Journal of Food Composition and Analysis 152 (2026) 109038

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

What are the key differences among Italian monovarietal extra virgin olive oils? Holistic clustering based on volatile, phenolic and terpene characterization and sensory profile: A four-year study

Tommaso Ugolini [a], Bruno Zanoni [a], Alessandro Parenti [a], Fabrizio Melani [b], Nadia Mulinacci [b], Carlotta Breschi [a], Marzia Migliorini [c], Lorenzo Guerrini [d], Lorenzo Cecchi [a,*]

[a] DAGRI – Department of Agricultural, Food, Environmental, and Forestry Sciences and Technologies – University of Florence, via Donizetti, 6, Firenze 50144, Italy
[b] Department of NEUROFARBA, University of Florence, Via Ugo Schiff 6, Sesto F.no, Florence 50019, Italy
[c] Carapelli Firenze S.p.A., Via Leonardo da Vinci 31, Tavarnelle Val di Pesa, Firenze 50028, Italy
[d] Dipartimento Territorio e Sistemi Agro-Forestali (TESAF), Università degli Studi di Padova, (Viale) dell'Università 16, Legnaro, PD 35020, Italy

ARTICLE INFO

Keywords:
Cultivar
Sesquiterpenes
Monoterpenes
Authentication
HS-SPME-GC-MS
Chemometric
Machine learning

ABSTRACT

This study suggests original holistic approach based on comprehensive analysis of phenolic compounds (HPLC-DAD), volatile (including pentene dimers) and terpene compounds (HS-SPME-GC-MS) and sensory analysis by the Panel Test to propose practical tools for clustering and quality differentiation of monovarietal extra virgin olive oils (MEVOO). The sample-set consisted of 356 MEVOO samples from 13 major Italian cultivars analyzed across five production years. Hierarchical Cluster Analysis following genetic algorithm-linear discriminant analysis and random forest methods allowed identifying three robust cultivar clusters with similar sensory features and selecting nine significant sensory attributes. Models for samples classification in these clusters were set and externally validated. Terpenes exhibited the best relationships with sensory clustering (reaching approx. 92% accuracy), but combining terpene, volatile and phenolic compounds improved accuracy up to 98%. Finally, two simplified chemometric models based either on 8 terpenes or 12 combined terpene, volatile, and phenolic compounds were proposed for predicting MEVOOs classification in the three clusters.

1. Introduction

Extra virgin olive oil (EVOO) is the highest-quality category of olive oil and is widely recognized as the vegetable oil with the most favorable sensory and nutritional properties (EFSA, 2009; EFSA, 2011; Peri, Peri, 2014). EVOO is appreciated globally, with its production historically concentrated in the Mediterranean basin, primarily in Spain and Italy, but it is recently expanding beyond the natural boundaries of this region (COI 2024b). However, the EVOO sector faces several challenges, starting with the fight against fraud and the issue of profitability for producers.

According to the EU Agri-Food Fraud Network Report (EU, 2020) on suspicious non-compliances and potential intentional violations of the EU agri-food chain legislation, "Fats and oils" was the most frequently reported product category, with majority of cases involving olive oil. Recorded olive oil non-compliances typically relate to a failure to meet legal limits of the intrinsic physical-chemical and sensory characteristics required by EU for EVOO (Barbieri et al., 2020; T. Cecchi and Alfei, 2013; COI, 2018; COI 2024a; EU 2022a; EU 2022b; Conte et al., 2020). These violations often result from fraudulent practices such as mixing with cheaper lower-quality oils, adulteration, mislabeling (Cecchi et al., 2017; Dias et al., 2014), or failure to apply good manufacturing

Abbreviations: **CEA**, Class Error Average; **COI**, International Olive Council; **"defect"**, monovarietal extra virgin olive oil samples with median of defect > 1; **EVOO**, Extra virgin olive oil commercial category; **"EVOO#"**, monovarietal extra virgin olive oil samples with median of defect ≤ 1; **GA**, genetic algorithm; **HCA**, Hierarchical Clustering Analysis; **HS-SPME-GC-MS**, HeadSpace-Solid Phase MicroExtraction-Gas Chromatography-Mass Spectrometry; **LDA**, Linear Discriminant Analysis; **LOX**, lipoxygenase; **MEVOO**, monovarietal extra virgin olive oil; **MISN**, Multiple Internal Standard Normalization; **non-EVOO**, samples not belonging to the Extra virgin olive oil commercial category **PLS-DA**, Partial Least Square-Discriminant Analysis; **RF**, Random Forest; **VHCs**, volatile hydrocarbon compounds; **VOCs**, volatile organic compounds; **VOOx**, virgin olive oil.

* Corresponding author.
E-mail address: lo.cecchi@unifi.it (L. Cecchi).

https://doi.org/10.1016/j.jfca.2026.109038
Received 9 October 2025; Received in revised form 20 February 2026; Accepted 25 February 2026
Available online 26 February 2026

practices.

Like other sectors, olive oil producers may adopt either price-based or quality-based competition strategies to improve profitability (Morikawa, 2021). Price-based competition is more common in production systems where a small number of large mills employ highly mechanized oil extraction processes using a limited number of olive varieties, thereby leading to reduced production costs. Conversely, quality-based competition is prevalent in production systems where numerous small mills process a wide array of cultivars, leading to differentiated quality types of EVOO closely linked to the local territory; Italy, with approx. 400 registered cultivars and a lot of small mills, is an example of the latter strategy (ISMEA, 2018).

Cultivar diversification appears to be a key factor in enhancing the value of EVOO (Campestre et al., 2017; COI, 2018; Ugolini et al., 2024; Žanetić et al., 2021). Interest in monovarietal EVOO (MEVOO) is growing (Guerrini et al., 2023; Montealegre et al., 2010; Olive Oil Times, 2023; Ugolini et al., 2024), driven by consumers' increasing willingness to pay premium prices for products with distinctive sensory qualities (Bajoub et al., 2018; Delgado & Guinard, 2011; Lukić et al., 2019; Stasi et al., 2018). Consequently, ensuring the authenticity and promoting the value of MEVOOs (starting with securing their rightful place in competitive EVOO awards) has become essential.

Research efforts should focus on addressing the current lack of information regarding the sensory characteristics of MEVOOs (Cecchi, Parenti, et al., 2022; Dekhili et al., 2011; Del Giudice et al., 2015; ISMEA, 2018; Piccolo et al., 2013) thus providing consumers with effective tools to understand and appreciate the sensory profiles of MEVOOs from different cultivars. In the EU, the Panel Test is mandatory to assess the sensory quality of virgin olive oil. Often, it is primarily focused in identifying any sensory defect and a limited set of positive attributes such as bitterness, pungency, and fruitiness (ripe or green), but it can also assess more in detail a broader range of positive attributes typical of samples like MEVOOs, as it happens in EVOO competitions (EU 2022a; Conte et al., 2020; Cecchi, Parenti, et al., 2022; Rébufa et al., 2021; Clodoveo et al., 2022; García-González et al., 2023).

Given that numerous compounds, especially volatile organic compounds (VOCs), phenolic compounds and terpenes, contribute to the sensory profile of EVOO (Aprea et al., 2018; Blasi et al., 2019; Campestre et al., 2017; Cecchi, Migliorini, et al., 2022; T. Cecchi & Alfei, 2013; Guth and Grosch, 1993; Kalua et al., 2005; Kesen et al., 2013; Ocakoglu et al., 2009; Torres-Cobos et al., 2021; Ugolini et al., 2024; Vichi et al., 2006; Žanetić et al., 2021), further research effort is needed to explore the relationship between chemical composition and sensory characteristics in order to develop models for authenticating and/or clustering MEVOOs.

Recent literature has provided promising insights into how specific classes of molecules contribute to quality differences among MEVOOs from different cultivars (Cecchi et al., 2022; Ugolini et al., 2024). Although a clear classification of individual cultivars has not yet been achieved, MEVOOs have been successfully classified into clusters of cultivars with similar chemical characteristics, regardless of pedoclimatic/agronomic and processing variables. In a study, VOCs have proven effective in distinguishing different MEVOO clusters (Cecchi, Migliorini, et al., 2022). In this study, 320 MEVOO samples from 9 cultivars across 4 production years were collected from the main producing countries in the Mediterranean region. Seventy-two VOCs were quantitated via HeadSpace-Solid Phase MicroExtraction-Gas Chromatography-Mass Spectrometry (HS-SPME-GC-MS), and chemometric methods (i.e., Hierarchical Clustering Analysis (*HCA*) and Partial Least Square-Discriminant Analysis (*PLS-DA*)) were applied, allowing 94% of samples to be correctly classified into three clusters of cultivars: (i) Arbequina and Nocellara; (ii) Coratina, Frantoio, Leccino and Moraiolo; (iii) Hojiblanca, Picual and Koroneiki. Ten VOCs, eight of which are intermediate in the lipoxygenase (LOX) pathway, were identified as key discriminators among these clusters. In a second study, volatile hydrocarbon compounds (VHCs) such as terpenes and pentene dimers (a group of molecules less studied in the literature) also showed strong discriminatory power (Ugolini et al., 2024). In this case, 225 MEVOOs from 14 Italian cultivars across 3 production years were analyzed, and 60 VHCs were quantitated via HS-SPME-GC-MS; chemometric methods such as *HCA* and Linear Discriminant Analysis (*LDA*) allowed achieving a 94% of correct classification across four clusters of cultivars: (i) Ascolana Tenera, Casaliva, Leccio del Corno, Moraiolo, Pendolino, Frantoio and Leccino; (ii) Coratina, Picholine, Tonda Iblea and Nocellara; (iii) Bosana and Peranzana; (iv) Itrana. Among VHCs, sesquiterpenes were the most significant compounds for clusters differentiation. Interestingly, some cultivars (i.e., Coratina and Nocellara) were grouped differently depending on whether classification was based on VOCs or VHCs.

The findings of these previous studies indicate that a more effective approach to clustering Italian MEVOOs should integrate all classes of chemical compounds associated with the sensory traits typical of monovarietal oils. Furthermore, compared to the previous studies, it is essential to integrate the chemical data related to these molecular classes with data concerning the sensory profiles of MEVOOs. For this reason, a new methodological framework based on the integration of chemical and sensory profile data for clustering groups of olive cultivars with oils showing similar sensory characteristics and for developing simplified chemical predictors is required to cover current literature research gap. The proposed approach introduces the following novelties: (i) the integration for the first time of sensory data together with multiple chemical classes, including both compounds widely used for EVOO characterization, such as VOCs and phenolic compounds, as well as less frequently considered classes such as terpenes and pentene dimers; (ii) the definition of cultivar clusters based on MEVOO sensory data; (iii) the development of chemical-based models for the classification of MEVOOs into the clusters defined in point (ii).

