are unreliable for these attributes, likely due to their dependence on compounds that are either present at very low concentrations or not detectable by NMR.

3.3. Direct determination of olive oil composition from NMR spectra

In this study, we also present a comprehensive spectral assignment and concentration analysis of various metabolites identified in the NMR spectra of olive oil samples. Detailed information regarding the assigned metabolites is provided in Supplementary Table 1, which delineates the specific spectral regions associated with each compound. The metabolites quantified in this analysis include cycloartenol, Δ7 sterols, β-sitosterol, squalene, α-CH2, sn-1,2 mono- and diglycerides (MG + DG), elenolic acid, oleacein, oleocanthal, formaldehyde, and three distinct terpenes (terpene1, terpene2 and terpene3). Additionally, we quantified 1-(S)-(*E*)-elenolide, (*E*)-2-alkenals, oleuropein, alkenals, and oleokoronal. The concentrations of these metabolites were calculated in accordance with Equation 1, as detailed in Section 2.3. The presence and quantification of these metabolites can be effectively determined from the ^{1}H NMR spectra of olive oil in $CDCl_3$, using the integrals of the peaks corresponding to the metabolites of interest, alongside the total TG content, which is derived from the area of the α-CH2 peak at 2.36 ppm in the NMR spectra. This methodology allows for a reliable assessment of the metabolite composition in olive oil, contributing valuable insights into its biochemical profile. For the evaluation of FA composition, we employed various mathematical equations described by [44], as detailed in Supplementary Table 4. The equations used for FA quantification were selected based on their reported accuracy and applicability. Fig. 1 displays the olive oil spectra acquired using the two NMR pulse sequences, along with the assignment of proton signals, as summarized in Supplementary Table 1. To identify the equations that best correlate with FA profiles determined by GC-FID, a correlation analysis was performed, the results of which are shown in Table 3. As presented in Table 3, equations 3 and 4 exhibit identical Pearson correlation coefficients (R = 0.995) for the prediction of linoleic acid concentration. However, model 4 shows a lower percentage error (6.25 %) compared to model 3 (11.34 %), indicating superior predictive accuracy. Conversely, the quantification of linolenic acid proved to be the most challenging, with the percentage error between NMR-calculated values and GC-FID measurements reaching 69.65 %. For oleic acid, equation 2 provided the most accurate results, with a low error rate of 4 %. In the case of SFA, model 4 demonstrated the best performance, yielding an error of

Table 2
RF regression models performed on sensorial profiles of olive oils (score points 1–10). For each analytical parameter the mean of the directly quantified value (analytic.) and the spectrum-specific predicted values (pred.) ± the standard deviation (sd), the Pearson's R, the *P*-value (*P* * *<0.001) and the median of the errors (real measured-predicted/real measured *100) expressed in % (e %) are reported.

Parameters	mean±sd (analytic.)	mean ± sd (pred.)	R	R2	e%	*MAE*	*RPD*	*P*-value
Bitter	2.72 ± 1.59	2.79 ± 1.18	0.75	0.56	14.9	0.97	1.47	* *
Sweet	1.84 ± 0.97	1.82 ± 0.36	−0.05	0.003	36.7	0.99	0.93	0.45
Pungent	3.31 ± 1.64	3.37 ± 0.69	0.48	0.23	22.2	1.2	1.12	* *
Gustatory intense	3.69 ± 0.89	3.75 ± 0.51	0.53	0.28	14.1	0.64	1.1	* *
Olfactory fruity	2.97 ± 1.13	3.03 ± 0.63	0.27	0.07	19.2	0.87	1.04	* *
Olfactory intense	3.42 ± 1.38	3.45 ± 1.04	0.76	0.58	19.1	0.69	1.45	* *
Overall pleasantness	5.23 ± 1.95	5.29 ± 1.24	0.70	0.49	15.4	1.14	1.33	* *
Developmental stage	2.36 ± 1.65	2.21 ± 0.75	0.55	0.30	35.9	1.02	1.19	* *

Table 3
Correlation (R) value of GC-FID measured FA and NMR mathematical equations [1–6] as reported by [44].

	1	2	3	4	5	6
Linoleic	0.992	0.992	0.995	0.995		
Linolenic	0.708	0.708	0.708	0.648		
Oleic	0.974	0.975	0.945	0.944		
SFA	0.838	0.854	0.711	0.871	0.499	0.431
MUFA	0.974	0.975	0.946	0.944	0.975	0.450
PUFA	0.991	0.992	0.995	0.994	0.992	0.969

8.84 %. The quantification of MUFA was highly accurate using both equations 2 and 5, with error percentages of 2.11 % and 2.07 %, respectively. For PUFA, equation 3 achieved the best correlation with an error of 5.12 %.

