

## Article

# Prediction of Fruity Aroma Intensity and Defect Presence in Virgin Olive Oil Using an Electronic Nose

Pablo Cano Marchal <sup>1,\*</sup>, Chiara Sanmartin <sup>2</sup>, Silvia Satorres Martínez <sup>1</sup>, Juan Gómez Ortega <sup>1</sup>,  
Fabio Mencarelli <sup>2</sup> and Javier Gámez García <sup>1</sup>

<sup>1</sup> Robotics, Automation and Computer Vision Group, University of Jaén, 23071 Jaén, Spain; satorres@ujaen.es (S.S.M.); juango@ujaen.es (J.G.O.); jggarcia@ujaen.es (J.G.G.)

<sup>2</sup> Department of Agriculture, Food and Environment, University of Pisa, 56126 Pisa, Italy; chiara.sanmartin@unipi.it (C.S.); mencarelli@unipi.it (F.M.)

\* Correspondence: pcano@ujaen.es

**Abstract:** The organoleptic profile of a Virgin Olive Oil is a key quality parameter that is currently obtained by human sensory panels. The development of an instrumental technique capable of providing information about this profile quickly and online is of great interest. This work employed a general purpose e-nose, in lab conditions, to predict the level of fruity aroma and the presence of defects in Virgin Olive Oils. The raw data provided by the e-nose were used to extract a set of features that fed a regressor to predict the level of fruity aroma and a classifier to detect the presence of defects. The results obtained were a mean validation error of 0.5 units for the prediction of fruity aroma using lasso regression; and 88% accuracy for the defect detection using logistic regression. Finally, the identification of two out of ten specific sensors of the e-nose that can provide successful results paves the way to the design of low-cost specific electronic noses for this application.

**Keywords:** virgin olive oil; quality; electronic nose



**Citation:** Cano Marchal, P.; Sanmartin, C.; Satorres Martínez, S.; Gómez Ortega, J.; Mencarelli, F.; Gámez García, J. Prediction of Fruity Aroma Intensity and Defect Presence in Virgin Olive Oil Using an Electronic Nose. *Sensors* **2021**, *21*, 2298. <https://doi.org/10.3390/s21072298>

Academic Editor: Tomislav Jemrić

Received: 2 March 2021

Accepted: 19 March 2021

Published: 25 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A common feature of the Mediterranean diet, which is a dietary pattern associated with longevity, improvement of life quality, and a reduction of the incidence of many chronic pathologies (i.e., cardiovascular diseases, cerebrovascular diseases, diabetes mellitus, metabolic syndrome, certain cancers, and neurodegenerative diseases), is the high consumption of olives and olive oil as the primary source of dietary fat [1–3].

Olive oil quality is deeply conditioned by extraction techniques, thereafter virgin olive oils are designated also as “oils obtained from the fruit of the olive tree solely by mechanical or other physical means under conditions, particularly thermal conditions that do not lead to alterations in the oil, and which have not undergone any treatment other than washing, decantation, centrifugation and filtration”. These oils are ready for direct human consumption (EC Reg ECC 2568/1991 as amended). European legislation provides three main merceological classes of olive oil, based on physical, chemical, and sensory characteristics: extra virgin (EVOO), virgin (VOO), and lampante olive oil (LOO). The olive oil features are strictly regulated by EU, which also establishes methods for their analysis (Reg (EEC) 2568/91, EC Reg 640/08) [4]. When EVOO is extracted from fresh and healthy olive fruits, it is properly processed and adequately stored [5,6]; it exhibits an organoleptical profile characterized by a unique fruity aroma and taste, with no defects (e.g., fusty, musty, winery, metallic, rancid).

Its volatile constituents are mainly C6 compounds (aldehydes, ketones, alcohols, and esters in particular), synthesized by means of the lipoxygenase metabolic pathway and responsible for EVOO green scents [7]. EVOO's fragrant and delicate flavor is ascribable to fruity sensation, which is a reminiscence of the healthy olive fruit perfectly ripe, and to

scents of cut grass, olive leaf, flowery scents, green fruits (e.g., apple, banana, almond) or vegetables (e.g., tomato, artichoke) [8,9].

Indeed, its peculiar taste is characterized by more or less intense bitterness and pungency notes, which are positive attributes depending on the content of bioactive compounds strictly related to important health benefits, such as phenols [10]. The EVOO sensory profile is strongly related to its quality, so there are other lower-quality olive oils, such as VOO, which is also fruity but has a few minor defects deriving from incorrect extraction or storage processes, and LOO, which exhibits an unpleasant smell and is non-edible [11].

According to the European legislation, the sensory analysis of virgin olive oils must be carried out by both smell and taste assessments by means of a panel test, constituted by a group of trained and selected assessors (EEC 2568/91; EC 640/08). Thus far, there is no methodology able to suitably replace human sensory panels, essentially because taste and olfactory properties are formed by a plethora of different compounds in varying concentrations, which deliver the holistic sensory impression [12–14]. Sensory assessment, however, needs many resources and time, as well as specialized panellists, which are not always at the disposal of small/medium-sized enterprises and cooperative societies, and should not be used for routine operations [13,15].

In this context, there is a need for the development of accurate instrumental techniques capable of performing measurements in real-time and generating the same information as a panel, in a reproducible and stable way, in order to achieve the correct classification of VOOs rapidly and efficiently [16]. Several analytical methods have been proposed in different studies carried out in the last decade, many including characterizations of the volatile markers for each sensory characteristics [17], and there have been recent efforts devoted to the quality testing of different types of oils using electronic noses [18–20]. The development of easy-to-use electronic noses, capable of being installed directly in the process line, appears extremely appealing in this field.

The aim of this work is to employ the electronic nose to develop an automatic online analysis system for olive oils' batches sensory classification immediately after the olive oil extraction process. The acquired knowledge could be applied to evaluate the fruity aroma and defect level of the olive oil batches at the end of the virgin olive oil production process and classify them, in order to avoid misclassification and to prevent the mixing low quality oil in the same tank with high quality oil. In this work, we analyze all the VOO produced by a company in a season in lab conditions, and assess its quality by means of predicting the value of the fruity aroma attribute and the existence of defects.

The rest of the paper is organized as follows: the next section provides a brief overview of relevant previous works, while Section 3 details the Materials and Methods employed to obtain the results discussed in Section 4. The paper ends with the conclusions presented in Section 5.

## 2. Related Works

The sensors usually denominated electronic noses are devices that employ one or more gas transducers that receive a flow of air extracted from the headspace of the samples and modify their electric impedance in the presence of different families of volatile molecular compounds, so that they provide an electric signal that depends on the presence or absence of these compounds [21].

The major applications in food products for this type of sensors are quality control and food safety [22,23], authenticity control [24], and detection of contaminants [25]. This way, in the literature, there are several reports that devote their attention to the application of e-noses to fruits and vegetables in order to identify and select cultivars, to discriminate among different ripening stages, to predict the optimal harvest date, to monitor shelf life [26] or to manage the post-harvest processes and to evaluate the occurrence of defects [16]. This technology has also been employed to evaluate the quality of fish [27,28], meat [29] and processed vegetables [25,30–32].

In addition, according to [23], e-noses find useful applications in several issues pertaining to classification, determination of botanical and geographical origin, detection of adulterations and evaluation of the degree of oxidation of edible oils [33]. For olive oil in particular, this technology has been proven useful for determining the geographical origin of the oil, detecting adulterations and quality determining the spoilage due to external factors, such as temperature, storage time, air composition in storage tanks, etc. [23].