Therefore, the aim of this research was to classify major Italian cultivars into clusters with similar sensory profiles and to identify which volatile, terpene and phenolic compounds best explain these sensory clusters. This study therefore proposes an original holistic approach consisting, for the first time, of the integration of combined chemical (terpenes, LOX VOCs, and phenolic compounds) and sensory data to better understand quality differences in MEVOOs. Samples from five production years were collected and analyzed. Given the substantial volume of experimental data, several chemometric methods (e.g., genetic algorithm-linear discriminant analysis, random forest) were employed.

2. Material and methods

2.1. Chemicals and standards preparation

Ultrapure water was produced using a Milli-Q-System (Millipore SA, Milsheim, France). Standards of tyrosol (purity > 99.5%) and syringic acid (purity > 95%), along with the reagents formic acid and phosphoric acid, and the HPLC-grade solvents methanol and acetonitrile, were purchased from Merck (Darmstadt, Germany). All commercial standards (including both external and internal standards) for VOCs and terpenes analysis via HS-SPME-GC-MS were purchased as previously reported when validating the relating methods (Cecchi et al., 2024; Cecchi et al., 2019)

Solutions of internal and external standards for VOCs and VHCs analyses were prepared according to the Multiple Internal Standard Normalization (MISN) approach, by dissolving appropriate amounts of the commercial standards in refined olive oil previously verified as free from interfering compounds. Following validated methods (Cecchi et al., 2019; Cecchi et al., 2024), solutions for building calibration curves of VOCs and VHCs were prepared by diluting the internal and external standard solutions. Each diluted solution contained a constant amount of internal standards and increasing amounts of external standards. Solutions were stored at −20 °C until HS-SPME-GC-MS analysis.

2.2. Samples

The MEVOO samples were collected from olive oil producers, public institutions, and associations. All samples were obtained from olives of Italian cultivars processed in Italy over five consecutive production years; for each production year, each sample was produced by a different producer.

A first set of 398 samples were collected during the 2020–2021, 2021–2022, 2022–2023, and 2023–2024 olive oil production years. Samples of cultivars represented by fewer than eight samples were excluded from the sample-set, to minimize underestimation of within-cultivar variability (they are listed in the annex I of the Supplementary file). This resulted in a sample-set of 282 monovarietal samples: for the purposes of this study, 240 of them were labelled as "**EVOO#**" samples, and 42 labelled as "**defect**" samples (Table 1A). Samples labelled as "**EVOO#**" included all samples with defect ≤ 1 (see Section 2.7), while samples labelled as "**defect**" included samples with defect > 1. This sample-set was used for defining the sensory clusters and as the training-set for building and internally validating the chemometric model for MEVOOs classification in the defined clusters. A further set of 74 samples was collected during the 2024–2025 production year and was used as the test-set for the external validation of the chemometric model (Table 1B). Samples were analyzed within one month after collection; during this period, they were stored in the dark at room temperature in sealed 500 mL bottles to minimize headspace and oxygen exposure.

Table 1A-B shows the sample-sets categorized by botanical origin, production years, and category (i.e., "**EVOO#**" *vs.* "**defect**" samples). The entire data-set is representative of the Italian MEVOO production scenario, as it includes 13 major Italian cultivars from different regions, as follows: Ascolana Tenera (**AT**, 19 samples), Bosana (**BO**, 19), Coratina (**CO**, 67), Frantoio (**FR**, 36), Itrana (**IT**, 25), Leccio del Corno (**LC**, 17), Leccino (**LE**, 31), Maurino (**MA**, 15), Moraiolo (**MO**, 29), Nocellara (**NO**, 32), Pendolino (**PE**, 10), Peranzana (**PR**, 27), and Tonda Iblea (**TI**, 29).

2.3. 2.3 Analyses for chemical compliance with EU legal limits

The EU legal chemical characteristics of the samples were measured according to European Regulations 2104/2022 and 2105/2022 for virgin olive oil classification.

2.4. 2.4 HPLC-DAD analysis of phenolic compounds

Phenolic compounds were analyzed according to the official method of the International Olive Council (COI, 2022). Briefly, an aliquot of 2 g of sample was extracted with MeOH:H_2O (80:20, 6 mL including syringic acid as the internal standard). A HP1100 liquid chromatography system equipped with an autosampler, a HP1100 Diode Array Detector (Agilent Technology, California, USA) and a SphereClone ODS column (5 μm, 250 × 4.6 mm i.d.) (Phenomenex, Bologna, Italy) was used to analyze the extract. The injection volume was 20 μL. The eluents were acidified water (pH = 2.0, H_3PO_4), MeOH, and acetonitrile, all with HPLC grade. Elution was performed according to the IOC official method, and chromatograms were acquired at 280 nm. The internal standard method was used to determine the contents of the total phenols (TPC) and individual phenolic compounds. Syringic acid was used as the internal standard, while tyrosol was employed as the reference compound; their response factors were calculated, as well as the relative response factor (RRF) given by the ratio of the two response factors. Following this, a total of 24 compounds were quantitated (Table S1 of

Table 1
A) The sample-set of monovarietal extra virgin olive oils (MEVOOs) used for defining the clusters of cultivars and building the models; B) the independent test-set of MEVOO samples used for external validation of the proposed model. "EVOO#": samples with defect ≤ 1; "Defect": samples with defect > 1. Under the "Production year" columns, the number in the brackets indicates the number of samples of that cultivar that were labelled in the category "defect".

A)

Cultivar (Code)	Total	Quality		Production year			
		EVOO#	Defect	2020–21	2021–22	2022–23	2023–24
Ascolana Tenera (**AT**)	19	15	4	5	7 (1)	7 (3)	0
Bosana (**BO**)	14	10	4	1 (1)	4	7 (3)	2
Coratina (**CO**)	47	46	1	15	9	16	7 (1)
Frantoio (**FR**)	27	25	2	10	4	10	3 (2)
Itrana (**IT**)	18	14	4	3	3	9 (3)	3 (1)
Leccio del Corno (**LC**)	12	12	0	2	4	5	1
Leccino (**LE**)	27	20	7	8	11 (4)	8 (3)	0
Maurino (**MA**)	10	10	0	4	1	5	0
Moraiolo (**MO**)	25	22	3	10	4	6 (2)	5 (1)
Nocellara (**NO**)	29	21	8	5	8 (3)	10 (2)	6 (3)
Pendolino (**PE**)	8	8	0	5	2	1	0
Peranzana (**PR**)	24	17	7	7 (1)	8 (3)	8 (3)	1
Tonda Iblea (**TI**)	22	20	2	7	2	4	9 (2)
Total	**282**	**240**	**42**	**82 (2)**	**67 (11)**	**96 (19)**	**37 (10)**

B)

Cultivar (Code)	Test-set (2024–25)	Quality	
		EVOO#	Defect
Ascolana Tenera (**AT**)	0	0	0
Bosana (**BO**)	5	5	0
Coratina (**CO**)	20	20	0
Frantoio (**FR**)	9	9	0
Itrana (**IT**)	7	7	0
Leccio del Corno (**LC**)	5	5	0
Leccino (**LE**)	4	4	0
Maurino (**MA**)	5	5	0
Moraiolo (**MO**)	4	4	0
Nocellara (**NO**)	3	3	0
Pendolino (**PE**)	2	2	0
Peranzana (**PR**)	3	3	0
Tonda Iblea (**TI**)	7	7	0
Total	**74**	**74**	**0**

Supplementary Material) using a single response factor, and all results were expressed as mg_{tyr}/Kg_{oil}.

2.5. HS-SPME-GC-MS analysis of volatile organic compounds

Volatile organic compounds (VOCs) associated with virgin olive oil (VOOx) sensory defects and the C5 and C6 VOCs originating from the LOX pathway (LOX VOCs) were analyzed using previous validated method (Cecchi et al., 2019). Briefly, 4.3 g of oil and 0.1 g of internal standard solution were weighted into a screw-cap vial (total volume: 20 mL). Volatile compounds were concentrated onto a DVB/CAR/PDMS 1-cm 50/30 μm SPME fiber (Agilent, Palo Alto, CA, USA) by exposing the vials headspace at 45 °C for 20 min. GC-MS conditions and quantitative details are described in previous article (Cecchi et al., 2019). Only data for the 18 C5 and C6 LOX VOCs were processed as described in the following sections. These compounds are listed in Table S2 of the Supplementary Material, along with the seven C10 pentene dimers analyzed as described in Section 2.6.

2.6. HS-SPME-GC-MS analysis of volatile hydrocarbon compounds

Volatile hydrocarbon compounds (VHCs), including monoterpenes, sesquiterpenes, and pentene dimers (i.e., seven C10 molecules originating from the LOX pathway), were analyzed using a validated method (Cecchi et al., 2024). Briefly, 3.27 g of oil and 0.1 g of internal standard solution were weighted into a 20-mL screw-cap vial. VHCs were concentrated onto a DVB/CAR/PDMS 1-cm 50/30 μm SPME fiber (Agilent, Palo Alto, CA, USA) by exposing the vials headspace at 90 °C for 60 min. GC-MS conditions and quantitative details are provided in previous article (Cecchi et al., 2024) A total of 59 VHCs were quantitated, including 24 monoterpenes, 28 sesquiterpenes (Table S3 of Supplementary Material), and seven C10 pentene dimers (see Table S2).

2.7. Sensory analysis by the Panel Test

Sensory analysis was performed using the Panel Test method (https://www.internationaloliveoil.org/wp-content/uploads/2024/07/III-2.2.COI-T20-Doc.-15-REV-11–2024_EN-1.pdf). The Sensory Panel employed in this study consisted of a panel leader and 8–12 trained assessors and was officially recognized by the Italian Ministry of Agricultural Policies (MIPAAF). This approval is obtained through participation in an annual national proficiency ring test, and its continuation is contingent upon the panel's performance in a yearly evaluation conducted by the Member State, with renewal granted only when all parameters assessing alignment with other national panels are met. The panel leader and the panellists who took part in the evaluations over the five-year period were the same throughout, and were those belonging to the professional ANAPOO panel recognized by the MIPAAF. Samples were evaluated immediately after bottle opening. Panellists rated each attribute on a scale from 0 to 10, and mean scores were determined.