3.4. Characterization of cultivars and year of harvest

RF classification models were also employed to classify olive oil cultivars (*Leccino cv., Moraiolo cv., Leccio del Corno cv., and Frantoio cv.*) and the year of harvest. The models were built using various data sources: the bucketed NMR spectra matrix, the NMR-quantified molecules matrix, the standard analytical values matrix, and combinations of these datasets. The performance of the RF models for cultivar classification is reported in Table 4, while the classification performance for the year of harvest is presented in Table 5. The aim of this analysis was to determine whether these two critical variables (cultivar and harvest year) could be identified quickly using a single comprehensive analysis, such as the entire NMR spectrum or a combination of molecular and analytical data. This could be a valuable tool to guarantee the authenticity of single-varietal oils and to study the influence of the harvest year on oil quality, thus promoting the optimization of agronomic practices and extraction processes. The distinctive characteristics of a particular olive oil are influenced by a combination of genetic makeup, environmental conditions, agricultural practices, and harvest period [23,24,70]. Consequently, tools capable of ascertaining details about the origin of olive oils are gaining importance in the field of olive oil authentication. To avoid classification bias due to sample similarity, RF models were constructed by iteratively and randomly excluding all technical replicates of the same olive oil samples from the training set. Among the proposed models, those utilizing NMR-quantified metabolites, either alone or in combination with analytical data, consistently demonstrated the highest classification accuracy for olive cultivars, reaching 92.9 % overall. Leccino, Moraiolo, and Frantoio cultivars were particularly well-classified, with accuracy exceeding 90 %, while Leccio del Corno exhibited an accuracy of 75 %. For harvest year classification, both the analytical data model and the combined NMR + analytical model achieved similar performances, with classification accuracies of 96.1 % and 97.3 %, respectively. However, it was not possible to evaluate and achieve accurate results for the year 2017, as only two samples were available for model training. Overall, the results suggest that a model based on a combination of NMR-identified metabolites and standard analytical data is comprehensive enough to accurately classify both the olive oil cultivar and the year of harvest. To further ensure that the high classification accuracy for harvest year recognition was not influenced by the inclusion of heterogeneous samples, an additional analysis was performed using a more homogeneous dataset. Specifically, three distinct models were built using samples from Moraiolo, Leccino, and Frantoio cultivars, all collected over multiple years from the same producer. The results, reported in Table 6, confirm an even stronger classification performance, with accuracies of approximately 99 % for the three cultivars. The individual classification accuracies were 98 % for Moraiolo samples collected in 2019, 2020, and 2021, 94 % for Leccino, and 100 % for Frantoio samples collected in the same years. These findings further reinforce the robustness and discriminative power of the NMR-based approach in capturing inter-annual variations in olive oil composition. The ability to distinguish harvest years with such high accuracy, even within oils derived from the same cultivar and producer, suggests that subtle chemical differences associated with environmental factors, maturation conditions, and seasonal variations are effectively detected by the model.

Table 4
RF classification models performed to classify olive oil cultivars built on NMR bucketed spectra, NMR quantified molecules, molecules quantified with standard analytical procedures (Analytical) and model created by combining NMR quantified and analytical molecules. Accuracy and F-score are reported.

	accuracy%; F-score%			
cultivar	**NMR spectra**	**NMR molecules**	**Analytical**	**NMRmol+Ana.**
Leccino	92.1; 62.0	92.8; 62.0	96.1; 64.0	93.5; 63.0
Moraiolo	94.4; 79.0	95.2; 83.0	92.1; 74.0	95.2; 81.0
Leccio del Corno	66.7; 80.0	75.0; 86.0	75.0; 86.0	75.0; 86.0
Frantoio	95.2; 97.0	97.6; 89.0	69.0; 79.0	92.9; 96.0
Overall accuracy	81.8	92.9	89.6	92.6

Table 5
RF classification models performed to classify olive oil harvest year built on NMR bucketed spectra, NMR quantified molecules, molecules quantified with standard analytical procedures (Analytical) and model created by combining NMR quantified and analytical molecules. Accuracy and F-score are reported.

	accuracy%; F-score%			
year	**NMR spectra**	**NMR molecules**	**Analytical**	**NMRmol+Ana.**
2017	0;0	50.0; 63.0	0; 0	0; 0
2018	20.0; 18.0	40.0; 56.0	40.0; 44.0	40; 33.0
2019	99.1; 76.0	92.1; 82.0	97.4; 67.0	100; 83.0
2020	91.7; 94.0	89.3; 90.0	96.4; 93.0	98.8; 99.0
2021	97.2; 87.0	100; 63.0	100; 95.0	100; 91.0
2022	100; 100	55.6; 71.0	100; 100	100; 100
Overall accuracy	93.5	91	96.1	97.3

4. Discussion

Ensuring the verification and traceability of olive oil presents a significant challenge due to its complex and variable chemical composition, which is influenced by cultivar, environmental conditions, and

Table 6
Confusion matrices for three RF classification models used to distinguish olive oil harvest years based on NMR bucketed spectra. The models were built for three cultivars:(A) Moraiolo, (B) Leccino, and (C) Frantoio, harvested from the same producer in three different years (2019, 2020, and 2021). The table displays the number of bottles/samples (n°/s), their actual harvest year, and the percentage of correctly classified bottles for each year. The overall classification accuracy for each model is reported in the last row.

n°/s	**A**	**2019**	**2020**	**2021**
14/28	**2019**	100	0	0
10/30	**2020**	6.7	93.3	0
18/54	**2021**	0	0	100
Accuracy	98.2 %			
n°/s	**B**	**2019**	**2020**	**2021**
14/28	**2019**	90.5	0	9.5
12/36	**2020**		83.3	16.7
25/75	**2021**	0	0	100
Accuracy	93.8 %			
n°/s	**C**	**2019**	**2020**	**2021**
4/8	**2019**	100	0	0
6/18	**2020**	0	100	0
4/12	**2021**	0	0	100
Accuracy	100 %			