Other recent studies of the application of this technology to olive oil [34,35] show the different responses provided by the sensors when confronted with different olive oil samples adulterated with other vegetal oils. Ref. [36] show the possibility of classifying virgin olive oil versus other chemically treated olive oils using only a single sensor, with a success rate of 91%. Other works, such as [37], show the use of e-nose to detect the quality loss over time of virgin olive oils stored at different temperatures, with a success rate close to 94%. In turn, ref. [38] used an e-nose to classify olive oils obtained from different cultivars, yielding a 82% accuracy in their predictions. Finally, ref. [39] demonstrated the potential of using this type of technology directly on intact olives to predict whether the oil to be produced from them would be virgin or lampante, with an accuracy of 90%.

The novelty of the work presented in this paper resides on the prediction of the intensity of fruity aroma and the detection of defect, which can be employed to classify olive oils prior to their storage.

### 3. Materials and Methods

#### 3.1. Olive Oil Samples

Forty-two olive oil samples, produced from *Olea europaea* fruits belonging to typical Spanish cultivars (mainly cv. Picual) and harvested during the 2018/2019 crop season from different growers of Jaén province (Spain), were provided by a local Olive Oil Cooperative (Picualia mill, Bailen—Jaén).

Each sample was divided in representative aliquots and stored in dark glass 125 mL containers with the aim to immediately evaluate its volatile emission by electronic-nose and to be delivered to the accredited CM Europe laboratory for the sensory and chemical analysis.

#### 3.2. Chemical and Sensory Analysis of Oil

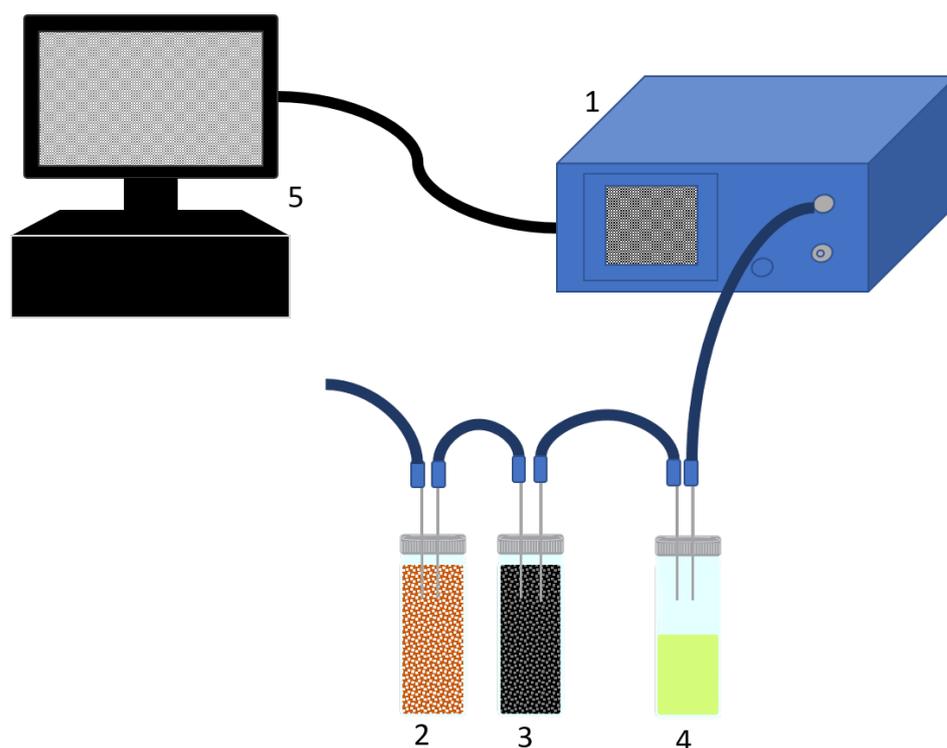
Each chemical determination was performed in triplicate, following the official methods—acidity index (AI), peroxide value (PV), K232, K270, and  $\Delta K$  according to EU Reg 2568/1991. The sensory characterization followed the method described by [40]. Table 1 includes the main chemical and sensory parameters of the considered samples.

**Table 1.** Chemical and sensory characterization of the samples (overall mean and standard deviation), and relevant quality degree, in accordance with current legislation.

Number of Samples	AI (g Oleic Acid/100 g Oil)	PV (meq O <sub>2</sub> /kg Oil)	K232	K270	$\Delta K$	EE (mg/kg)	Fruity	Defect	Quality Grade
21	0.16 (0.03)	4.4 (0.6)	1.50 (0.10)	0.15 (0.01)	<0.01	7 (5)	4.1 (0.8)	0 (0)	EVOO
19	0.24 (0.05)	5.1 (0.8)	1.50 (0.10)	0.14 (0.01)	<0.01	8 (3)	2.4 (0.4)	2.6 (0.3)	VOO
2	0.25 (0.01)	4.3 (0.4)	1.57 (0.01)	0.17 (0.01)	<0.01	7 (1)	1.3 (1.7)	3.6 (0)	LOO

#### 3.3. Volatile Emission by Electronic-Nose

The analysis of the headspaces of each olive oil sample was performed using an electronic nose apparatus PEN3 (Airsense Analytics GmbH, Schwerin, Germany), consisting of a gas sampling unit (maximum flow rate of 600 mL/min), an integrated sensor array composed of 10 different thermo-regulated (200–500 °C) metal oxide thick film sensors (MOS) sensitive to different classes of chemical compounds and a software (Win Muster v. 1.6.2) [41]. Figure 1 depicts a schematic of the setup used.