The analysis was carried out adopting a detailed profile sheet kindly provided by the ANAPOO association (Montevarchi, Arezzo, Italy), which allows the evaluation both of sensory defects and the fruity, bitter, and pungent attributes, and of a series of positive attributes potentially related to the sensory differences that may occur in high-quality monovarietal extra virgin olive oils. In particular, the profile sheet is divided in two consecutive sections. The first section is designed to evaluate negative and positive sensory attributes for virgin olive oil classification according to European Regulations 2104/2022 and 2105/2022, covering the first seven attributes listed in Table S4. Samples were initially classified as EVOO if the median of defect score was 0 and the median of fruity score was greater than 0, or as non-EVOO if the median of fruity score was 0 or the median of defect score exceeded 0. This approach, based on the requirement that an oil must be completely free of defects to be classified as EVOO, is aimed at commercial classification; however, this research is focused on the recognizability of the intrinsic characteristics of the cultivar to which the oil belongs, not on its commercial categorization. Since a barely perceptible defect does not prevent the recognition of the cultivar's intrinsic traits, it was considered inappropriate to exclude an oil in this case. Simultaneously, once a defect becomes predominant, it prevents the cultivar fingerprint from being recognized. After preliminary clustering experiments showing that choosing a median defect threshold of 0.5, 1, or 1.5 led to separate defective samples from the other clusters, for the purpose of this study, it was assumed that a sample loses the sensory characteristics typical of its cultivar of origin when the median defect score exceeds 1. This value was adopted as a compromise between a threshold that is not too low (to minimize the risk of excluding oils in which the intrinsic characteristics of the cultivar are still clearly recognizable) and one that is not too high (to minimize the risk of including oils in which the defect is clearly predominant over the cultivar characteristics). Therefore, each sample with a defect median score between 0 and 1 was still labelled with its cultivar name ("**EVOO#**" in Table 1), while samples with a defect median score greater than 1 were labelled as defective samples ("**defect**" in Table 1). The second section was designed to evaluate several orthonasal and retronasal positive attributes that specifically characterize the unique sensory features of MEVOOs (i.e., from attribute 8–54 in Table S4).

2.8. Data analysis

Various data analysis methods were applied using R software (version 4.3.1, The R Foundation for Statistical Computing), including the Shapiro–Wilk test, Kruskal–Wallis test, Dunn's test, Random Forest (RF), Linear Discriminant Analysis (LDA), Hierarchical Cluster Analysis (HCA), Jaccard cluster stability analysis and consensus matrix computation. An original computer program was also developed in Visual Basic (VB6, Microsoft, USA) to process data using a genetic algorithm (*GA*) method (Forrest, 1993), and the pseudocode of the GA-LDA program can be found in Annex II of the Supplementary file. Classification of samples was performed using two supervised approaches: *GA-LDA* and *RF*.

GA-LDA: the *LDA* method calculates a discriminant function that best separates samples into predefined classes by combining a pre-established number of variables. Since literature data indicate that *LDA*-based models should be coupled with an appropriate stepwise algorithm for variables selection (Aparicio-Ruiz et al., 2019), in this study the *GA* method was used for variable selection. *GA* is a metaheuristic, iterative method inspired by natural selection processes, commonly applied to finding high-quality solutions to complex problems. It is particularly well-suited to the problem under study, where numerous descriptors of different types were measured across many samples. The iterative process begins with a set (the population) of randomly generated combinations of variables (the individuals). The performance (fitness) of each individual in the population is evaluated, and those with better performance are selected. The selected individuals (elite individuals) undergo genetic operations such as crossover (recombination) and mutation, generating a new population. This process is repeated until an optimal number of generations is reached. The following *GA* parameters, selected after some preliminary trials, were set in this study:

- Number of variables per individual (combination of variables): 4–20
- Number of individuals in the initial population (population size): 100
- Number of generations (populations) was established when plateauing of the score value
- Number of selected individuals (elite individuals): 50
- Crossover probability: 1 (all elite individuals)
- Mutation probability: random.

The *LDA* method was used as the fitness function to evaluate individuals. Accuracy and Class Error Average (*CEA*) were obtained via 5-fold random cross-validation, and the score (S) of the fitness function

was calculated by "S = 100 - *CEA*" so that it would not be influenced by class imbalance. Specifically, accuracy was calculated as the number of correctly classified samples divided by the total number of samples, while *CEA* was calculated as the percentage of misclassified samples in each class, averaged across the total number of classes, where misclassified samples are those that were assigned to the wrong class. Figure S1 in the Supplementary Material shows the flowchart of the applied *GA-LDA* method, with also a more detailed explanation of how parameters were set.

RF: the *RF* method is a machine learning algorithm that combines the output of multiple decision trees to reach a single result (Breiman, 2001). The decision trees are built independently using different subsets of data and random combinations of variables. This process, known as "bagging" (Bootstrap Aggregating), ensures that each tree differs slightly from the others, thereby reducing correlation between trees and improving the overall accuracy of the model. In this study, RF models were trained using the default parameters of the "caret" package in R. Specifically, the number of tree was 500, number of variables tried at each split mtry = sqrt(p) with tuneLenght = 3, nodesize = 1, sampsize = nrow(data), importance = TRUE, and trainControl (method= 'repeatedcv', number = 5).

The *RF* results were optimized by minimizing the Out-of-Bag (OOB) error obtained via 5-fold random cross-validation. The overall accuracy was also calculated as the number of correctly classified samples divided by the total number of samples.

Both the models were externally validated using the external test-set of samples from the 2024–2025 production year.

The unsupervised *HCA* method was applied to determine the clustering of MEVOO samples using the normalized confusion matrices from the *GA-LDA* and *RF* methods applied to the data from the Panel Test. This unconventional clustering approach, in which the confusion matrices are interpreted as element of similarity among cultivars, was adopted following the aim of grouping cultivars based on the similarity of their sensory descriptors, rather than generating clusters of individual oils; further details can be found in Annex III of Supplementary file. Hierarchical clustering was performed using two distance measures (Euclidean and Manhattan) and three linkage methods (Average, Complete and Ward2). Since the *GA-LDA* and *RF* methods were each repeated three times, a total of $2 \times 3 \times 3 = 18$ dendrograms were generated for each method, yielding a total of 36 outputs.

The list of compounds included in every model is summarized in Table S5.The Shapiro-Wilk test was used to assess the normality of the data distributions, revealing that they were not normally distributed. Consequently, the Kruskal-Wallis test, a non-parametric method suitable for non-normal distributions, was applied to evaluate which sensory or chemical variables significantly discriminated between the defined clusters of cultivars based on *p*-values. Dunn's test was then used to determine, for each variable, the significance of differences among the clusters of cultivars at $p > 0.05$. Box-and-whisker plots were also generated.

3. Results and discussion

3.1. From sensory profiles to MEVOO clustering

A data matrix was created with the 282 samples summarized in Table 1A as rows and the 54 sensory attributes (i.e., the variables, Table S4) as columns. Rows were labelled with either the name of sample's cultivar or as "**defect**" for samples with a median defect score > 1, regardless of cultivar. The columns contained the scores for each sensory attribute. *GA-LDA* and *RF* methods were applied to test the ability of the sensory attributes to classify samples into 14 classes (i.e., one for each cultivar included in the research along with the class "**defect**"). No model was able to satisfactorily assign sample to their correct cultivar, confirming findings from previous studies (Cecchi et al., 2022; Ugolini et al., 2024).

The confusion matrices obtained from the above application of *GA-LDA* and *RF* methods were subsequently applied in combination to *HCA* to group the cultivars into clusters based on similarities in their sensory profiles. To assess the robustness of clustering, *HCA* was performed across multiple methodological perturbations, including two distance measures and three linkage methods, on the 6 confusion matrices generated by repeated runs of *GA-LDA* and *RF* methods (see Section 2.8), yielding a total of 36 clustering solutions, each given by four clusters of cultivars. Four different combinations of clusters were obtained (Table S6 of Supplementary Material), and clusters stability was here evaluated by identifying identical clustering solutions and calculating their frequency of occurrence across all solutions (that is a consensus-based approach where the frequency represents the percentage a given clustering solution emerged across all 36 combinations of models, repetitions, distances, and linkage criteria). Three out of these four clustering solutions, with overall frequencies of 72.5%, 16.5%, and 5.5% respectively (i.e., totalling 94.5%), identified a cluster exclusively composed of samples labelled as "**defect**" regardless of cultivar, suggesting that, when pronounced sensory defects are present, the positive attributes characteristic of MEVOOs are masked and confirming the appropriateness of the selected threshold (i.e., median > 1).

Based on this finding, the 42 "**defect**" samples (Table 1A) were removed from the data matrix, and the clustering process was repeated. Using the same approach, *HCA* applied to the confusion matrices derived from *GA-LDA* and *RF* resulted in six different combinations of three clusters of cultivars (Table 2). The stability of the different clustering solutions obtained was evaluated both as described above (frequencies calculation) and as Jaccard stability (Table S7 in Supplementary file); clustering #1 and #2 consistently exhibited the highest overall frequencies (61.5% and 19.5%, respectively) and the higher stability according to Jaccard index (0.940 and 0.879, respectively). They differed only in the placement of the Ascolana Tenera (**AT**) cultivar. In clustering #2, this cultivar displayed sensory profile similarities with Itrana (**IT**), Nocellara (**NO**) and Tonda Iblea (**TI**), while in clustering #1, it clustered with Bosana (**BO**) and Peranzana (**PR**). This situation is furtherly confirmed by the consensus matrix reported in Table S8 of Supplementary file: a first block of cultivars (PE, CO, LE, FR, LC, MA, MO) showed perfect stability (stability = 1), a second block (TI,

Table 2
Clusters based on the similarity of sample sensory profiles, after removing "**Defect**" samples from the data matrix. The percentage indicates the frequency of each clustering, which was determined by processing the confusion matrices generated by Genetic Algorithm-Linear Discriminant Analysis (*GA-LDA*) and Random Forest (*RF*) methods. In the last column, the percentage represents the overall frequency given by the mean outputs of the *GA-LDA* and *RF* methods.