production methods. While traditional analytical methods, including titrations for acidity, GC for fatty acids, and HPLC for polyphenols, are well-established and validated by regulatory bodies such as the IOC, they require extensive sample preparation, large sample volumes, and organic solvents, raising environmental concerns [71,72]. To assess the quality of virgin olive oil, the IOC endorses a panel test that evaluates the organoleptic properties of the oil through sensory evaluation. This method involves trained panelists who assess different characteristics. While effective in identifying adulterations (such as the addition of refined oils or synthetic mixtures) this approach falls short of definitively determining the authenticity of olive oil with respect to its geographical and botanical origins [73]. Furthermore, as highlighted by Circi et al. [74] and Barbieri et al. [73], the panel test often lacks reproducibility when applied to commercial extra virgin olive oils with chemical parameters near regulatory thresholds. Their study demonstrated significant inconsistencies among different IOC-recognized laboratories, where the same olive oil sample was classified as extra virgin, virgin, or even lampante depending on the laboratory conducting the sensory evaluation. This discrepancy raises concerns about the subjectivity and variability of sensory assessments, particularly in cases where the chemical profile does not strongly distinguish between quality grades. The introduction of ^{1}H NMR spectroscopy in olive oil quality assessment marks a significant advancement in analytical efficiency and sustainability [10]. NMR spectroscopy allows for rapid and reliable data acquisition while preserving the chemical environment of the analyte. This method eliminates the need for labour-intensive separation processes and reduces chemical waste, facilitating the simultaneous evaluation of multiple parameters without complex instrumental setups [24,71,75,76]. The increasing demand for stringent quality controls and product traceability in the food sector creates an ideal environment for the development of NMR-based techniques, which have proven effective in identifying various substances, uncovering complex frauds, and studying the effects of food processing and storage, as well as confirming labels and geographic provenance [20,77–82]. The application of RF regression models to predict standard analytical values from NMR spectra represents a promising avenue for fast and non-destructive analysis. Previous research supports the efficacy of spectroscopic techniques in characterizing olive oil [9,14,75]. Moreover, there is a growing number of studies based on the use of spectroscopic techniques that offer an untargeted approach to predicting qualitative attributes directly from a measurement [26,27,83–85]. Notably, to the best of our knowledge, this study is the first to utilize RF decision trees for predicting the chemical composition of olive oil across a broad spectrum of qualitative parameters, offering a novel approach to evaluating olive oil quality and composition. The application of RF regression models to predict physicochemical quality parameters from ^{1}H NMR spectra yielded promising results, with a relative error of less than 15 % for 17 chemical parameters. Specifically, it has proven to be an effective tool for the rapid and accurate prediction of major fatty acids (C16:0, C16:1, C18:0, C18:1, C18:2, C18:3). This is particularly relevant given the importance of accurate fatty acid profiling in determining olive oil quality and authenticity [86,87]. By leveraging the comprehensive information contained within the entire NMR spectrum, the RF models exhibited superior quantification performance compared to methods based on isolated spectral regions. This underscores the importance of utilizing the complete spectral fingerprint for extracting quantitative information, consistent with findings from other studies employing similar methodologies for olive oil analysis [88]. The results for the prediction of total polyphenols and tocopherols highlight the potential of NMR spectroscopy as a rapid assessment tool for olive oil quality, which is essential for both producers and consumers [89,90]. While NMR demonstrated correlations with several chemical and sensory attributes, predicting certain sensory characteristics, such as sweetness and pungency, showed lower accuracy. This highlights the need for a more diverse dataset that encompasses a wider range of sensory variations, potentially including oils from different geographical and climatic regions. Our findings suggest that NMR, with its holistic analytical approach, has the potential to offer a more comprehensive and nuanced representation of sensory perception compared to the analysis of individual compounds. However, the challenge in predicting sensory attributes underscores the importance of expanding the dataset to capture greater organoleptic variability. Indeed, our results suggest that NMR, with its ability to analyse a sample holistically, could provide a more comprehensive and human-like description of sensory perception, compared to the analysis of individual compounds. However, predicting sensory attributes proved to be more challenging, indicating a need for an expanded dataset characterized by greater organoleptic variability. Despite this limitation, the results affirm the potential of NMR as a rapid analytical tool for assessing olive oil quality. Additionally, the exploration of cultivar classification and harvest year using RF models contributes to the growing literature on the authenticity and traceability of olive oils [24,91–93]. In this study, by comparing various available data, we demonstrate that NMR coupled with machine learning algorithms, such as RF, significantly contributes to improving the predictive accuracy of models, compared to the use of analytical variables alone. While this study showcases the potential of NMR offering valuable insights into the chemical and sensory profiling of olive oil, several limitations need to be addressed to enhance the robustness and the generalizability of the findings. One limitation is given by the limited sample size for certain cultivar (such as Leccio del Corno) and certain harvest year (such as 2017). A small dataset may reduce the statistical power of the predictive models and limit their applicability to a broader range of olive oils. Our results also indicate variability across different cultivars and year of harvest. For example, a few 2020 Moraiolo bottles were misclassified as 2019 bottles, suggesting that the annual variability for this cultivar may not be as pronounced as for others, such as Frantoio, which showed perfect 100 % classification. Similarly, for the Leccino cultivar, the variability between years was more pronounced, although some samples were confused between 2019 and 2020. These observations suggest that RF models may be sensitive to inter-annual variation in chemical characteristics, but also that certain cultivars, such as Leccino and Moraiolo, could present additional challenges in situations of high variability, requiring a larger dataset to further improve classification accuracy. The reasons for this confusion could include the overlap of similar chemical compositions between harvest years or variability related to environmental factors and agronomic practices, which may not be fully captured by the current models. Another significant challenge concerns the variability in sensory panel evaluations, which can introduce a degree of subjectivity into the dataset. However, it is important to highlight that NMR-based predictions surpass the inherent inconsistencies of human sensory panels, which are often influenced by individual biases and environmental factors. By integrating chemical and spectral data, this approach provides a more consistent and objective alternative, reducing the reliance on subjective panel assessments. While refining sensory evaluation protocols or expanding panel training could further improve reproducibility, the use of NMR as an analytical tool represents a substantial advancement in olive oil quality assessment. Despite the strong performance of the NMR-based model, certain sensory attributes, such as sweetness, exhibited lower predictive accuracy, with R values close to zero. This limitation likely arises from three key factors. First, the complexity of sensory perception: unlike bitterness or pungency, which correlate with well-defined chemical markers such as polyphenols, sweetness may be influenced by compounds present in extremely low concentrations, masked by dominant molecules [74]. Second, dataset constraints: the limited variability in sweetness scores across the dataset may have impeded the ability to train a robust predictive model. Future studies should incorporate a broader range of sensory profiles to improve prediction accuracy for subtle attributes. Third, panel subjectivity: sweetness is a highly nuanced and context-dependent sensory attribute, making it more susceptible to individual perception and bias. Such variability in training data further complicates the development of precise predictive models. Predicting