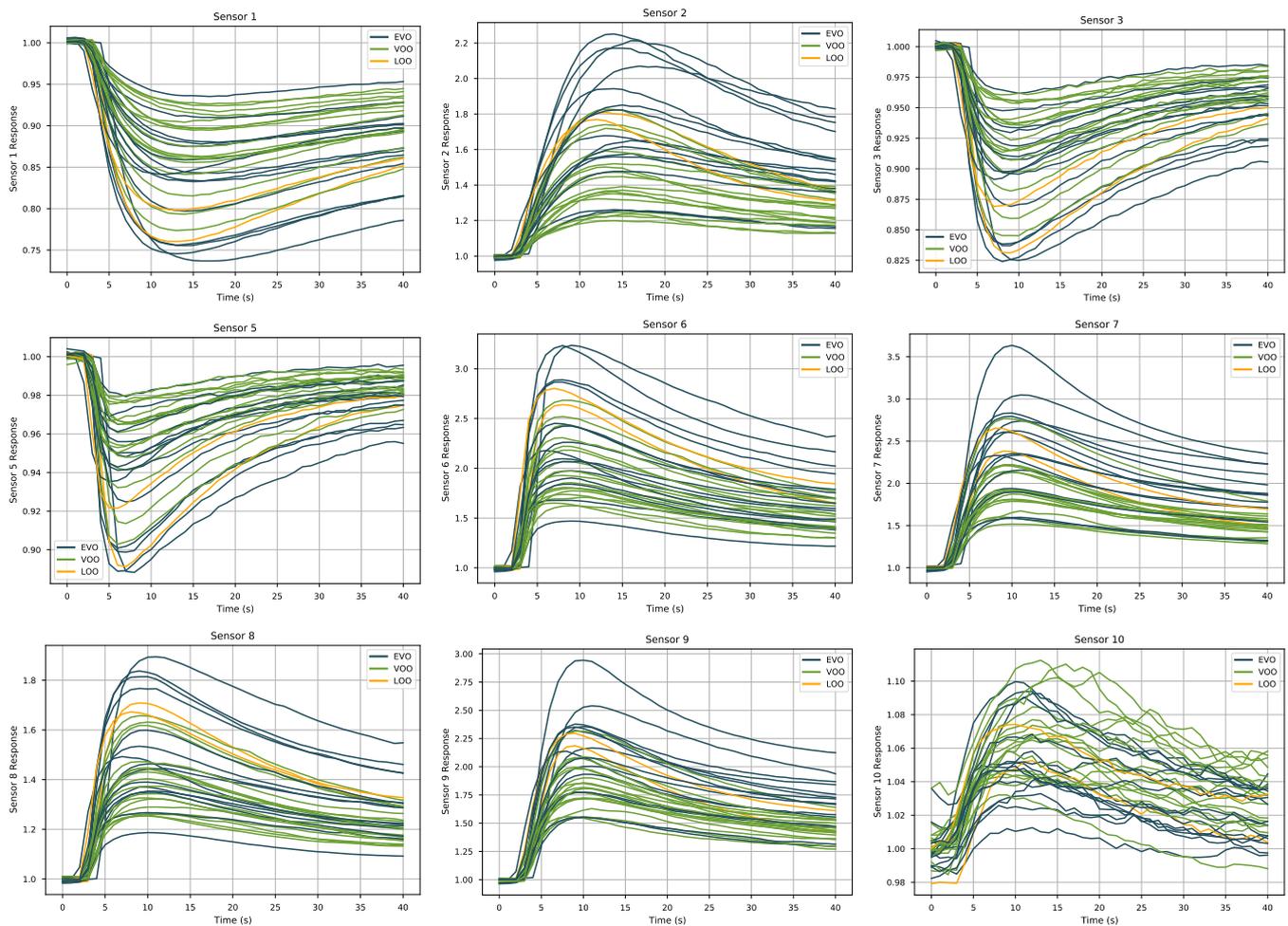
**Figure 1.** Schematic representation of the experimental setup of the electronic nose used in sample measurements (1 = PEN3 e-nose; 2 = silica gel filter; 3 = activated carbon filter; 4 = tested sample; 5 = computer).

For each sample, 5 g of olive oil were poured in a dark glass 13.5 mL vial, hermetically sealed with rubber closure and conditioned to  $30 \pm 1$  °C before the measurement process in order to increase the volatile concentration in the headspace.

As previously described by [41], each measurement process starts with a sensor array cleaning stage, when air, after passing through an activated carbon filter, reaches the sensor array. During the second step, air crosses another time through an activated carbon filter and then through the sample; after passing through a moisture and particle filter, air finally reaches the array of sensors. When the sample volatile compounds react with the sensing film of the sensor, an oxygen exchange occurs resulting in a change of electrical conductivity, detectable by a transducer element (electrode) attached to each sensor.

### 3.4. Data Analysis

The primary source of information about the samples are the response curves for each of the sensors in the PEN3 device. These curves represent the sensor measurement as a function of time. Figure 2 shows the raw response of the nine sensors finally used for the data analysis as provided by the PEN3 software for all the samples. These plots show that the noise level of the signals was very modest, so no filtering preprocessing step was required. The exception to this pattern was sensor 4 (not shown in Figure 2), which provided data that were confined in a very narrow range and with a very poor signal-to-noise ratio, so this sensor was not considered in the following data analysis.



**Figure 2.** Curves of the responses of the nine sensors finally employed in the data analysis for the different samples.

The first step of the data analysis consisted of the extraction of features that summarized the information contained in the curves, as is common practice in applications that employ electronic noses [42]. The reason for this step is that the value at each time step is highly correlated with its neighbors, so directly using the data provided by the sensors at each time as an independent feature would grant an extremely high-dimensional feature vector with high correlations between their components, which is not desirable.

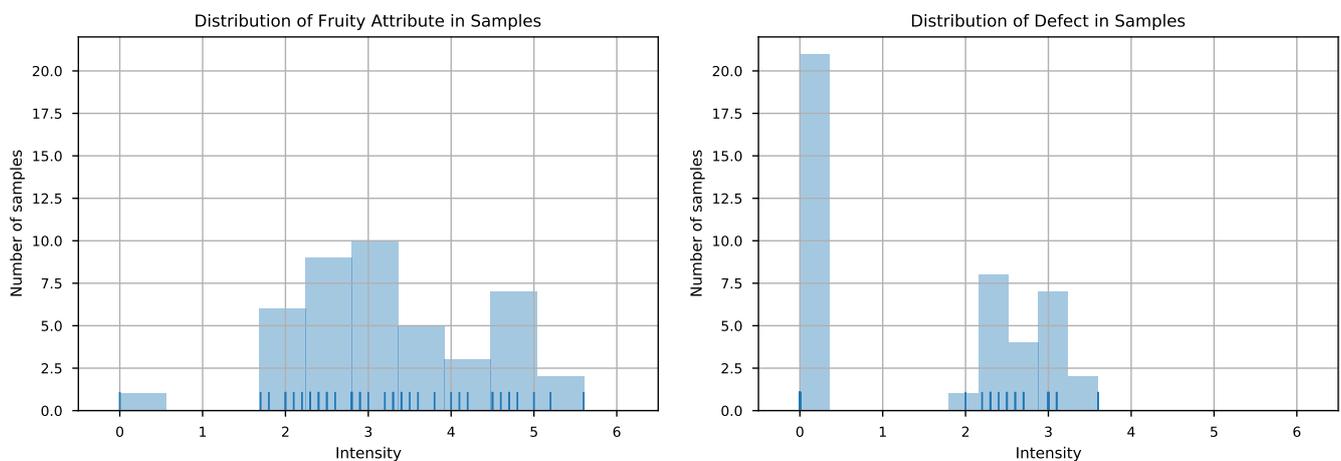
A visual inspection of the sensor response plots, along with a review of features commonly used in the literature for the characterization of this type of curves [42], provided the basis the selection of the preliminary set of features. This way, two features were extracted from the time series:

1. Peak value of the curve (SxM),
2. Final value of the curve (SxF).

This way, as an example, the peak value of the curve for sensor 1 is denoted S1M, while the final value of the curve for sensor 2 is denoted S2F. The two features extracted from the time response of each of the nine sensors finally used provide a feature vector of dimension 18, which is still relatively high when compared to the total number of samples available for the analysis (42). In order to attenuate the risk of overfitting, further regularization and dimensionality reduction techniques were applied, as detailed below. All of the computations were carried out using scikit-learn [43], a machine learning library for Python.

### 3.5. Prediction of Fruity Aroma

As the value of fruity aroma is a continuous parameter, its prediction constitutes a data regression problem. The left-hand panel of Figure 3 shows the distribution of the fruity aroma values of the samples. As depicted in the plot, the values are reasonably spread, so they should not be problematic for standard regression tools. However, as commented above, the feature vector dimension was high related to the number of samples, so it was necessary to consider candidate techniques that included regularization parameters so that overfitting did not impair the generalization of the model. Furthermore, the fact that the features essentially represent the response of sensors to the concentration of volatiles makes it plausible to consider that a linear model could be appropriate for the prediction of fruity aroma. Thus, taking these considerations into account, three candidate techniques were implemented: ridge, lasso, and elastic-net regressions.



**Figure 3.** Distribution of fruity aroma and defect intensity for the studied samples.

The three methods are very similar in nature, as they all essentially fit a linear model solving an optimization problem that includes some penalty in the objective function related to the size of the coefficients in the linear model, thus including regularization [44].