	Clustering	% (GA-LDA)	% (RF)	% (TOT)
#1	1. PE, CO, LE, FR, LC, MA, MO 2. TI, IT, NO 3. AT, BO, PR	56	67	**61.5**
#2	1. PE, CO, LE, FR, LC, MA, MO 2. TI, IT, NO, AT 3. BO, PR	17	22	**19.5**
#3	1. PE, CO, LE, FR, LC, MA, MO 2. TI, NO 3. AT, BO, PR, IT	22	-	**11**
#4	1. PE, CO, LE, FR, LC, MA, MO 2. TI, NO, IT, BO 3. AT, PR	-	6	**3**
#5	1. PE, CO, LE, FR, LC, MA, MO 2. TI, IT, NO, PR, AT 3. BO	5	-	**2.5**
#6	1. PE, CO, LE, FR, LC, MA, MO 2. TI, IT, NO, BO, AT 3. PR	-	5	**2.5**

AT, Ascolana Tenera; **BO**, Bosana; **CO**, Coratina; **FR**, Frantoio; **IT**, Itrana; **LC**, Leccio del Corno; **LE**, Leccino; **MA**, Maurino; **MO**, Moraiolo; **NO**, Nocellara; **PE**, Pendolino; **PR**, Peranzana; **TI**, Tonda Iblea

IT, NO) and a third block of cultivars (BO, PR) were almost always grouped together (high stability, 0.89–1). In contrast, the cultivar AT exhibited intermediate co-clustering values (0.25–0.778), confirming its unstable positioning discussed above. Notably, the sensory similarity of **AT** with **IT**, **NO**, and **TI** had already been reported (https://www.olimonovarietali.it/database/le-tipologie-sensoriali/). To determine which of the two defined clustering solutions was more reliable, the consistency of the predictive models based on chemical data was assessed, as described in Section 3.2

The role of sensory variables in discriminating among the defined clusters in both clustering #1 and #2 was preliminarily evaluated. The importance of variables determined from application of GA-LDA and RF was considered. The nine most important sensory variables were: "Bitterness", "coffee", "fruity", "lettuce", "pepper", "thistle", "tomato fruit", "tomato leaf" and "yellow apple". The Kruskal-Wallis test was applied to these variables, followed by Dunn's test to assess significant differences among clusters at $p < 0.05$. Finally, box-and-whisker plots were generated (Fig. 1). Overall, all clusters exhibited medium-to-high intensity of legally recognized sensory attributes, such as "fruity" and "bitterness". The attribute most significant in discriminating the three clusters of cultivars was "tomato fruit". Specifically, cluster 1 in both clustering types (namely #1 and #2) was characterized by the highest intensities of "bitterness", "thistle" and "coffee", and by the lowest intensities of "tomato fruit" and "yellow apple". Cluster 2 displayed more complex behaviour, attributable to the absence or presence of the **AT** cultivar in clustering #1 and #2, respectively. It was distinguished by the highest intensities of "tomato fruit", "tomato leaf", "fruity", "lettuce" and "pepper", and the lowest intensity of "bitterness". The intensities of "lettuce", "pepper" and "bitterness" significantly differentiated clusters 2 from clusters 1 in both clustering #1 and #2, while these attributes distinguished cluster 2 from cluster 3 only in clustering #1. Finally, cluster 3 exhibited significantly higher intensity of "tomato fruit" than clusters 1 in both clustering #1 and #2. It was also characterized by the highest intensity of "yellow apple", which significantly differentiated clusters 3 from clusters 1 in both clustering #1 and #2, although it was not significant in differentiating clusters 3 from clusters 2.

3.2. Relationships between sensory clusters and minor chemical compounds

Chemical data on molecules most frequently associated in the literature with EVOO sensory attributes, namely phenolic compounds, LOX VOCs and terpenes (Campestre et al., 2017; Cecchi & Alfei, 2013; Kesen et al., 2013; Torres-Cobos et al., 2021; Ugolini et al., 2024) were analyzed to assess the relationships between these compounds and the cultivar clusters defined based on sensory data (see Section 3.1), identifying which compounds were most effective in explaining the differences among clustering #1 and #2 (Table 2).

A data matrix was created with samples in rows and chemical compounds in columns, reporting the quantitative data of each compound in each sample. The 240 samples in the rows, excluding the 42 "**defect**" samples (Table 1A), were labelled according to their cultivar cluster membership. The chemical compounds in the columns comprised 24 phenolic compounds, 25 LOX VOCs (including 6 C5, 12 C6, and 7 C10 compounds), and 52 terpenes (Table S1-S3 of Supplementary Material); VOCs associated with virgin olive oil sensory defects were not included, as they were considered unsuitable for differentiating MEVOO samples based on their typical sensory profiles. Both *GA-LDA* and *RF* methods were applied, and the discriminatory capabilities of phenolic compounds, LOX-related VOCs and terpenes were initially examined separately. The confusion matrices obtained are shown in Table 3, and Class Error Average (CEA) and accuracy for MEVOOs clustering #1 and #2 were compared in Table 4.

The strongest relationship with sensory clusters was observed for terpenes in both clustering #1 and #2, confirming the crucial role of these molecules in the sensory nuances of MEVOOs (Ugolini et al., 2024). For both *GA-LDA* and *RF* methods, the best results (i.e., highest accuracy of approx. 93% and lowest CEA of 8%) were achieved for terpenes and clustering #2, with correctly classified samples percentages never falling below 84% (see confusion matrices "RF - Clustering 2"

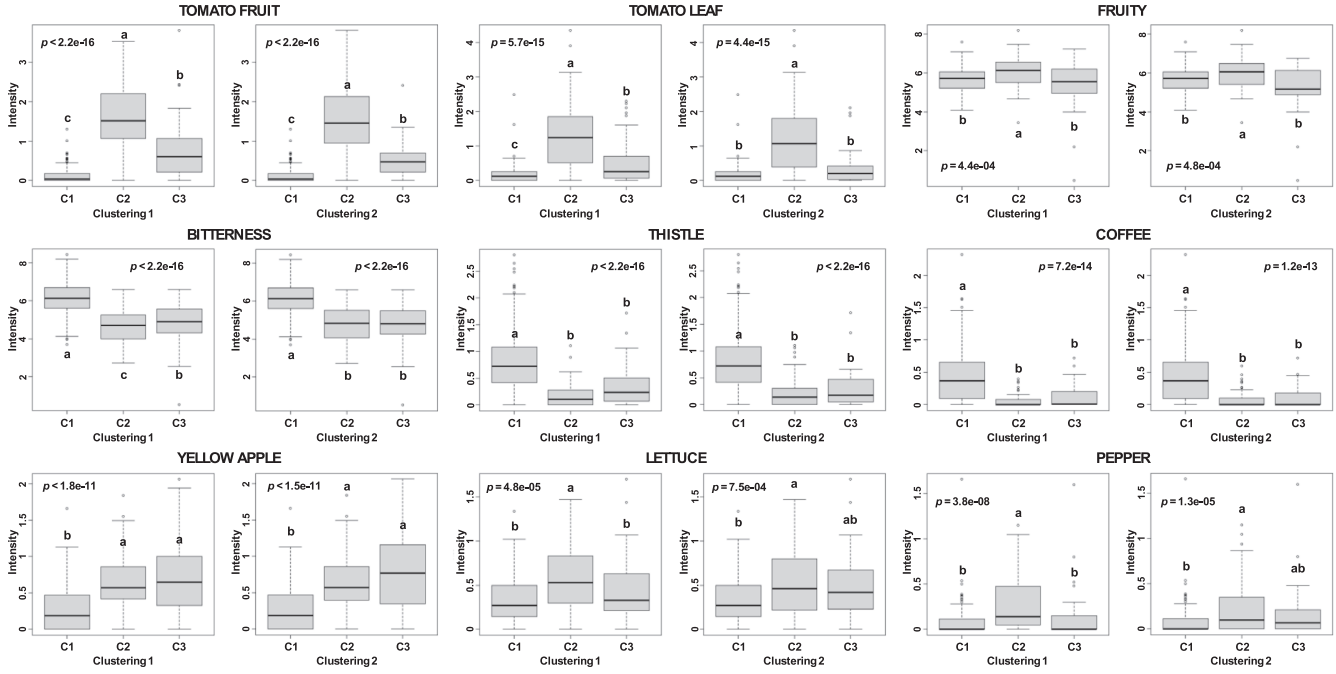


Fig. 1. Box-and-whisker plot of the sensory attributes selected as able in discriminating clusters of cultivars. For each attribute, different letters indicate statistically significant differences among cultivars at $p < 0.05$, according to the Dunn's test. The *p*-value from the Kruskal-Wallis test is also reported. **Clustering #1** – Cluster 1 (**C1**): Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; **C2**: Tonda Iblea, Itrana, Nocellara; **C3**: Ascolana Tenera, Bosana, Peranzana. **Clustering #2** – **C1**: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; **C2**: Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; **C3**: Bosana, Peranzana.

Table 3
Confusion matrices obtained applying Genetic Algorithm-Linear Discriminant Analysis (*GA-LDA*) and Random Forest (*RF*) methods to test the ability of (**A**) phenolic compounds, (**B**) LOX VOCs and (**C**) terpenes in discriminating samples in the three clusters of cultivars according to both clustering #1 and #2. *CEA,* Class Error Average.

A)Phenolic compounds

GA-LDA - Clustering 1					**GA-LDA - Clustering 2**				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**127 (88.8%)**	4 (2.8%)	12 (8.4%)	1	143	**132 (92.3%)**	6 (4.2%)	5 (3.5%)
2	55	2 (3.6%)	**47 (85.5%)**	6 (10.9%)	2	70	6 (8.6%)	**60 (85.7%)**	4 (5.7%)
3	42	6 (14.3%)	5 (11.9%)	**31 (73.8%)**	3	27	5 (18.5%)	7 (25.9%)	**15 (55.6%)**
Total	**240**	Accuracy = 85.4%			**Total**	**240**	Accuracy = 86.3%		
		Class Error Average = 17.3%					Class Error Average = 22.1%		
RF - Clustering 1					**RF - Clustering 2**				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**132 (92.3%)**	3 (2.1%)	8 (5.6%)	1	143	**134 (93.7%)**	6 (4.2%)	3 (2.1%)
2	55	3 (5.4%)	**47 (85.5%)**	5 (9.1%)	2	70	9 (12.9%)	**60 (85.7%)**	1 (1.4%)
3	42	14 (33.3%)	8 (19.0%)	**20 (47.7%)**	3	27	10 (37.0%)	11 (40.8%)	**6 (22.2%)**
Total	**240**	Accuracy = 82.9%			**Total**	**240**	Accuracy = 83.3%		
		Class Error Average = 24.8%					Class Error Average = 32.8%		

B)LOX VOCs

GA-LDA - Clustering 1					**GA-LDA - Clustering 2**				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**136 (95.1%)**	4 (2.8%)	3 (2.1%)	1	143	**133 (93.0%)**	10 (7.0%)	0 (0%)
2	55	3 (5.5%)	**50 (90.9%)**	2 (3.6%)	2	70	6 (8.6%)	**62 (88.6%)**	2 (2.8%)
3	42	10 (23.8%)	5 (11.9%)	**27 (64.3%)**	3	27	4 (14.8%)	1 (3.7%)	**22 (81.5%)**
Total	**240**	Accuracy = 88.8%			**Total**	**240**	Accuracy = 90.4%		
		Class Error Average = 16.6%					Class Error Average = 12.3%		
RF - Clustering 1					**RF - Clustering 2**				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**132 (92.3%)**	6 (4.2%)	5 (3.5%)	1	143	**137 (95.8%)**	6 (4.2%)	0 (0%)
2	55	4 (7.3%)	**48 (87.3%)**	3 (5.4%)	2	70	10 (14.3%)	**58 (82.8%)**	2 (2.9%)
3	42	9 (21.4%)	4 (9.6%)	**29 (69.0%)**	3	27	2 (7.4%)	5 (18.5%)	**20 (74.1%)**
Total	**240**	Accuracy = 87.1%			**Total**	**240**	Accuracy = 89.6%		
		Class Error Average = 17.1%					Class Error Average = 15.8%		