sensory parameters is inherently more complex than predicting chemical compounds, even in cases where they cannot be directly quantified via NMR. While some sensory attributes exhibit clear correlations with specific spectral peaks or their combinations, allowing for more reliable predictions, others are influenced by compounds present at very low concentrations, often below the detection limits of standard NMR sensitivity. This limitation makes it difficult for the model to capture the full spectrum of sensory perceptions. Furthermore, certain attributes may arise from complex molecular interactions rather than from the presence of specific, easily identifiable markers. In such cases, the model hardly achieves a reasonable predictive accuracy. These findings emphasize the need for further methodological refinements and for an increase in dataset diversity to improve model generalizability. To address these challenges, several opportunities for improvement emerge. Expanding the dataset to include a broader range of cultivars, harvest years, and sensory diversity will be crucial for improving model accuracy and ensuring its generalizability across different olive oil types. Beyond its limitations, this study demonstrates the potential of NMR spectroscopy as a transformative tool for quality control and product traceability in the industry. The ability to rapidly assess quality, detect authenticity markers, and classify olive oils by cultivar and harvest year in a single measurement, positions NMR as a powerful alternative to conventional methods. As global interest in traceability and quality assurance continues to grow, NMR-based approaches, coupled with advanced machine learning algorithms, could redefine the future of olive oil authentication, ensuring greater transparency and consumer confidence in the marketplace. Future studies should continue to explore the implications of these findings and refine NMR and statistical methodologies for enhanced reliability and applicability in olive oil quality assessment.

Funding

This work was financed by the project grant TIMONE-ID11 (https://www.cerm.unifi.it/timone/timone/), funded by the Italian Ministry of Agricultural, Food and Forestry Policies (MiPAAF) to Claudio Luchinat and Pietro Tonutti.

CRediT authorship contribution statement

Luchinat Claudio: Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Tonutti Pietro:** Writing – review & editing, Resources. **Cherubini Chiara:** Writing – review & editing, Investigation. **Mazzanti Laura:** Writing – review & editing, Investigation. **MEONI GAIA:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tenori Leonardo:** Writing – review & editing, Validation, Supervision, Software, Methodology, Investigation, Conceptualization. **Di Cesare Francesca:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Brizzolara Stefano:** Writing – review & editing, Resources.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that no AI technologies were used to generate the content of this manuscript. All content has been edited solely by the authors.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

Gaia Meoni was supported by a research contract co-funded by the European Union - PON Research and Innovation 2014–2020, in accordance with Article 24, paragraph 3a) of Law No. 240 of December 30, 2010, as amended, and Ministerial Decree No. 1062 of August 10, 2021. This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy center, as well as the project “Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC, Acronym “ITACA.SB” (Project no. IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union – NextGenerationEU.

The authors would like to express their sincere gratitude to Prof. Francesca Venturi and the Food Technology group at the Department of Agriculture, Food and Environment (DAFE), University of Pisa, for their support and expertise in conducting the panel test assessments for sensory evaluations.

Additionally, the authors extend their heartfelt appreciation to the donors and producers who share a deep interest in exploring the fascinating world of olive oil and are committed to certifying the quality of their products. Their generosity in providing us with samples of their oils was invaluable, and without their contribution, this work would not have been possible. A special thanks goes to the olive oil producers of the Tuscany region, including those from the "Selezione degli Oli Extra Vergine di Oliva DOP e IGP della Toscana", the producers from the hills of Fiesole, Fattoria di Volmiano Citernesi, Puneti, Torre Bianca, Famiglia Pasquini, and the New York University at Villa la Pietra. Their dedication to excellence in olive oil production and their willingness to support scientific research have been of the utmost importance for the success of this study.

CL would also like to thank his personal friends, who enthusiastically and proudly shared samples of Extra Virgin Olive Oil from their own production: Carlo Eugenio Casini, Giovanni Donnini, Paolo Fresco and Mauro Pagliai. CL would particularly like to mention the donation of several samples of his Villa Campestri oil by Paolo Pasquali, who sadly is no longer with us. Finally, special thanks go to Donatella Lippi, colleague and friend of CL, who organized the collection of several oil samples from the Associazione Le Colline di Careggi.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.03.045.