The ridge regression model finds the parameters  $w$  of the model minimizing the following expression:

$$J(w) = \|y - w^T x\|^2 + C\|w\|^2.$$

Here,  $C$  is the coefficient that determines the emphasis put on the regularization term. If  $C = 0$ , then we would have a regular least squares fitting problem.

In turn, the lasso minimizes the following expression:

$$J(w) = \|y - w^T x\|^2 + C\|w\|_1.$$

In this case, it is the  $l_1$  norm that penalizes the size of the parameters, thus yielding solutions that tend to be sparse, i.e., have many coefficients being exactly zero.

Finally, the elastic net is a mixture of both methods, minimizing

$$J(w) = \|y - w^T x\|^2 + C_1\|w\|_1 + C_2\|w\|^2,$$

with  $C_1$  and  $C_2$  being the parameters that determine the relative weight of the  $l_1$  and  $l_2$  norms, respectively.

### 3.5.1. Prediction of Defect

The prediction of defect required a different approach to the one used in the prediction of fruity aroma intensity. The distribution of the defect intensity values in the samples is shown in the right-hand panel of Figure 3. As depicted in the plot, many values are exactly zero—those corresponding to EVOO samples—and the rest of values are lumped between roughly 2 and 3.6. This data distribution is problematic for a regression approach due to the fact that many samples are exactly zero and the rest are concentrated in a fairly narrow range. For situations like this, it is usually best practice to start with a classification problem to determine whether the point belongs to the cluster at 0 or the other one, and then use a regression model trained exclusively with the data not being zero. However, the fact that the values are confined in a very narrow range, together with the limited number of samples available in this range, hindered the interest in training a regressor to predict these values. This way, only the classification step was pursued.

The first step in the data analysis was to perform Principal Component Analysis (PCA) to the set of 18 preliminary features, with the objective of visualizing the distribution of the classes in the feature space and obtaining an intuitive grasp of the difficulty of the task. Figure 4 shows the first four components of the PCA-transformed feature set, computed using the whole dataset available.

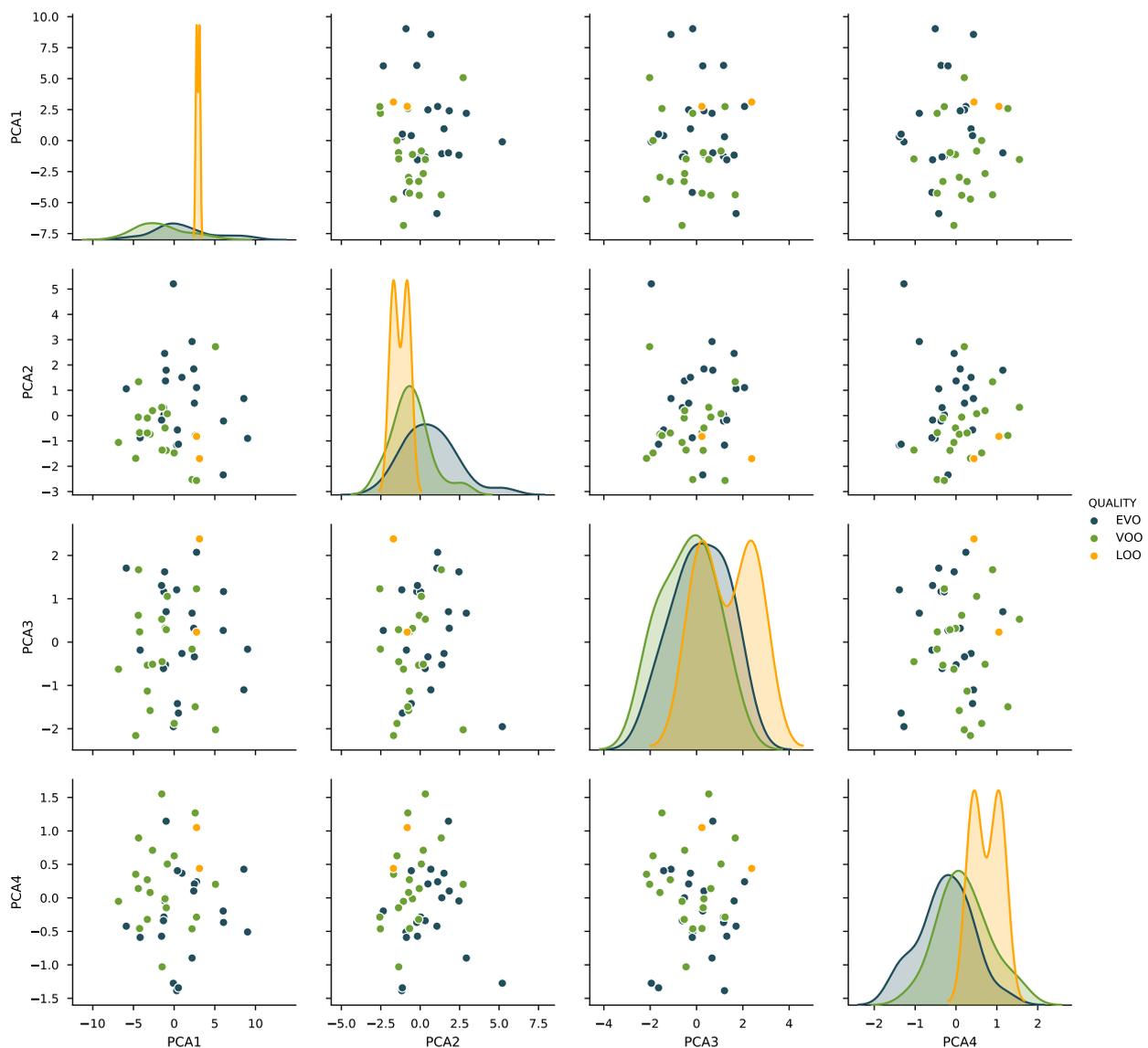


Figure 4. Visualization of the four first PCA components, computed for the whole dataset.

The inspection of the sample distribution shows that the data are effectively somewhat clustered according to its quality—which is in fact directly linked with the presence of defect—and that using a linear classifier could offer satisfactory results with good generalization properties.

This way, Support Vector Machines (SVM) were the first classifier class selected for the task. In the linearly separable case, SVM defines the separating hyperplane that maximizes the distance between the classes; for the nonlinearly separable case, as is ours, they provide an hyperplane that separates the classes including a penalty for the misclassified samples [44]. Thus, the optimization problem to be solved is:

$$\begin{aligned} & \underset{w, b, \xi}{\text{minimize}} && \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ & \text{subject to} && y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & && \xi_i \geq 0, i = 1, \dots, n \end{aligned}$$

Here,  $C$  is the parameter that defines the penalization of the misclassification of each sample. SVM allows for using different values of  $C$  for each sample, which is useful when facing an unbalanced problem, i.e., a problem where samples from one class are significantly more abundant than the other. However, we did not need to use this possibility, as roughly half of the samples are EVOO and half are not, so the value of  $C$  was the same for all the samples and chosen via cross-validation from a candidate set.

Another classification algorithm known to perform well for linear problems is Logistic Regression [44]. For this classifier, the optimization problem to be solved is:

$$\underset{w}{\text{minimize}} \quad \sum_{i=1}^n \log(1 + e^{-y_i w^T x_i}) + C \|w\|^2$$

Here,  $C$  is a regularization parameter that penalizes the size of the weight vector  $w$ . In this algorithm, unlike SVM, there is no explicit coefficient penalizing the misclassification of points, as it is implicit in the loss function in the term  $e^{-y_i w^T x_i}$ , which yields a large positive value whenever  $w^T x_i$  and  $y_i$  do not have the same sign.