C)Terpenes

GA-LDA - Clustering 1					GA-LDA - Clustering 2				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**137 (95.8%)**	4 (2.8%)	2 (1.4%)	1	143	**135 (94.4%)**	7 (4.9%)	1 (0.7%)
2	55	6 (10.9%)	**49 (89.1%)**	0 (0%)	2	70	11 (15.7%)	**59 (84.3%)**	0 (0%)
3	42	15 (35.7%)	2 (4.8%)	**25 (59.5%)**	3	27	1 (3.7%)	0 (0%)	**26 (96.3%)**
Total	**240**	Accuracy = 87.9%			**Total**	**240**	Accuracy = 91.7%		
		Class Error Average = 18.5%					Class Error Average = 8.3%		
RF - Clustering 1					**RF - Clustering 2**				
Cluster	Samples	1	2	3	Cluster	Samples	1	2	3
1	143	**139 (97.2%)**	2 (1.4%)	2 (1.4%)	1	143	**138 (96.5%)**	4 (2.8%)	1 (0.7%)
2	55	8 (14.5%)	**47 (85.5%)**	0 (0%)	2	70	9 (12.9%)	**61 (87.1%)**	0 (0%)
3	42	9 (21.4%)	5 (11.9%)	**28 (66.7%)**	3	27	2 (7.4%)	0 (0%)	**25 (92.6%)**
Total	**240**	Accuracy = 89.2%			**Total**	**240**	**Accuracy = 93.3%**		
		Class Error Average = 16.9%					**Class Error Average = 7.9%**		

Clustering #1: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara; 3: Ascolana Tenera, Bosana, Peranzana. **Clustering #2**: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; 3: Bosana, Peranzana.

Table 4
Comparison of accuracy (A) and Class Error Average (CEA) of chemometrics models based on the Genetic Algorithm-Linear Discriminant Analysis (*GA-LDA*) and Random Forest (*RF*) approaches to predict relationships between chemical compounds content and Monovarietal Extra Virgin Olive Oils (MEVOOs) classification in MEVOOs clustering #1 and #2.

Chemical Classes	***GA-LDA***				***RF***			
	Clustering #1		**Clustering #2**		**Clustering #1**		**Clustering #2**	
	A (%)	***CEA (%)***	***A (%)***	***CEA (%)***	***A (%)***	***CEA (%)***	***A (%)***	***CEA (%)***
Phenolic compounds (PCs)	85.4	17.3	86.3	22.1	82.9	24.8	83.3	32.8
LOX VOCs	88.8	16.6	90.4	12.3	87.1	17.1	89.6	15.8
Terpenes	87.9	18.5	91.7	8.3	89.2	16.9	93.3	7.9
Terpenes + LOX VOCs	90.8	13.9	95	5.8	91.7	14.2	95.8	5.3
Terpenes + LOX VOCs + PCs	93.3	10.5	96.3	4.8	94.6	9.2	97.5	3.1

Clustering #1: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara; 3: Ascolana Tenera, Bosana, Peranzana. **Clustering #2**: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; 3: Bosana, Peranzana.

and "GA-LDA - Clustering 2" in Table 3C). Conversely, the weakest relationship was observed between terpenes and clustering #1 (with lower accuracy and higher CEA), due to a high rate of misclassification in cluster 3 (comprising **AT**, **BO** and **PR** cultivars), where correctly classified samples percentages never exceeded 67% (see confusion matrices "RF - Clustering 1" and "GA-LDA - Clustering 1" in Table 3C). Thus, terpene content provides a better explanation for the placement of AT in cluster 2 alongside TI, IT and NO cultivars in clustering #2.

To further improve the discriminatory power of terpenes, combinations of chemical compound classes was tested, as follows: (i) terpenes and LOX VOCs; (ii) terpenes, LOX VOCs and phenolic compounds. The confusion matrices obtained are presented in Table 5, and CEA and Accuracy percentage values are compared in Table 4. For both *GA-LDA* and *RF* methods, combining terpenes with other compound classes enhanced the discriminating ability. The best performance was confirmed for clustering #2: combining terpenes, LOX VOCs, and phenolic compounds yielded an accuracy of up to 98% and a CEA of 3%, with correctly classified samples percentages never dropping below 92% (see confusion matrices "RF - Clustering 2" and "GA-LDA - Clustering 2" in Table 5B). Performance for clustering #1 was also improved, achieving 95% accuracy and 9% CEA when using the combination of terpenes, LOX VOCs and phenolic combination, although a relatively high rate of misclassification persisted for cluster 3, where correctly classified samples percentages never exceeded 79% (see confusion matrices "RF - Clustering 1" and "GA-LDA - Clustering 1" in Table 5B). Based on this evidence, Clustering 2 was selected as the more reliable.

The models using all the three classes of compounds (i.e., terpenes, LOX VOCs and phenolic compounds) for classifying samples in clusters of clustering #2 were those which gave the best performance. Therefore, they were two-fold validated: a leave-one-year-out validation was conducted to assess potential effect of the production year, while for assessing potential model overfitting, an independent test-set comprising samples from the 2024–2025 production year was used (Table 1B). Table S9 summarizes results of the leave-one-year-out validation, pointing out good stability over production years of both clusters and model performance; better results were obtained using RF for all the four production years with accuracy ranging from 92.9% (2021–2022) to 98.8% (2020–2021) and *CEA* ranging from 1.7% (2020–2021) to 10.0% (2020–2021). Results of external validation with samples of 2024–2025 are shown in Table 6, confirming very satisfactory performances for both models, again better for RF, with accuracy 93.2% and *CEA* 8.2%. As expected, these results are slightly lower than those obtained in internal validation, indicating a generalization gap due to the complexity of the model. Indeed, while these good results indicates that the models are reasonably robust to year-to-year variations and that cluster composition is stable over the years, a slight effect of the year also emerges probably together with that of other factors that may affect oil characteristics.

3.3. Simplified chemometric models for MEVOOs classification in clustering #2

Based on the relationships between chemical compounds and cultivar clusters described in Section 3.2, simplified chemometric models were developed to classify MEVOO samples into the three clusters of clustering #2. The aim was to preliminarily identify the smallest combination of variables capable of discriminating among the three clusters of cultivars. This strategy was applied using both terpenes alone and the combination of terpenes, LOX VOCs and phenolic compounds.

Table 5
Confusion matrices obtained applying Genetic Algorithm-Linear Discriminant Analysis (*GA-LDA*) and Random Forest (*RF*) methods to test ability of combined contents of (**A**) terpenes + LOX VOCs and (**B**) terpenes + LOX VOCs + phenolic compounds to differentiate samples in the clustering #1 and #2. *CEA*, Class Error Average.

A) Terpenes + LOX VOCs

GA-LDA - Clustering 1

Cluster	Samples	1	2	3
1	143	**137 (95.8%)**	5 (3.5%)	1 (0.7%)
2	55	1 (1.8%)	**53 (96.4%)**	1 (1.8%)
3	42	10 (23.8%)	6 (14.3%)	**26 (61.9%)**
Total	**240**	Accuracy = 90%		
		Class Error Average = 15.3%		

GA-LDA - Clustering 2

Cluster	Samples	1	2	3
1	143	**135 (94.4%)**	7 (4.9%)	1 (0.7%)
2	70	6 (8.6%)	**64 (91.4%)**	0 (0%)
3	27	2 (7.4%)	0 (0%)	**25 (92.6%)**
Total	**240**	Accuracy = 93.3%		
		Class Error Average = 7.2%		

RF - Clustering 1

Cluster	Samples	1	2	3
1	143	**141 (98.6%)**	1 (0.7%)	1 (0.7%)
2	55	2 (3.6%)	**52 (94.6%)**	1 (1.8%)
3	42	11 (26.2%)	4 (9.5%)	**27 (64.3%)**
Total	**240**	Accuracy = 91.6%		
		Class Error Average = 14.2%		

RF - Clustering 2

Cluster	Samples	1	2	3
1	143	**139 (97.2%)**	3 (2.1%)	1 (0.7%)
2	70	4 (5.7%)	**66 (94.3%)**	0 (0%)
3	27	2 (7.4%)	0 (0%)	**25 (92.6%)**
Total	**240**	**Accuracy = 95.8%**		
		Class Error Average = 5.3%		

B) Terpenes + LOX VOCs + Phenolic compounds

GA-LDA - Clustering 1

Cluster	Samples	1	2	3
1	143	**138 (96.5%)**	2 (1.4%)	3 (2.1%)
2	55	0 (0%)	**54 (98.2%)**	1 (1.8%)
3	42	9 (21.4%)	4 (9.5%)	**29 (69.1%)**
Total	**240**	Accuracy = 92.1%		
		Class Error Average = 12.1%		

GA-LDA - Clustering 2

Cluster	Samples	1	2	3
1	143	**139 (97.2%)**	4 (2.8%)	0 (0%)
2	70	5 (7.1%)	**65 (92.9%)**	0 (0%)
3	27	1 (3.7%)	1 (3.7%)	**25 (92.6%)**
Total	**240**	Accuracy = 95.4%		
		Class Error Average = 5.8%		

RF - Clustering 1

Cluster	Samples	1	2	3
1	143	**142 (99.3%)**	0 (0%)	1 (0.7%)
2	55	1 (1.8%)	**52 (94.6%)**	2 (3.6%)
3	42	4 (9.5%)	5 (11.9%)	**33 (78.6%)**
Total	**240**	Accuracy = 94.5%		
		Class Error Average = 9.2%		

RF - Clustering 2

Cluster	Samples	1	2	3
1	143	**141 (98.6%)**	1 (0.7%)	1 (0.7%)
2	70	3 (4.3%)	**67 (95.7%)**	0 (0%)
3	27	1 (3.7%)	0 (0%)	**26 (96.2%)**
Total	**240**	**Accuracy = 97.5%**		
		Class Error Average = 3.2%		

Clustering #1: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara; 3: Ascolana Tenera, Bosana, Peranzana. **Clustering #2**: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; 3: Bosana, Peranzana

Table 6
Confusion matrices obtained during external validation of the Genetic Algorithm-Linear Discriminant Analysis (*GA-LDA*) and Random Forest (*RF*) models applied on the combined data of terpenes + LOX VOCs + phenolic compounds to differentiate samples of the test-set (Table 2B) in clustering #2. *CEA*, Class Error Average.