References

[1] International Olive Council [Internet]. [cited 2025]. Unità Economia e Promozione. Available from: ⟨https://www.internationaloliveoil.org/il-nostro-lavoro/unita-economia-e-promozione/?lang=it⟩.
[2] Grigg D, Olive oil. the Mediterranean and the world. GeoJournal 2001;53(2):163–72.
[3] Uylaşer V, Yildiz G. The historical development and nutritional importance of olive and olive oil constituted an important part of the mediterranean diet. Crit Rev Food Sci Nutr 2014 Jan 1;54(8):1092–101.
[4] Chiavarini M, Rosignoli P, Giacchetta I, Fabiani R. Health outcomes associated with olive oil intake: an umbrella review of meta-analyses. Foods 2024 Aug 21;13(16):2619.
[5] Vazquez-Aguilar A, Sanchez-Rodriguez E, Rodriguez-Perez C, Rangel-Huerta OD, Mesa MD. Metabolomic-based studies of the intake of virgin olive oil: a comprehensive review. Metabolites 2023 Apr;13(4):472.
[6] EU agricultural outlook - for markets, income and environment 2021 - 2031 | FAO [Internet]. [cited 2025 Jan 27]. Available from: ⟨https://www.fao.org/family-farming/detail/en/c/1459441/⟩.
[7] Casadei E., Valli E., Bendini A., Barbieri S., Tucci R., Ferioli F., et al. Valorization of monovarietal Nostrana di Brisighella extra virgin olive oils: focus on bioactive compounds. Front Nutr [Internet]. 2024 Apr 4 [cited 2025 Jan 27];11. Available from: ⟨https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2024.1353832/full⟩.

[8] Di Lecce G, Piochi M, Pacetti D, Frega NG, Bartolucci E, Scortichini S, et al. Eleven monovarietal extra virgin olive oils from olives grown and processed under the same conditions: effect of the cultivar on the chemical composition and sensory traits. Foods 2020 Jul;9(7):904.

[9] Tena N, Wang SC, Aparicio-Ruiz R, García-González DL, Aparicio R. In-depth assessment of analytical methods for olive oil purity, safety, and quality characterization. J Agric Food Chem 2015 May 13;63(18):4509–26.

[10] Hashempour-baltork F, Vali Zade S, Mazaheri Y, Mirza Alizadeh A, Rastegar H, Abdian Z, et al. Recent methods in detection of olive oil adulteration: state-of- the-art. J Agric Food Res 2024 Jun 1;16:101123.

[11] Silva BS, Schmiele M. From olive to olive oil: a general approach. Res, Soc Dev 2021 Mar 17;10(3). e32210313408–e32210313408.

[12] González-Pereira A, Otero P, Fraga-Corral M, Garcia-Oliveira P, Carpena M, Prieto MA, et al. State-of-the-art of analytical techniques to determine food fraud in olive oils. Foods 2021 Feb 24;10(3):484.

[13] Lioupi A, Nenadis N, Theodoridis G. Virgin olive oil metabolomics: a review. J Chromatogr B 2020 Aug 1;1150:122161.

[14] Zaroual H, Chénè C, El Hadrami EM, Karoui R. Application of new emerging techniques in combination with classical methods for the determination of the quality and authenticity of olive oil: a review. Crit Rev Food Sci Nutr 2022 Jun 1; 62(16):4526–49.

[15] Willenberg I, Matthäus B, Gertz C. A new statistical approach to describe the quality of extra virgin olive oils using near infrared spectroscopy (NIR) and traditional analytical parameters. Eur J Lipid Sci Technol 2019;121(2):1800361.

[16] Borràs E, Mestres M, Aceña L, Busto O, Ferré J, Boqué R, et al. Identification of olive oil sensory defects by multivariate analysis of mid infrared spectra. Food Chem 2015 Nov 15;187:197–203.

[17] Zaroual H, Hadrami EME, Karoui R. A preliminary study on the potential of front face fluorescence spectroscopy for the discrimination of Moroccan virgin olive oils and the prediction of their quality. Anal Methods 2021 Jan 28;13(3):345–58.

[18] Guzmán E, Baeten V, Pierna JAF, García-Mesa JA. Evaluation of the overall quality of olive oil using fluorescence spectroscopy. Food Chem 2015 Apr 15;173:927–34.

[19] Arroyo-Cerezo A, Yang X, Jiménez-Carvelo AM, Pellegrino M, Felicita Savino A, Berzaghi P. Assessment of extra virgin olive oil quality by miniaturized near infrared instruments in a rapid and non-destructive procedure. Food Chem 2024 Jan 1;430:137043.

[20] Ghini V, Meoni G, Vignoli A, Di Cesare F, Tenori L, Turano P, et al. Fingerprinting and profiling in metabolomics of biosamples. Prog Nucl Magn Reson Spectrosc 2023 Nov 1;138–139:105–35.

[21] Mannina L, Sobolev AP. High resolution NMR characterization of olive oils in terms of quality, authenticity and geographical origin. Magn Reson Chem 2011;49(S1): S3–11.

[22] Alonso-Salces RM, Moreno-Rojas JM, Holland MV, Reniero F, Guillou C, Héberger K. Virgin olive oil authentication by multivariate analyses of ^{1}H NMR fingerprints and δ13C and δ2H data. J Agric Food Chem 2010 May 12;58(9): 5586–96.

[23] Longobardi F, Ventrella A, Napoli C, Humpfer E, Schütz B, Schäfer H, et al. Classification of olive oils according to geographical origin by using ^{1}H NMR fingerprinting combined with multivariate analysis. Food Chem 2012 Jan 1;130 (1):177–83.

[24] Maestrello V, Solovyev P, Bontempo L, Mannina L, Camin F. Nuclear magnetic resonance spectroscopy in extra virgin olive oil authentication. Compr Rev Food Sci Food Saf 2022;21(5):4056–75.