Logistic Regression can also be performed using  $l_1$  regularization

$$\underset{w}{\text{minimize}} \quad \sum_{i=1}^n \log(1 + e^{-y_i w^T x_i}) + C \|w\|_1$$

This formulation, analogous to the lasso regression, tends to provide a sparse  $w$ , which can be used to provide models using only a subset of the sensors. This approach was also tested and the results are included in the next section.

## 4. Results and Discussion

This section presents and comments on the results obtained for the prediction of fruity aroma intensity and the detection of defects.

### 4.1. Prediction of Fruity Aroma

According to the results of the sensory analysis, the fruity attribute was perceived in all the samples tested but one. Both for the prediction of fruity aroma intensity and the detection of defects, five-fold cross validation was used to test the generalization capacity of the models fitted. Within each of these first five sets, which were randomly generated from the complete set of data, four-fold cross validation was applied with the training samples to select the best parameters for the models fitted.

As commented in the previous section, the initial approach was to fit ridge regression, lasso, and elastic net models to the 18 features presented in the previous section. Figure 5 shows the fitted parameters for each of models and the five data folds. As seen in the plot, there is a strong coherence in the values picked for all the models, with the lasso

model showing the least fluctuation in the parameters found. In the lasso panel, two features clearly stand out, namely *S2F* and *S10M*. A second model using only those as input features was fitted, and the results obtained are included in Figure 6 and Table 2, together with the results of the other models.

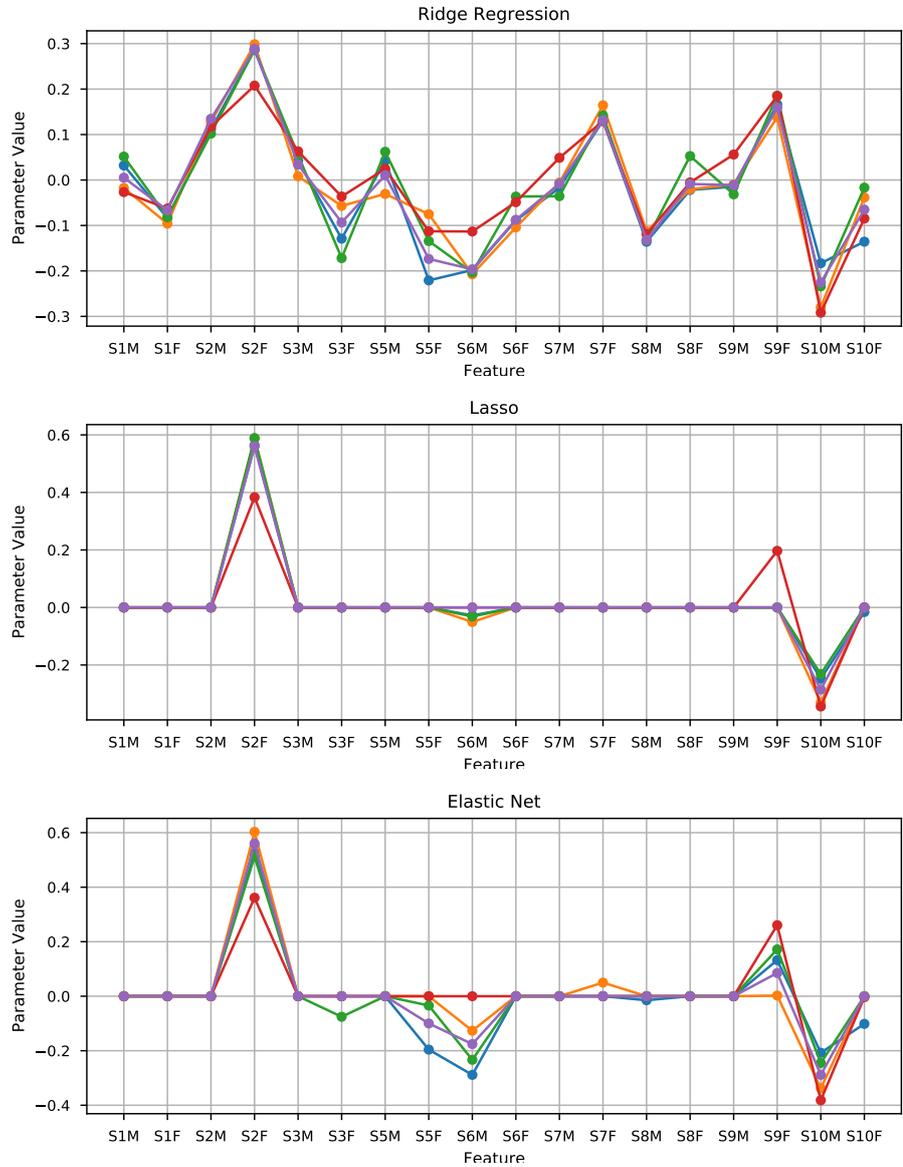
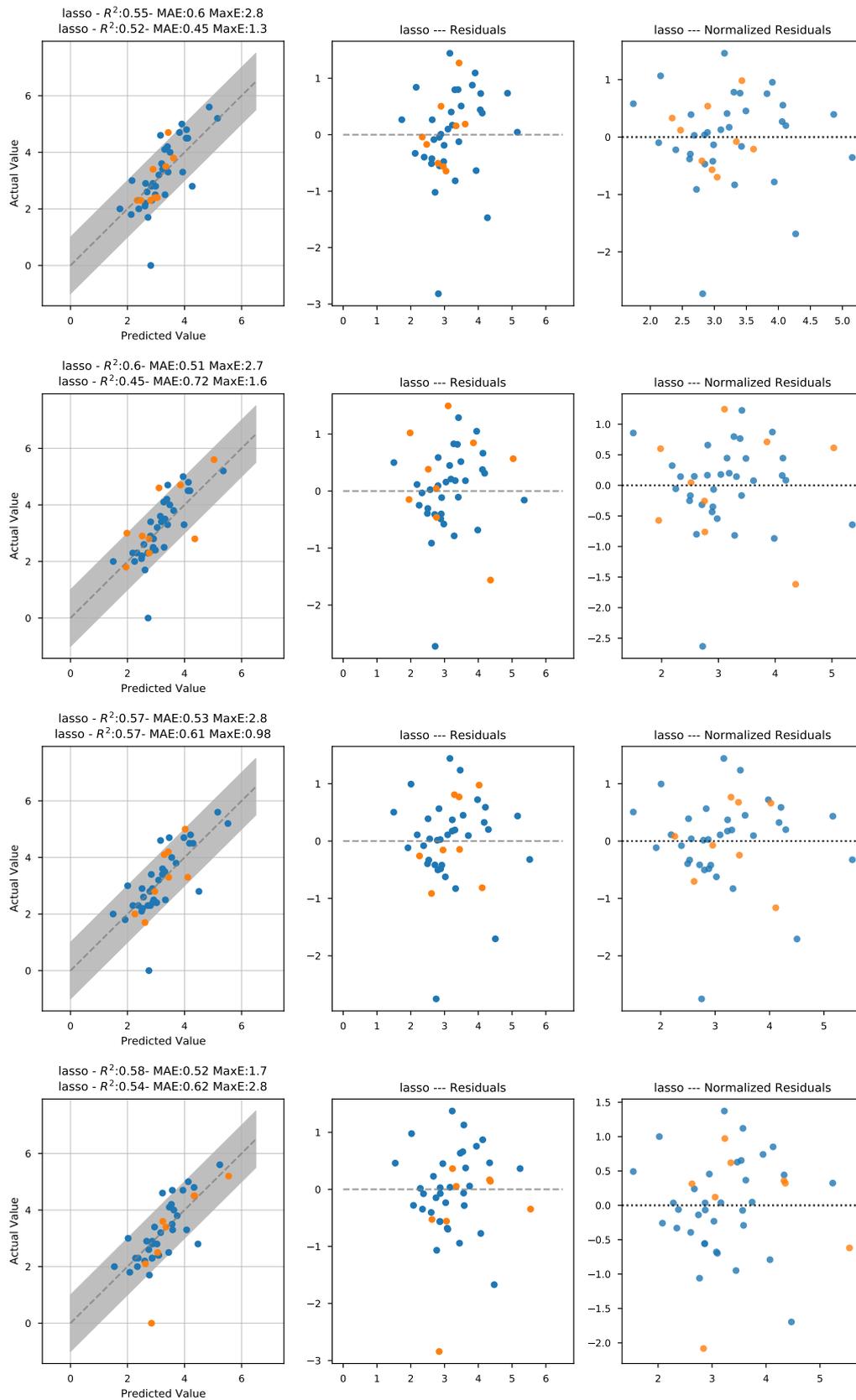