GA-LDA - Clustering 2				
Cluster	Samples	1	2	3
1	49	**46 (93.9%)**	3 (6.1%)	0 (0%)
2	17	2 (11.8%)	**15 (88.2%)**	0 (0%)
3	8	2 (25%)	1 (12.5%)	**5 (62.5%)**
Total	**74**	Accuracy = 89.2%		
		Class Error Average = 18.5%		
RF - Clustering 2				
Cluster	Samples	1	2	3
1	49	**46 (93.9%)**	3 (6.1%)	0 (0%)
2	17	1 (5.9%)	**16 (94.1%)**	0 (0%)
3	8	1 (12.5%)	0 (0%)	**7 (87.5%)**
Total	**74**	Accuracy = 93.2%		
		Class Error Average = 8.2%		

Clustering #2: 1: Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; 2: Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; 3: Bosana, Peranzana

Since both GA-LDA and RF consistently showed very satisfactorily performance, the simplified chemometric models were developed exclusively using the *GA-LDA* method because it allows the number of variables to be selected to be defined a priori. For variables selection, all variables were ranked according to the frequency with which they appeared in the 10 best combinations of 12 variables selected by the *GA*. Only variables with a frequency of at least 30% were retained for subsequent analyses, that was 14 variables in the case of terpenes (Fig. 2A) and 16 variables when combining terpenes, LOX VOCs, and phenolic compounds (Fig. 2B). Subsequently, the *LDA* model was run multiple times in leave-one-out cross-validation mode, initially using a single variable, then progressively adding variables in decreasing order of selection frequency. Accordingly, the *LDA* model was run 14 times for the terpene data-set and 16 times for the combined data-set.

The predictive performance of the models as a function of the increasing number of variables is shown in Fig. 2. Good results were already obtained using only four variables: with terpenes, accuracy of 83.3% and CEA of 19.1% were achieved, while the combination of terpenes, LOX VOCs, and phenolic compounds yielded 87.9% accuracy and a CEA of 13.7%. Better predictive performance was obtained as the number of variables increased. For terpenes, significant improvement was observed up to eight variables (accuracy 90.0%; CEA, 10.3%), while for the combined data-set, significant improvement continued up to twelve variables (accuracy 95.4%; CEA 6.8%), approaching the accuracy and CEA achieved by the *GA-LDA* models without variable selection (see Section 3.2). In summary, the simplified chemometric models for classifying MEVOO samples into the three clusters of clustering #2 (which need to be further validated in future steps of the research) were based on the following discriminating molecules (shown in bold in Figs. 2A and 2B):

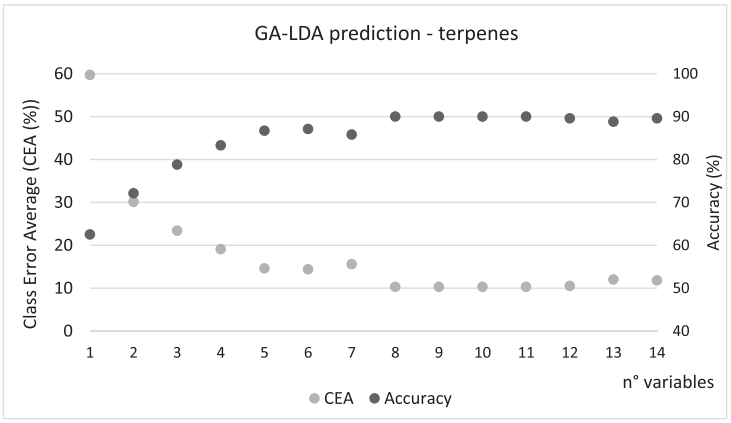
- Predictive model with terpenes only: four polycyclic sesquiterpenes with at least two fused rings (α-copaene, β-copaene, β-caryophylene, ylangene), one bicyclic bridged sesquiterpene (*cis*-bergamotene), one acyclic sesquiterpene ((*Z,E,E*)-allofarnesene), one cyclic monoterpene (γ-terpinene), and one oxygenated monoterpene (linalool).
- Predictive model with terpenes, LOX VOCs and phenolic compounds: two secoiridoid derivatives of oleuropein and ligstroside (dialdehydic form of oleuropein aglycone and oxidized oleocanthal, respectively), five LOX VOCs including one ester ((*Z*)-3-hexenyl acetate), three aldehydes ((*E*)-2-hexenal, (*E*)-2-pentenal and hexanal), and one alcohol (1-penten-3-ol), and five terpenes, comprising two bicyclic bridged sesquiterpenes (*cis*-bergamotene, sesquisabinene), one polycyclic sesquiterpene (α-cubebene), one acyclic sesquiterpene ((*E,Z,E*)-allofarnesene) and one acyclic monoterpene ((*E,E*)-cosmene).

To better understand the role of each individual molecule in characterizing the different clusters of cultivars, the Kruskal-Wallis test followed by the Dunn's test was applied, and box-and-whisker plots were generated (Fig. 3). Considering the molecules selected for the predictive model using only terpenes (Fig. 3A), cluster 1 (PE, CO, LE, FR, LC, MA, MO) was characterized by low levels of all selected terpenes, particularly exhibiting lower levels of (*Z,E,E*)-allofarnesene content compared to clusters 2 (TI, IT, NO, AT) and 3 (BO, PR). High contents of polycyclic sesquiterpenes with at least two fused rings (α-copaene, β-copaene, β-caryophyllene, ylangene) distinguished cluster 2. Cluster 3 exhibited the highest levels of *cis*-bergamotene, consistent with previous literature reporting elevated *cis*- and *trans*-bergamotene concentrations in Bosana and Peranzana MEVOOs (Ugolini et al., 2024). Regarding the molecules selected for the predictive model using combined terpenes, LOX VOCs and phenolic compounds (Fig. 3B), cluster 1 was primarily characterized by high levels of phenolic compounds such as oxidized oleocanthal and dialdehydic form of oleuropein aglycone, and low levels of (*Z*)-3-hexenyl acetate and (*E,Z,E*)-allofarnesene. Cluster 2 was distinguished by high contents of hexanal, (*E*)-2-pentenal, α-cubebene, and low levels of *cis*-bergamotene. Finally, cluster 3 was clearly characterized by much higher levels of sesquisabinene and *cis*-bergamotene, again consistent with earlier findings (Ugolini et al., 2024), along with high levels of (*E,Z,E*)-allofarnesene and (*Z*)-3-hexenyl acetate. The higher levels of C6 LOX VOC in the (*Z*)-3- configuration ((*Z*)-3-hexenyl acetate) in clusters 2 and 3, compared to cluster 1, and the higher content of a C6 LOX VOC in the (*E*)-2- configuration ((*E*)-2-hexenal) in cluster 1 than in clusters 2 and 3, suggest greater activity of the isomerase enzyme in cultivars belonging to cluster 1 (Cecchi et al., 2021; Peres et al., 2017).

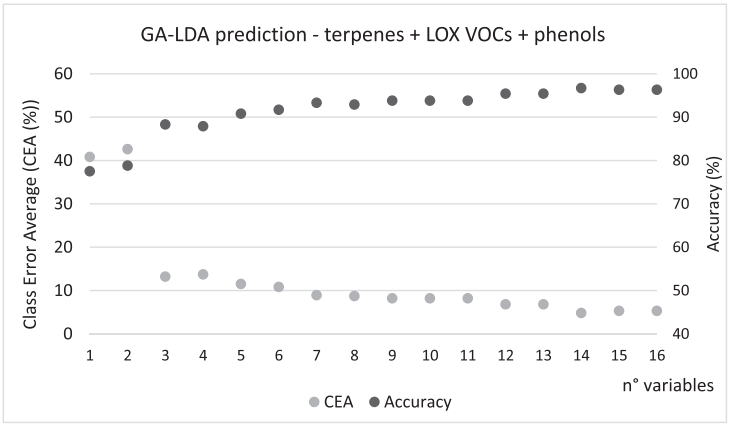
4. Conclusions

The *GA-LDA-HCA* and *RF-HCA* approaches robustly allowed grouping MEVOO samples in the same three clusters of cultivars with similar sensory and chemical features: Cluster 1 - Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; Cluster 2 - Tonda Iblea, Itrana, Nocellara, Ascolana Tenera; Cluster 3 - Bosana, Peranzana. Relationships between sensory profiles and chemical compounds were also pointed out. Cluster 1, characterized by the highest "bitterness", also exhibited the highest concentrations of phenolic compounds, aligning with the literature. Cluster 2, characterized by the highest values of "tomato leaf" and "tomato fruit", showed the highest concentration of specific terpenes (α-cubebene) and LOX VOCs (hexanal, (*E*)-2-pentenal, (*E*)-2-hexenal). Cluster 3, characterized by the highest value of "yellow apple", exhibited the highest concentrations of some terpenes (primarily *cis*-bergamotene and sesquisabinene) and a LOX ester ((*Z*)-3-hexenyl acetate), known for contributing to the sweet sensory notes of EVOO.

The clusters identified in this study are not definitive: other cultivars may be added to each of the three clusters, and additional clusters could also be defined should different sensory or chemical profiles emerge in cultivars not included in this study. However, possible practical applications already result from the proposed clustering approach, once transferability across independent laboratories has been demonstrated. The definition of entry classes based on MEVOOs with similar sensory profiles will lead to more homogeneous competitions in Italian EVOO competitive awards, preventing MEVOOs with markedly different profiles from competing against each other. Moreover, confirmation of the MEVOO's membership in one of the three sensory clusters could be supported by the proposed GA-LDA and RF models using combined LOX VOCs, terpenes and phenolic compounds data; simplified *GA-LDA* predictive models based on either eight terpenes or twelve combined



n°	Terpene	Frequency (%)
1	**α-copaene**	100
2	***cis*-bergamotene**	100
3	**β-copaene**	100
4	**β-caryophyllene**	100
5	**ylangene**	90
6	**γ-terpinene**	60
7	**linalool**	60
8	**(*Z*,*E*,*E*)-allofarnesene**	60
9	α-phellandrene	50
10	(*E*,*Z*)-alloocimene	50
11	(*E*,*Z*,*E*)-allofarnesene	50
12	(+)-cyclosativene	40
13	eucalyptol	30
14	α-muurolene	30



n°	Compound	Frequency (%)
1	**oxidized oleocanthal**	100
2	**hexanal**	90
3	***cis*-bergamotene**	70
4	**1-penten-3-ol**	60
5	**(*Z*)-3-hexenyl acetate**	60
6	**α-cubebene**	60
7	**dialdehydic oleur agl**	50
8	**(*E*)-2-pentenal**	50
9	**(*E*,*E*)-cosmene**	50
10	**sesquisabinene**	50
11	**(*E*,*Z*,*E*)-allofarnesene**	40
12	**(*E*)-2-hexenal**	30
13	(*E*)-2-hexenol	30
14	*trans*-bergamotene	30
15	δ-cadinene	30
16	α-curcumene	30

Fig. 2. A) Effect of the increasing number of terpenes selected by Genetic Algorithm (*GA*) on percentage accuracy and class error average (*CEA*) achieved by running *LDA*, and frequency with which each terpene was selected by *GA* (the variables with at least 30% frequency were reported). B) Effect of the increasing number of combined compounds (terpenes, LOX VOCs, and phenolic compounds) selected by Genetic Algorithm (*GA*) on percentage accuracy and class error average (*CEA*) achieved by running *LDA*, and frequency with which each compound was selected by *GA* (the variables with at least 30% frequency were reported). Accuracy and CEA were evaluated from classifications performed running *LDA* in leave-one-out cross-validation mode (CV-LOO).

terpenes, LOX VOCs, and phenolic compounds were also preliminarily proposed for this purpose. In order for these potential applications to become truly practical, a next validation step is required to demonstrate transferability across independent laboratories while maintaining a sample-set representative of the different Italian production areas and producers.