[25] Mannina L, Marini F, Gobbino M, Sobolev AP, Capitani D. NMR and chemometrics in tracing European olive oils: the case study of Ligurian samples. Talanta 2010 Mar 15;80(5):2141–8.

[26] Li X, Muñoz-Díez C, Miho H, Zhang L, Li P, Priego F, et al. Evaluation of phenolics in the analysis of virgin olive oil using near infrared spectroscopy. Spectrochim Acta Part A: Mol Biomol Spectrosc 2025 Feb 5;326:125262.

[27] Reda R, Saffaj T, Bouzida I, Saidi O, Belgrir M, Lakssir B, et al. Optimized variable selection and machine learning models for olive oil quality assessment using portable near infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2023 Dec 15;303:123213.

[28] Fang P, Wang H, Wan X. Olive oil authentication based on quantitative β-carotene Raman spectra detection. Food Chem 2022 Dec 15;397:133763.

[29] Vali Zade S, Forooghi E, Jannat B, Hashempour-baltork F, Abdollahi H. A combined classification modeling strategy for detection and identification of extra virgin olive oil adulteration using Raman spectroscopy. Chemom Intell Lab Syst 2023 Sep 15;240:104903.

[30] Laouni A, El Orche A, Elhamdaoui O, Karrouchi K, El Karbane M, Bouatia M. A preliminary study on the potential of FT-IR spectroscopy and chemometrics for tracing the geographical origin of moroccan virgin olive oils. J AOAC Int 2023 Jun 1;106(3):804–12.

[31] Rifna EJ, Pandiselvam R, Kothakota A, Subba Rao KV, Dwivedi M, Kumar M, et al. Advanced process analytical tools for identification of adulterants in edible oils – a review. Food Chem 2022 Feb 1;369:130898.

[32] Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, et al. High-throughput metabolomics by ^{1}D NMR. Angew Chem Int Ed Engl 2019 Jan 21;58(4):968–94.

[33] Wei F, Furihata K, Miyakawa T, Tanokura M. A pilot study of NMR-based sensory prediction of roasted coffee bean extracts. Food Chem 2014;152:363–9.

[34] Malmendal A, Amoresano C, Trotta R, Lauri I, De Tito S, Novellino E, et al. NMR Spectrometers as "Magnetic Tongues": prediction of sensory descriptors in canned tomatoes. J Agric Food Chem 2011 Oct 26;59(20):10831–8.

[35] Gbashi S, Njobeh PB. Enhancing food integrity through artificial intelligence and machine learning: a comprehensive review. Appl Sci 2024 Jan;14(8):3421.

[36] TIMONE [Internet]. [cited 2025 Feb 27]. Trattamenti Innovativi in Pre-frangitura delle Olive per il Miglioramento delle Caratteristiche Organolettiche e delle Proprietà Nutraceutiche dell'Olio. Available from: ⟨https://www.cerm.unifi.it/timone/timone/⟩.

[37] Cecchi L, Migliorini M, Cherubini C, Giusti M, Zanoni B, Innocenti M, et al. Phenolic profiles, oil amount and sugar content during olive ripening of three typical Tuscan cultivars to detect the best harvesting time for oil production. Food Res Int 2013 Dec 1;54(2):1876–84.

[38] Macaluso M, Taglieri I, Venturi F, Sanmartin C, Bianchi A, De Leo M, et al. influence of the atmosphere composition during malaxation and storage on the shelf life of an unfiltered extra virgin olive oil: preliminary results. Eur J Lipid Sci Technol 2021;123(2):2000122.

[39] Commission Delegated Regulation (EU) 2015/1830 of 8 July 2015 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis [Internet]. OJ L Jul 8, 2015. Available from: ⟨http://data.europa.eu/eli/reg_del/2015/1830/oj/eng⟩.

[40] Alharbi H, Kahfi J, Dutta A, Jaremko M, Emwas AH. The detection of adulteration of olive oil with various vegetable oils – a case study using high-resolution 700 MHz NMR spectroscopy coupled with multivariate data analysis. Food Control 2024 Dec 1;166:110679.

[41] Cinzia I, Antonella C, Mannina L, Simone C, Silvia V, Donatella C, et al. Extra-Virgin Olive Oils Nine Ital Reg: H-1 NMR-Chemom Charact 2019 [cited 2025 Jan 29]; Available from: ⟨https://iris.cnr.it/handle/20.500.14243/379259⟩.

[42] Dourou AM, Brizzolara S, Meoni G, Tenori L, Famiani F, Luchinat C, et al. The inner temperature of the olives (cv. Leccino) before processing affects the volatile profile and the composition of the oil. Food Res Int 2020 Mar 1;129:108861.

[43] Naveja JJ, Rico-Hidalgo MP, Medina-Franco JL. Analysis of a large food chemical database: chemical space, diversity, and complexity. F1000Res 2018 Aug 10;7. Chem Inf Sci-993.

[44] Castejón D, Mateos-Aparicio I, Molero MD, Cambero MI, Herrera A. Evaluation and optimization of the analysis of fatty acid types in edible Oils by ^{1}H NMR. Food Anal Methods 2014 Jul 1;7(6):1285–97.

[45] Alonso-Salces RM, Gallo B, Collado MI, Sasía-Arriba A, Viacava GE, García-González DL, et al. 1H–NMR fingerprinting and supervised pattern recognition to evaluate the stability of virgin olive oil during storage. Food Control 2021 May 1; 123:107831.

[46] Ruiz-Aracama A, Goicoechea E, Guillén MD. Direct study of minor extra-virgin olive oil components without any sample modification. ^{1}H NMR multisupression experiment: a powerful tool. Food Chem 2017 Aug 1;228:301–14.