Figure 5. Visualization of the parameters of the lasso classifier for each of the input features.



**Figure 6.** Visualization of the training and testing prediction results for the first four of the five iterations for the final models constructed using the S2F and S10M features. The gray band represents the zone of  $\pm 1$ . Training points are blue and testing points are orange.

**Table 2.** Results obtained. for the prediction of fruity aroma. The final model structure selected was the fourth row (Lasso S2F, S10M).

Model	Mean Average Error (Training)	Mean Average Error (Testing)	Mean Max Error (Training)	Mean Max Error (Testing)
Ridge	1.09	1.06	5.12	3.4
ElasticNet	0.54	0.59	2.31	1.58
Lasso	0.56	0.59	2.49	1.57
Lasso (S2F, S10M)	0.55	0.55	2.55	1.49
Lasso (S2F, S10M) (VOO, EVO)	0.49	0.51	1.52	1.18

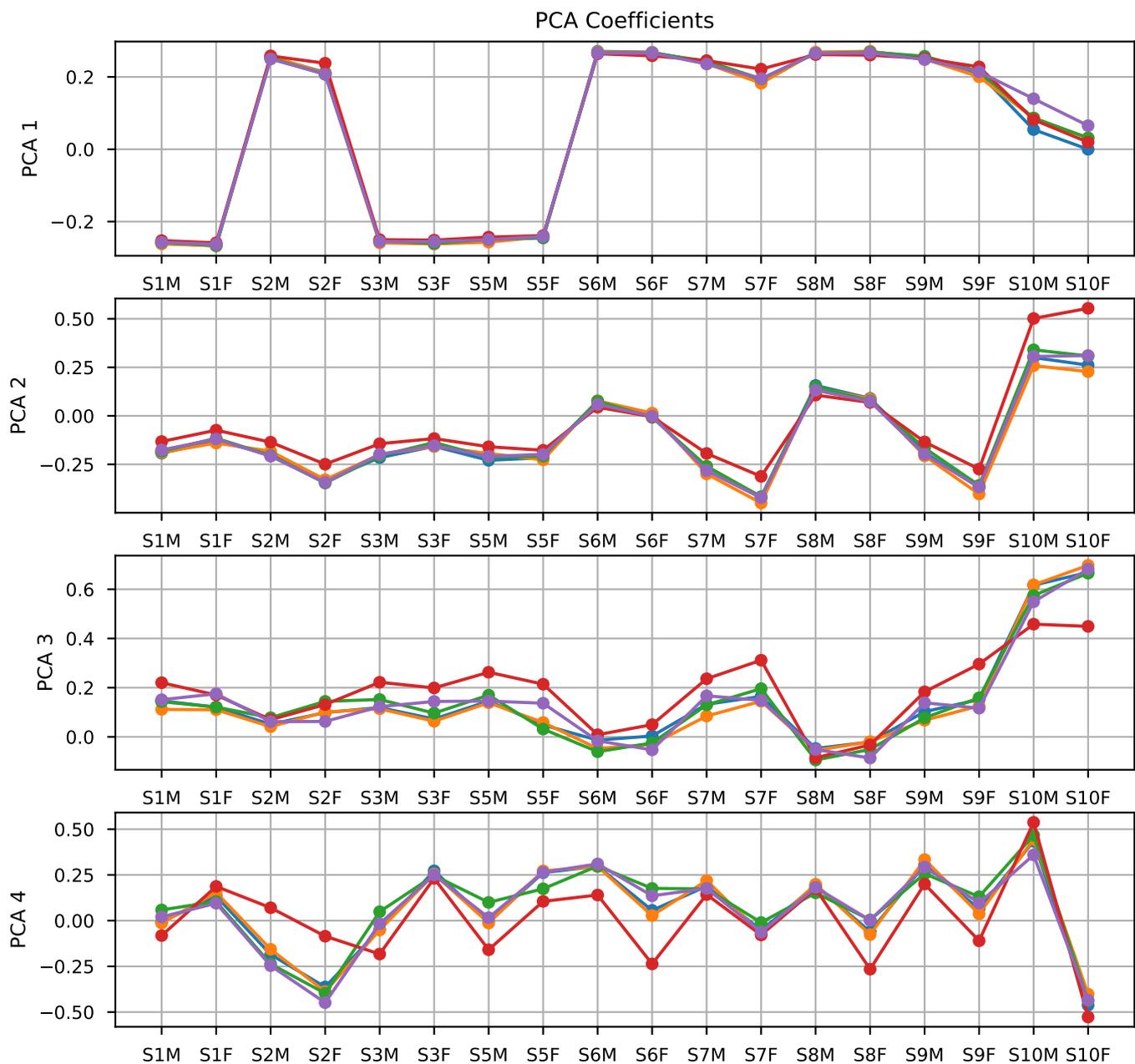
The results obtained are fairly similar for all the models and quite satisfactory: with the final lasso model yielding a mean average error of less than 0.6 and a mean maximum error around 1.5 in the testing subsets. One final test was carried out fitting a model removing the two lampante olive oil (LOO) samples in the dataset, as an inspection of Figure 6 shows a very clear outlier corresponding to the LOO sample with 0 fruity aroma intensity. The results obtained are included in the last column of Table 2. As included in the table, the testing mean average error did not improve, but the maximum error did decrease. The most likely explanation for this point behaving as an outlier is that, in panel test analysis, it is common practice not to actually taste the olive oil if the perceived level of defect is large enough for the sample to be labeled as lampante. Thus, sometimes, there is some level of fruity aroma in the sample that might not be included in the reference results.

#### 4.2. Detection of Defects

As commented in Section 3.2, the visualization of the first four components of the PCA transformation of the dataset suggested that linear classification algorithms using these transformed features should provide good results. This PCA decomposition, however, was not used for the fitting of the models.