Finally, this study paves the way for optimization studies of the MEVOO production process aimed at enhancing (or at least preserving) the sensory differences identified among the three clusters through the selective monitoring of chemical compounds content throughout the entire MEVOO supply chain.

CRediT authorship contribution statement

Tommaso Ugolini: Writing – original draft, Validation, Methodology, Investigation, Data curation. **Lorenzo Cecchi:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nadia Mulinacci:** Supervision, Resources, Project administration. **Fabrizio Melani:** Writing – review & editing, Validation, Software, Formal analysis. **Alessandro Parenti:** Methodology, Investigation, Conceptualization. **Bruno Zanoni:** Writing – review & editing, Writing – original draft, Supervision. **Lorenzo Guerrini:** Investigation, Data curation. **Marzia Migliorini:** Methodology, Conceptualization. **Carlotta Breschi:** Investigation, Data curation.

Consent for publication

All authors have reviewed and approved the final version of the manuscript for publication.

Funding

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Declaration of Competing Interest

The authors declare that they have no known competing financial

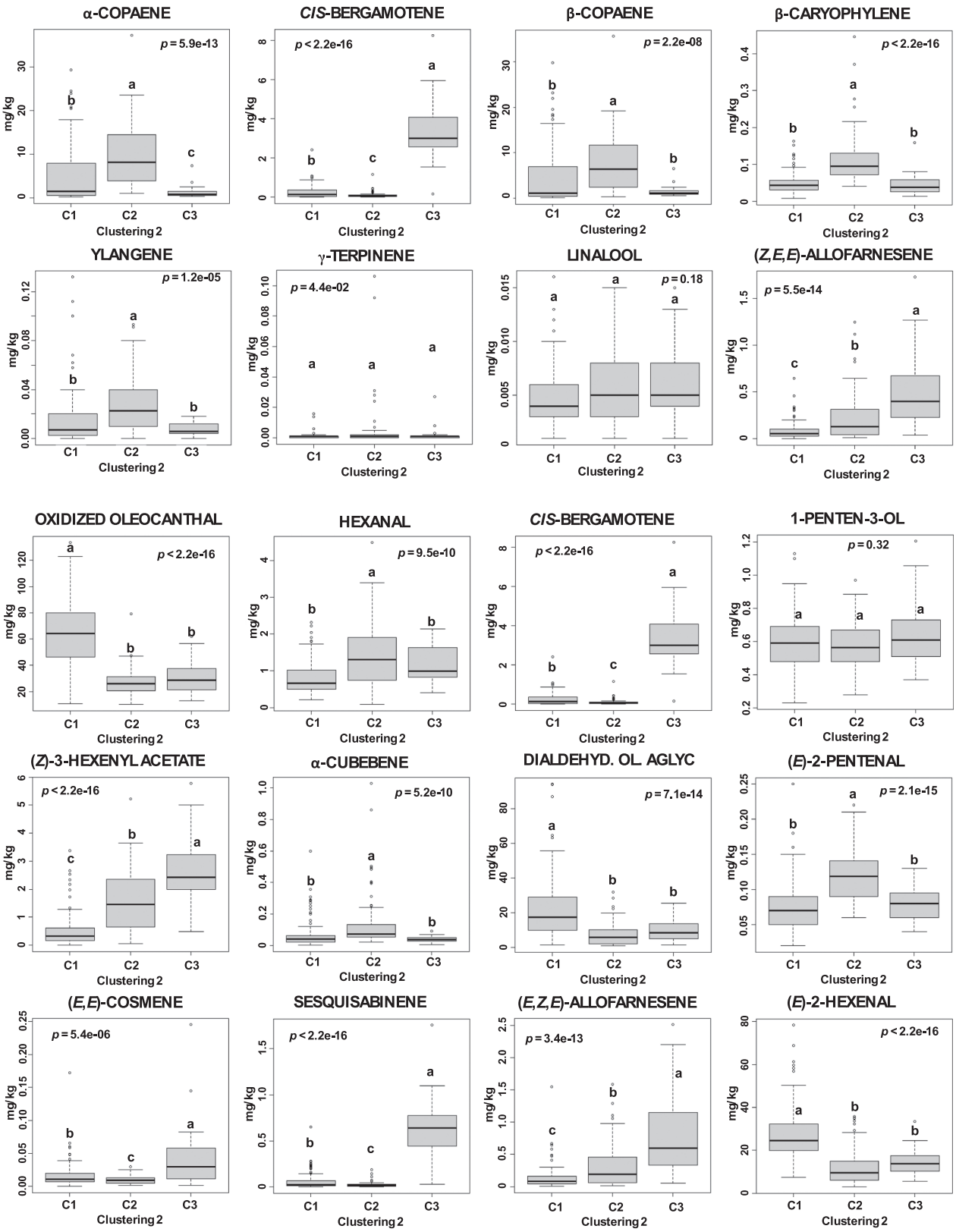


Fig. 3. Box plots and significance of the molecules related to clusters of cultivars in the clustering #2. For each molecule, different letters indicate statistically significant differences among cultivars at $p < 0.05$ according to the Dunn's test. The *p*-value from the Kruskal-Wallis test is also reported. (A), terpenes; (B), terpenes + LOX VOCs + phenolic compounds. **Clustering 2**: **C1**, Pendolino, Coratina, Leccino, Frantoio, Leccio del Corno, Maurino, Moraiolo; **C2**, Ascolana Tenera, Tonda Iblea, Itrana, Nocellara; **C3**, Bosana, Peranzana.

interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2026.109038.

Data availability

Data will be made available on request.

References

Aparicio-Ruiz, R., Morales, M.T., Aparicio, R., 2019. Does authenticity of virgin olive oil sensory quality require input from chemistry? Eur. J. Lipid Sci. Technol. 121 (12). https://doi.org/10.1002/ejlt.201900202.

Aprea, E., Gasperi, F., Betta, E., Sani, G., Cantini, C., 2018. Variability in volatile compounds from lipoxygenase pathway in extra virgin olive oils from Tuscan olive germoplasm by quantitative SPME/GC-MS. J. Mass Spectrom. 53 (9), 824–832. https://doi.org/10.1002/jms.4274.

Bajoub, A., Bendini, A., Fernández-Gutiérrez, A., Carrasco-Pancorbo, A., 2018. Olive oil authentication: A comparative analysis of regulatory frameworks with especial emphasis on quality and authenticity indices, and recent analytical techniques developed for their assessment. A review. Crit. Rev. Food Sci. Nutr. 58 (5), 832–857. https://doi.org/10.1080/10408398.2016.1225666.

Barbieri, S., Brkić Bubola, K., Bendini, A., Bučar-Miklavčič, M., Lacoste, F., Tibet, U., Winkelmann, O., García-González, D.L., Gallina Toschi, T., 2020. Alignment and Proficiency of Virgin Olive Oil Sensory Panels: The OLEUM Approach. Foods 9 (3), 355. https://doi.org/10.3390/foods9030355.

Blasi, F., Pollini, L., Cossignani, L., 2019. Varietal authentication of extra virgin olive oils by triacylglycerols and volatiles analysis. Foods 8 (2), 58. https://doi.org/10.3390/foods8020058.

Breiman, 2001. Random For. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Campestre, C., Angelini, G., Gasbarri, C., Angerosa, F., 2017. The Compounds Responsible for the Sensory Profile in Monovarietal Virgin Olive Oils. Molecules 22 (11), 1833. https://doi.org/10.3390/molecules22111833.

Cecchi, L., Innocenti, M., Melani, F., Migliorini, M., Conte, L., Mulinacci, N., 2017. New isobaric lignans from Refined Olive Oils as quality markers for Virgin Olive Oils. Food Chem. 219, 148–157. https://doi.org/10.1016/j.foodchem.2016.09.132.

Cecchi, L., Migliorini, M., Giambanelli, E., Cane, A., Zanoni, B., Canuti, V., Mulinacci, N., Melani, F., 2022. Is the volatile compounds profile a suitable tool for authentication of virgin olive oils (Olea europaea L.) according to cultivars? A study by using HS-SPME-GC-MS and chemometrics. Food Control 139, 109092. https://doi.org/10.1016/j.foodcont.2022.109092.

Cecchi, L., Migliorini, M., Giambanelli, E., Rossetti, A., Cane, A., Melani, F., Mulinacci, N., 2019. Headspace Solid-Phase Microextraction–Gas Chromatography–Mass Spectrometry Quantification of the Volatile Profile of More than 1200 Virgin Olive Oils for Supporting the Panel Test in Their Classification: Comparison of Different Chemometric Approaches. J. Agric. Food Chem. 67 (32), 9112–9120. https://doi.org/10.1021/acs.jafc.9b03346.

Cecchi, L., Migliorini, M., Mulinacci, N., 2021. Virgin Olive Oil Volatile Compounds: Composition, Sensory Characteristics, Analytical Approaches, Quality Control, and Authentication. J. Agric. Food Chem. 69 (7), 2013–2040. https://doi.org/10.1021/acs.jafc.0c07744.

Cecchi, L., Orlandini, S., Balli, D., Zanoni, B., Migliorini, M., Giambanelli, E., Catola, S., Furlanetto, S., Mulinacci, N., 2024. Analysis of Volatile Hydrocarbons (Pentene Dimers and Terpenes) in Extra Virgin Olive Oil: Optimization by Response Surface Methodology and Validation of HS-SPME-GC-MS Method. J. Agric. Food Chem. 72 (5), 2813–2825. https://doi.org/10.1021/acs.jafc.3c07430.