[47] Ün İ, OK S. Analysis of olive oil for authentication and shelf life determination. J Food Sci Technol 2018 Jul 1;55(7):2476–87.

[48] Goicoechea-Oses E, Ruiz-Aracama A. Usefulness of the ^{1}H NMR multisuppression approach for the global characterization of monovarietal extra-virgin olive oils. Foods 2024 Jan;13(14):2298.

[49] Dourou AM, Brizzolara S, Meoni G, Tenori L, Famiani F, Luchinat C, et al. The inner temperature of the olives (cv. Leccino) before processing affects the volatile profile and the composition of the oil. Food Res Int 2020 Mar 1;129:108861.

[50] Salvo A, Rotondo A, La Torre GL, Cicero N, Dugo G. Determination of 1,2/1,3-diglycerides in Sicilian extra-virgin olive oils by ^{1}H NMR over a one-year storage period. Nat Prod Res 2017 Apr;31(7):822–8.

[51] Di Pietro ME, Mannu A, Mele A. NMR determination of free fatty acids in vegetable oils. Processes 2020 Apr;8(4):410.

[52] Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. Int J Comput Sci Issues(IJCSI) 2012 Sep 1:9.

[53] Breiman L. Random forests. Mach Learn 2001 Oct 1;45(1):5–32.

[54] Amaratunga D, Cabrera J, Lee YS. Enriched random forests. Bioinformatics 2008 Sep 15;24(18):2010–4.

[55] Nguyen JM, Jézéquel P, Gillois P, Silva L, Azzouz FB, Lambert-Lacroix S, et al. Random forest of perfect trees: concept, performance, applications and perspectives. Bioinformatics 2021 Aug;37(15):2165.

[56] Scott IM, Lin W, Liakata M, Wood JE, Vermeer CP, Allaway D, et al. Merits of random forests emerge in evaluation of chemometric classifiers by external validation. Anal Chim Acta 2013 Nov 1;801:22–33.

[57] Bobbo T, Meoni G, Niero G, Tenori L, Luchinat C, Cassandro M, et al. Nuclear magnetic resonance spectroscopy to investigate the association between milk metabolites and udder quarter health status in dairy cows. J Dairy Sci 2021 Oct 13; (21). S0022-0302.

[58] Hauke J, Kossowski T. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. Quaest Geogr 2011 Jun 1;30(2): 87–93.

[59] rlm function - RDocumentation [Internet]. [cited 2025 Feb 24]. Available from: ⟨https://www.rdocumentation.org/packages/MASS/versions/7.3-64/topics/rlm⟩.

[60] Iglewicz B, Heiberger RM, Moore D, Tan Y. Algorithms, routines, and s functions, for robust statistics. Technometrics [Internet] 1995 Aug 1 [cited 2025 Feb 24]; Available from: ⟨https://www.tandfonline.com/doi/abs/10.1080/00401706.1995.10484353⟩.

[61] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30(1):79–82.

[62] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger JM, McBratney A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends Anal Chem 2010 Oct 1;29(9):1073–81.

[63] RStudio | Open source & professional software for data science teams [Internet]. [cited 2022 Jun 14]. Available from: ⟨https://www.rstudio.com/⟩.