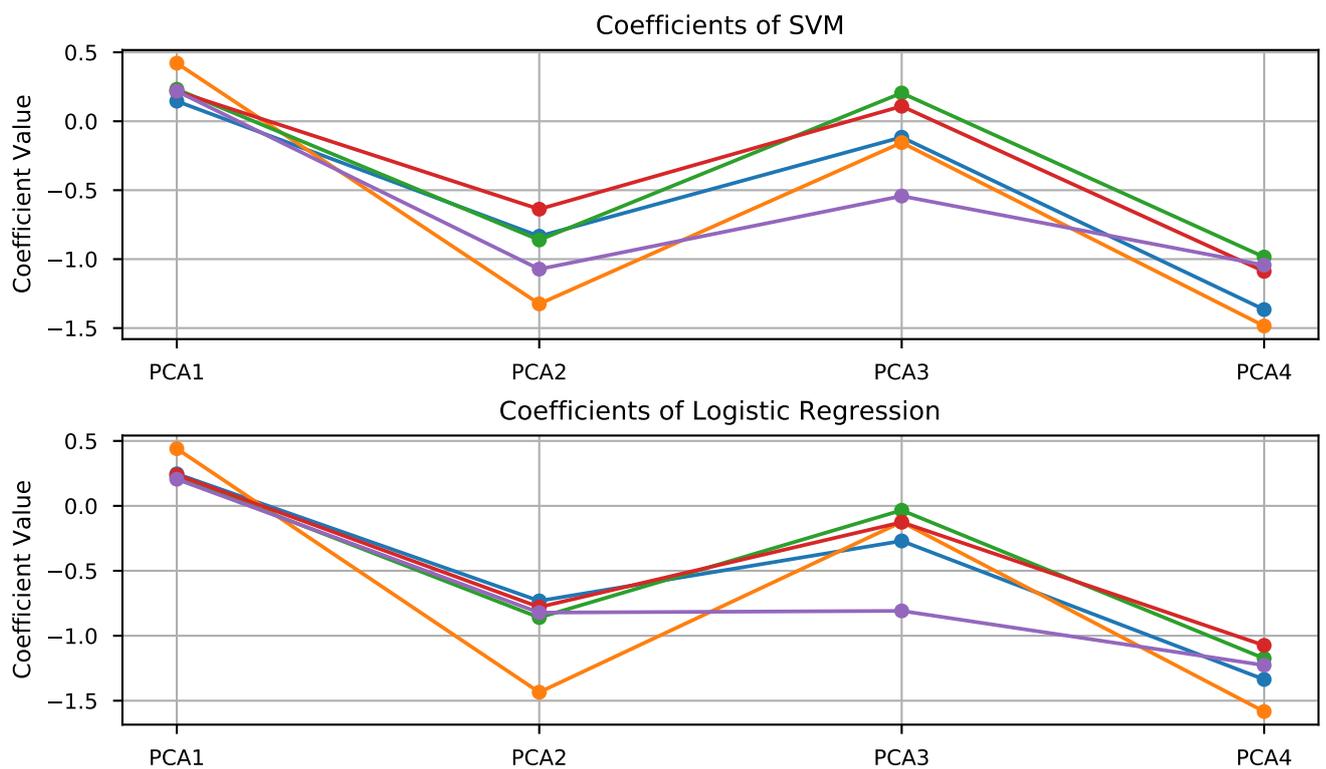
This way, the first approach was to train an SVM and a logistic regression using PCA elements in a five-fold cross-validation setup. Figure 7 shows the coefficients of the PCA transformation of each of the raw features for each fold of the cross-validation procedure. As depicted in the plot, the components provided by each subset are very similar.

A closer inspection of the discriminative component 2 (second row of Figure 7) shows that sensor 10 obtains the largest coefficient, followed by sensors 9 and 7. PCA component 4 (fourth row of Figure 7) again shows a large weight assigned to sensor 10, this time providing basically the difference between the two features associated with the sensor, as the coefficients are roughly the same magnitude and opposite sign. In this PCA component, sensor 2 also has a relatively large weight, particularly the final value feature (S2F). It is worth noting that these two sensors were the ones selected by the lasso regression for the prediction of fruity aroma intensity.



**Figure 7.** Visualization of the coefficients for the first four PCA components.

Furthermore, Figure 8 shows the weights of the fitted models for both the SVM (top panel) and logistic regression (bottom panel). The inspection of these plots suggests that PCA components 2 and 4 are the most discriminative, as the coefficients associated with these components present a higher value than the others, particularly so for the SVM. The similarity of the coefficients yielded by both methods is noticeable, although not surprising taking into account the similarity of the optimization problems solved to obtain these parameters. The results obtained using this approach are included in Table 3. These results are not bad, but far from satisfying, particularly due to a lack of robustness shown as a large fluctuation of the classification score depending on the particular seed chosen for the random number generator responsible for the five-fold partition.



**Figure 8.** Visualization of the coefficients associated with each PCA component for the Support Vector Machine (upper panel) and Logistic Regression (lower panel).

The good results obtained for the prediction of fruity aroma using the lasso motivated exploring a slightly different approach for the defect prediction problem: fitting a logistic regression with  $l_1$  regularization using the original set of 18 features. The results obtained improved those using PCA features, yielding 89.32 and 85.56% accuracy for training and testing, respectively. Figure 9 shows the coefficients yielded by the algorithm. The inspection of this plot suggested using the same approach as with the prediction of fruity aroma, i.e., fitting a new model including only the features that got assigned weights of certain magnitude. This way, two more models were fitted, one including the features S2F, S6M, S9F and S10M; and another including just S2F and S10M. The results obtained are also included in Table 3. As seen there, the approach taking only S9F and S10M offered the best results overall (85.72% and 88.06% for training and testing, respectively) and a remarkable independence of the particular partition for the five-fold validation. The reason for these good results can be appreciated observing Figure 10. This plot shows that the samples are almost linearly separable in the space spanned by these two features. A final test was performed by fitting an SVM using just these two features, but the results obtained were poorer than those offered by the logistic regression, but are also included in the last row of Table 3.

These results are in line with the performance obtained in previous works that employed e-noses for different tasks in olive oil, such as the 91% accuracy reported to classify VOO versus other chemically treated oils [36], the 84% accuracy for the discrimination between cultivars [38], the 90% accuracy in the prediction of the olive oil quality measuring intact olives [39], and the 81% accuracy in discerning between different fruity aroma profiles [20].

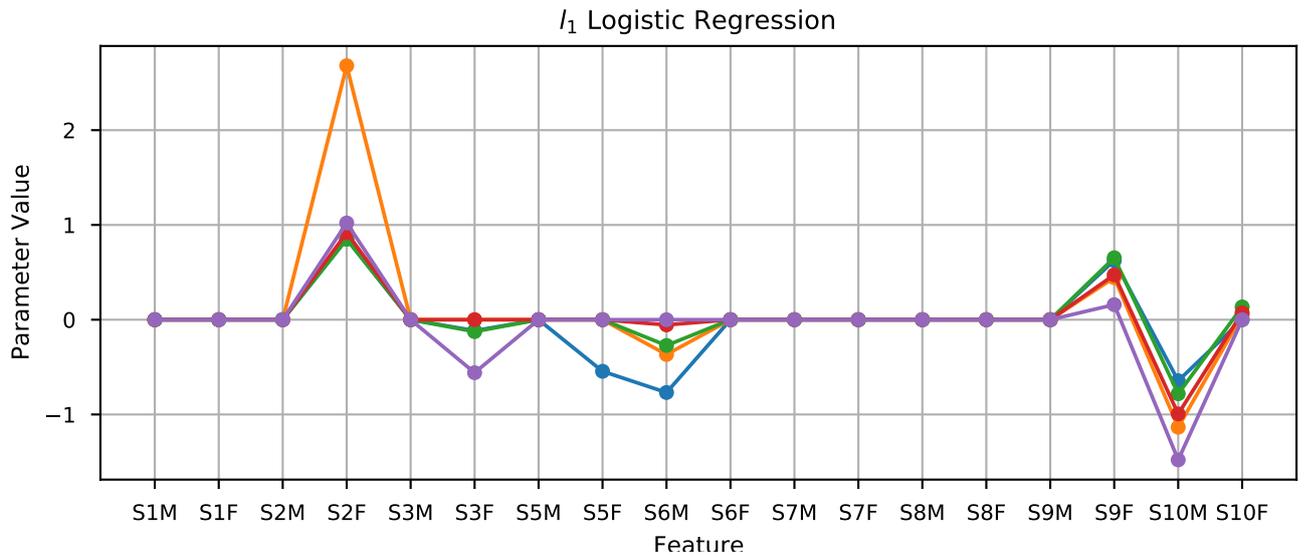


Figure 9. Visualization of the coefficients of the original features for the  $l_1$  logistic regression.