Cecchi, L., Parenti, A., Bellumori, M., Migliorini, M., Mulinacci, N., Guerrini, L., 2022. Clustering Monovarietal Extra Virgin Olive Oil According to Sensory Profile, Volatile Compounds, and *k*-Mean Algorithm. Eur. J. Lipid Sci. Technol. 124 (11). https://doi.org/10.1002/ejlt.202200038.

Cecchi, T., Alfei, B., 2013. Volatile profiles of Italian monovarietal extra virgin olive oils via HS-SPME–GC–MS: Newly identified compounds, flavors molecular markers, and terpenic profile. Food Chem. 141 (3), 2025–2035. https://doi.org/10.1016/j.foodchem.2013.05.090.

Clodoveo, M.L., Muraglia, M., Crupi, P., Hbaieb, R.H., De Santis, S., Desantis, A., Corbo, F., 2022. The Tower of Babel of Pharma-Food Study on Extra Virgin Olive Oil Polyphenols. Foods 11 (13), 1915. https://doi.org/10.3390/foods11131915.

COI 2018. ⟨https://www.internationaloliveoil.org/wp-content/uploads/2019/11/COI-T20-Doc.-15-REV-10-2018-Eng.pdf⟩, Accessed 23 July 2025.

COI 2022. ⟨https://www.internationaloliveoil.org/wp-content/uploads/2022/06/Doc.-No-29-REV-2_ENK.pdf⟩, Accessed 23 July 2025.

COI 2024a. ⟨https://www.internationaloliveoil.org/olive-world/olive-oil/⟩, Accessed 23 July 2025.

COI 2024b. ⟨https://www.internationaloliveoil.org/wp-content/uploads/2023/12/IOC-Olive-Oil-Dashboard.html#production-1⟩, Accessed 23 July 2025.

Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., Lacoste, F., Brereton, P., García-González, D.L., Moreda, W., Gallina Toschi, T., 2020. Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future. Trends Food Sci. & Technol. 105, 483–493. https://doi.org/10.1016/j.tifs.2019.02.025.

Dekhili, S., Sirieix, L., Cohen, E., 2011. How consumers choose olive oil: The importance of origin cues. Food Qual. Prefer. 22 (8), 757–762. https://doi.org/10.1016/j.foodqual.2011.06.005.

Del Giudice, T., Cavallo, C., Caracciolo, F., Cicia, G., 2015. What attributes of extra virgin olive oil are really important for consumers: A meta-analysis of consumers' stated preferences. Agric. Food Econ. 3 (1). https://doi.org/10.1186/s40100-015-0034-5.

Delgado, C., Guinard, J.-X., 2011. How do consumer hedonic ratings for extra virgin olive oil relate to quality ratings by experts and descriptive analysis ratings? Food Qual. Prefer. 22 (2), 213–225. https://doi.org/10.1016/j.foodqual.2010.10.004.

Dias, L.G., Fernandes, A., Veloso, A.C.A., Machado, A.A.S.C., Pereira, J.A., Peres, A.M., 2014. Single-cultivar extra virgin olive oil classification using a potentiometric electronic tongue. Food Chem. 160, 321–329. https://doi.org/10.1016/j.foodchem.2014.03.072.

EFSA, 2009. Labelling reference intake values for n-3 and n-6 polyunsaturated fatty acids. EFSA J. EFSA J. https://doi.org/10.2903/j.efsa.2009.1176.

EFSA, 2011. Scientific Opinion on the substantiation of health claims related to polyphenols in olive and protection of LDL particles from oxidative damage (ID 1333, 1638, 1639, 1696, 2865), maintenance of normal blood HDL-cholesterol concentrations (ID 1639), maintenance of normal blood pressure (ID 3781), "anti-inflammatory properties" (ID 1882), "contributes to the upper respiratory tract health" (ID 3468), "can help to maintain a normal function of gastrointestinal tract" (3779), and "contributes to body defences against external agents" (ID 3467) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J. 2011 9 (4), *2033*. https://doi.org/10.2903/j.efsa.2011.2033.

EU, 2020. Directorate General for Health and Food Safety. The EU agri-food fraud network and the administrative assistance and cooperation system: 2020 annual report. Publications Office. ⟨https://data.europa.eu/doi/10.2875/20163⟩.

EU 2022a. Commission Delegated Regulation (EU) 2022/2104 of 29 July 2022 supplementing Regulation (EU) No 1308/2013 of the European Parliament and of the Council as regards marketing standards for olive oil, and repealing Commission Regulation (EEC) No 2568/91 and Commission Implementing Regulation (EU) No 29/2012.

EU 2022b. Commission Implementing Regulation (EU) 2022/2105 of 29 July 2022 laying down rules on conformity checks of marketing standards for olive oil and methods of analysis of the characteristics of olive oil.

Forrest, S., 1993. *Genetic Algorithms*. Forrest 1993 Princ. Nat. Sel. Appl. Co. Putation 261, 872–878. https://doi.org/10.1126/science.8346439.

García-González, A., Quintero-Flórez, A., Ruiz-Méndez, M.-V., Perona, J.S., 2023. Virgin Olive Oil Ranks First in a New Nutritional Quality Score Due to Its Compositional Profile. Nutrients 15 (9), 2127. https://doi.org/10.3390/nu15092127.

Guerrini, L., Corti, F., Masella, P., Calamai, L., Angeloni, G., Spadi, A., Zanoni, B., Parenti, A., 2023. Cross-batch contamination in a continuous horizontal decanter centrifuge during virgin olive oil production. Eur. J. Lipid Sci. Technol. 125 (2). https://doi.org/10.1002/ejlt.202200174.

Guth, H., Grosch, W., 1993. Quantitation of potent odorants of virgin olive oil by stable-isotope dilution assays. J. Am. Oil Chem. ' Soc. 70 (5), 513–518. https://doi.org/10.1007/bf02542586.

ISMEA 2018. Gusto in-EVOluzione. Recuperato 23 luglio 2025, da ⟨https://www.ismeamercati.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/8572⟩.

Kalua, C.M., Allen, M.S., Bedgood, D.R., Bishop, A.G., Prenzler, P.D., 2005. Discrimination of Olive Oils and Fruits into Cultivars and Maturity Stages Based on Phenolic and Volatile Compounds. J. Agric. Food Chem. 53 (20), 8054–8062. https://doi.org/10.1021/jf051233i.

Kesen, S., Kelebek, H., Sen, K., Ulas, M., Selli, S., 2013. GC–MS–olfactometric characterization of the key aroma compounds in Turkish olive oils by application of the aroma extract dilution analysis. Food Res. Int. 54 (2), 1987–1994. https://doi.org/10.1016/j.foodres.2013.09.005.

Lukić, I., Lukić, M., Žanetić, M., Krapac, M., Godena, S., Brkić Bubola, K., 2019. Inter-varietal diversity of typical volatile and phenolic profiles of croatian extra virgin olive oils as revealed by GC-IT-MS and UPLC-DAD analysis. Foods 8 (11), 565. https://doi.org/10.3390/foods8110565.

Montealegre, C., Marina Alegre, M.L., García-Ruiz, C., 2010. Traceability markers to the botanical origin in olive oils. J. Agric. Food Chem. 58 (1), 28–38. https://doi.org/10.1021/jf902619z.

Morikawa, M., 2021. Price competition vs. quality competition: Evidence from firm surveys. J. Econ. Bus. 116, 106007. https://doi.org/10.1016/j.jeconbus.2021.106007.

Ocakoglu, D., Tokatli, F., Ozen, B., Korel, F., 2009. Distribution of simple phenols, phenolic acids and flavonoids in Turkish monovarietal extra virgin olive oils for two harvest years. Food Chem. 113 (2), 401–410. https://doi.org/10.1016/j.foodchem.2008.07.057.

Olive Oil Times 2023. *How Monovarietal Olive Oils Promote Ecological Farming, Safeguard Landscapes.* ⟨https://www.oliveoiltimes.com/world/how-monovarietal-olive-oils-promote-ecological-farming-safeguard-landscapes/124357⟩.

Peres, F., Martins, L.L., Ferreira-Dias, S., 2017. Influence of enzymes and technology on virgin olive oil composition. Crit. Rev. Food Sci. Nutr. 57 (14), 3104–3126. https://doi.org/10.1080/10408398.2015.1092107.

Peri, 2014. The extra-virgin olive oil chain. (2014). In: Peri, C. (Ed.), The Extra-Virgin Olive Oil Handbook. Wiley, pp. 3–10. https://doi.org/10.1002/9781118460412.ch1.

Piccolo, D., Capecchi, S., Iannario, M., Corduas, M., 2013. Modelling consumer preferences for extra virgin olive oil: the italian case. Int. Agric. Policy 2013 (1).

Rébufa, C., Pinatel, C., Artaud, J., Girard, F., 2021. A comparative study of the main international extra virgin olive oil competitions: Their impact on producers and consumers. Trends Food Sci. & Technol. 107, 445–454. https://doi.org/10.1016/j.tifs.2020.11.014.

Stasi, A., Diotallevi, F., Marchini, A., Nardone, G., 2018. Italian Extra-Virgin Olive Oil: Impact on Demand on Being Market Leaders, Private Labels or Small Producers. Rev. Econ. 13 (3).

Torres-Cobos, B., Quintanilla-Casas, B., Romero, A., Ninot, A., Alonso-Salces, R.M., Toschi, T.G., Bendini, A., Guardiola, F., Tres, A., Vichi, S., 2021. Varietal authentication of virgin olive oil: proving the efficiency of sesquiterpene fingerprinting for Mediterranean Arbequina oils. Food Control 128, 108200. https://doi.org/10.1016/j.foodcont.2021.108200.

Ugolini, T., Mattagli, F., Melani, F., Zanoni, B., Migliorini, M., Trapani, S., Giambanelli, E., Parenti, A., Mulinacci, N., Cecchi, L., 2024. HS-SPME-GC-MS and chemometrics for the quality control and clustering of monovarietal extra virgin olive oil: a 3-year study on terpenes and pentene dimers of Italian cultivars. J. Agric. Food Chem. 72 (19), 11124–11139. https://doi.org/10.1021/acs.jafc.4c00610.

Vichi, S., Guadayol, J.M., Caixach, J., López-Tamames, E., Buxaderas, S., 2006. Monoterpene and sesquiterpene hydrocarbons of virgin olive oil by headspace solid-phase microextraction coupled to gas chromatography/mass spectrometry. J. Chromatogr. A 1125 (1), 117–123. https://doi.org/10.1016/j.chroma.2006.05.029.

Žanetić, M., Jukić Špika, M., Ožić, M.M., Brkić Bubola, K., 2021. Comparative study of volatile compounds and sensory characteristics of dalmatian monovarietal virgin olive oils. Plants 10 (10), 1995. https://doi.org/10.3390/plants10101995.