[64] Dais P, Hatzakis E. Quality assessment and authentication of virgin olive oil by NMR spectroscopy: a critical review. Anal Chim Acta 2013 Feb 26;765:1–27.
[65] Ray N.B., Hilsabeck K.D., Pitsillou E., Mann A., Karagiannis T.C., McCord D.E. Chapter 13 - Olive fruit and olive oil bioactive polyphenols in the promotion of health. In: Singh R.B., Watanabe S., Isaza A.A., editors. Functional Foods and Nutraceuticals in Metabolic and Non-Communicable Diseases [Internet]. Academic Press; 2022 [cited 2022 Sep 9]. p. 203–220. Available from: https://www.sciencedirect.com/science/article/pii/B9780128198155000148.
[66] Antoniadi L, Angelis A, Stathopoulos P, Bata EM, Papoutsaki Z, Halabalaki M, et al. Oxidized forms of olive oil secoiridoids: semisynthesis, identification and correlation with quality parameters. Planta Med 2022 Aug;88(9/10):805–13.
[67] Mousavi S, Mariotti R, Stanzione V, Pandolfi S, Mastio V, Baldoni L, et al. Evolution of extra virgin olive oil quality under different storage conditions. Foods 2021 Aug 21;10(8):1945.
[68] Tuck KL, Hayball PJ. Major phenolic compounds in olive oil: metabolism and health effects. J Nutr Biochem 2002 Nov 1;13(11):636–44.
[69] Yalcin S, Schreiner M. Stabilities of tocopherols and phenolic compounds in virgin olive oil during thermal oxidation. J Food Sci Technol 2018 Jan;55(1):244–51.
[70] Nikou T, Witt M, Stathopoulos P, Barsch A, Halabalaki M. Olive oil quality and authenticity assessment aspects employing FIA-MRMS and LC-orbitrap MS metabolomic approaches. Front Public Health [Internet] 2020 Sep 25 [cited 2024 Nov 8];8. Available from: ⟨https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2020.558226/full⟩.
[71] Ciampa A, Danesi F, Picone G. NMR-based metabolomics for a more holistic and sustainable research in food quality assessment: a narrative review. Appl Sci 2023 Jan;13(1):372.
[72] Guillaume C, Ravetti L. Shelf-life prediction of extra virgin olive oils using an empirical model based on standard quality tests. J Chem 2016;2016(1):6393962.
[73] Barbieri S, Brkić Bubola K, Bendini A, Bučar-Miklavčič M, Lacoste F, Tibet U, et al. Alignment and proficiency of virgin olive oil sensory panels: the OLEUM approach. Foods 2020 Mar;9(3):355.
[74] Circi S, Capitani D, Randazzo A, Ingallina C, Mannina L, Sobolev AP. Panel test and chemical analyses of commercial olive oils: a comparative study. Chem Biol Technol Agric 2017 Jul 26;4(1):18.
[75] Tsolis T, Kyriakou D, Sifnaiou E, Thomos D, Glykos D, Tsiafoulis CG, et al. NMR analysis of extra virgin olive oil of the epirus region of greece with emphasis on selected phenolic compounds. Molecules 2024 Jan;29(5):1111.
[76] Louppis AP, Kontominas MG. Analytical insights for ensuring authenticity of Greek agriculture products: unveiling chemical marker applications. Food Chem 2024 Jul 1;445:138758.
[77] Cao R, Liu X, Liu Y, Zhai X, Cao T, Wang A, et al. Applications of nuclear magnetic resonance spectroscopy to the evaluation of complex food constituents. Food Chem 2021 Apr 16;342:128258.
[78] Meoni G, Tenori L, Luchinat C. Nuclear magnetic resonance-based metabolomic comparison of breast milk and organic and traditional formula milk brands for infants and toddlers. OMICS A J Integr Biol 2020;24(7):424–36.
[79] Meoni G, Luchinat C, Gotti E, Cadena A, Tenori L. Phenotyping green and roasted beans of nicaraguan coffea arabica varieties processed with different post-harvest practices. Appl Sci 2021 Jan;11(24):11779.
[80] Uncu O, Ozen B, Tokatli F. Authentication of Turkish olive oils by using detailed pigment profile and spectroscopic techniques. J Sci Food Agric 2020;100(5):2153–65.
[81] Beteinakis S, Papachristodoulou A, Kolb P, Rösch P, Schwarzinger S, Mikros E, et al. NMR-based metabolite profiling and the application of STOCSY toward the quality and authentication assessment of European EVOOs. Molecules 2023 Jan;28(4):1738.
[82] Tang F, Polari JJ, Green HS, Wang SC, Hatzakis E. NMR-based metabolomics for olive oil cultivar classification: a comparison with standard targeted analysis of fatty acids and triglycerides. Food Control 2022 Jul 1;137:108939.
[83] Martín G, Francisco J. Potential of near-infrared spectroscopy for the determination of olive oil quality. Sensors 2022 Jan;22(8):2831.
[84] Venturini F, Sperti M, Michelucci U, Gucciardi A, Martos VM, Deriu MA. Extraction of physicochemical properties from the fluorescence spectrum with 1D convolutional neural networks: Application to olive oil. J Food Eng 2023 Jan 1;336:111198.
[85] Zaroual H, El Hadrami EM, Chénè C, Karoui R. Fourier transform infrared spectroscopy coupled with chemometrics for the monitoring of virgin olive oil quality during storage up to 18 months. Eur Food Res Technol 2024 Jul 1;250(7):1969–86.
[86] Rodríguez-López P, Lozano-Sanchez J, Borrás-Linares I, Emanuelli T, Menéndez JA, Segura-Carretero A. Structure–biological activity relationships of extra-virgin olive oil phenolic compounds: health properties and bioavailability. Antioxidants 2020 Aug;9(8):685.
[87] Sanmartin C, Taglieri I, Macaluso M, Sgherri C, Ascrizzi R, Flamini G, et al. Cold-pressing olive oil in the presence of cryomacerated leaves of olea or citrus: nutraceutical and sensorial features. Molecules 2019 Jul 19;24(14):2625.
[88] Allouche Y, López EF, Maza GB, Márquez AJ. Near infrared spectroscopy and artificial neural network to characterise olive fruit and oil online for process optimisation. J Infrared Spectrosc 2015 Apr 1;23(2):111–21.
[89] Alves E, Rey F, Melo T, Barros MP, Domingues P, Domingues R. Bioprospecting bioactive polar lipids from olive (Olea europaea cv. Galega vulgar) fruit seeds: LC-HR-MS/MS fingerprinting and sub-geographic comparison. Foods 2022 Mar 25;11(7):951.
[90] Che Idris CA, Wai Lin S, Abdull Razis AF. Hypocholesterolaemic and anti-atherogenic effects of palm-based oils (NoveLin I and NoveLin II) in cholesterol-fed rabbits. Int J Environ Res Public Health 2020 May;17(9):3226.
[91] García-Vico L, Belaj A, Sánchez-Ortiz A, Martínez-Rivas JM, Pérez AG, Sanz C. Volatile compound profiling by HS-SPME/GC-MS-FID of a core olive cultivar collection as a tool for aroma improvement of virgin olive oil. Molecules 2017 Jan;22(1):141.
[92] Girelli CR, Calò F, Angilè F, Mazzi L, Barbini D, Fanizzi FP. ^{1}H NMR spectroscopy to characterize italian extra virgin olive oil blends, using statistical models and databases based on monocultivar reference oils. Foods 2020 Dec;9(12):1797.
[93] Talhaoui N, Gómez-Caravaca AM, León L, De la Rosa R, Fernández-Gutiérrez A, Segura-Carretero A. From olive fruits to olive oil: phenolic compound transfer in six different olive cultivars grown under the same agronomical conditions. Int J Mol Sci 2016 Mar 4;17(3):337.