Table 3. Results obtained for the detection of defects. The best results are obtained using a logistic regression with  $l_1$  using the features S2M and S10F.

Model	Training Accuracy (%)	Test Accuracy (%)
SVM PCA	85.76	80.83
Logistic Regression PCA	87.54	80.83
Logistic Regression $l_1$	89.32	85.56
Logistic Regression $l_1$ (S2, S6, S9, S10)	88.11	85.56
Logistic Regression $l_1$ (S2, S10)	85.72	88.06
SVM $l_1$ (S2, S10)	86.36	83.06

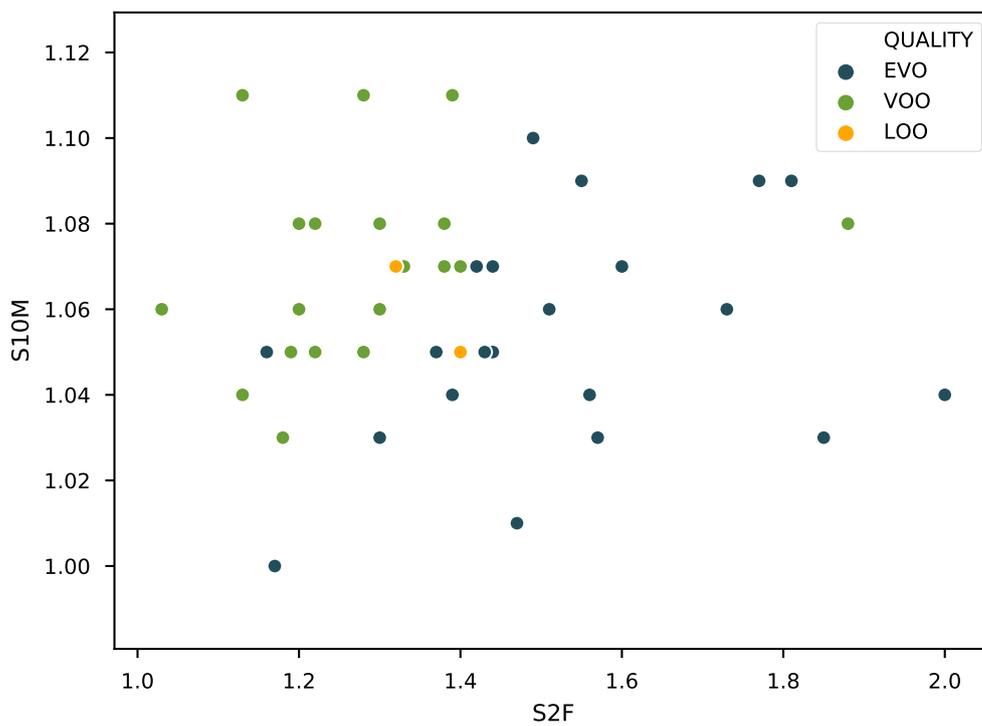


Figure 10. Visualization of the feature space yielded by S2F and S10M. The classes are almost completely linearly separable.

## 5. Conclusions

The obtained results show that the use of electronic noses is a viable approach to perform a fast screening of the quality of produced VOO, paving the way for the development of devices capable of providing at-line or online assessments of the quality of the VOO directly in the factory. These assessments could be used to store the produced VOO in tanks with other VOO of similar quality and to apply some feedback to the process.

The plots that depict the parameters of the classifiers and regressors for the different data folds demonstrate the low variability in these values, which suggests that the classifiers and regressors are robust and should perform well when applied to new samples.

Finally, a remarkable observation is the good results obtained using  $l_1$  regularization techniques and the comparable results obtained using only a subset of the available sensors. This is, in our opinion, very interesting for the construction of low-cost specific purpose electronic noses, as the number of sensors included in the device directly affects its total cost.

**Author Contributions:** Conceptualization, P.C.M., C.S., S.S.M., J.G.O., F.M. and J.G.G.; Formal analysis, P.C.M., C.S. and S.S.M.; Funding acquisition, J.G.O. and J.G.G.; Investigation, P.C.M., C.S. and S.S.M.; Writing—original draft, P.C.M., C.S. and S.S.M.; Writing—review and editing, J.G.O., F.M. and J.G.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the Spanish Ministry of Science and Innovation under the project PID2019-110291RB-I00.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to agreement terms with the company providing the samples.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Boskou, D. *Olive and Olive Oil Bioactive Constituents*; Elsevier Inc.: Amsterdam, The Netherlands, 2015; pp. 1–397. [[CrossRef](#)]
2. Hernaez, R.; Kramer, J.R.; Liu, Y.; Tansel, A.; Natarajan, Y.; Hussain, K.B.; Ginès, P.; Solà, E.; Moreau, R.; Gerbes, A.; et al. Prevalence and short-term mortality of acute-on-chronic liver failure: A national cohort study from the USA. *J. Hepatol.* **2019**, *70*, 639–647. [[CrossRef](#)]
3. Battino, M.; Forbes-Hernández, T.Y.; Gasparrini, M.; Afrin, S.; Cianciosi, D.; Zhang, J.; Manna, P.P.; Reboredo-Rodríguez, P.; Varela Lopez, A.; Quiles, J.L.; et al. Relevance of functional foods in the Mediterranean diet: The role of olive oil, berries and honey in the prevention of cancer and cardiovascular diseases. *Crit. Rev. Food Sci. Nutr.* **2019**, *59*, 893–920. [[CrossRef](#)]
4. Conte, L.; Bendini, A.; Valli, E.; Lucci, P.; Moret, S.; Maquet, A.; Lacoste, F.; Brereton, P.; García-González, D.L.; Moreda, W.; et al. 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.* **2019**. [[CrossRef](#)]
5. Sanmartin, C.; Venturi, F.; Sgherri, C.; Nari, A.; Macaluso, M.; Flamini, G.; Quartacci, M.F.; Taglieri, I.; Andrich, G.; Zinnai, A. The effects of packaging and storage temperature on the shelf-life of extra virgin olive oil. *Heliyon* **2018**, *4*, e00888. [[CrossRef](#)]
6. Sanmartin, C.; Venturi, F.; Macaluso, M.; Nari, A.; Quartacci, M.F.; Sgherri, C.; Flamini, G.; Taglieri, I.; Ascrizzi, R.; Andrich, G.; et al. Preliminary Results About the Use of Argon and Carbon Dioxide in the Extra Virgin Olive Oil (EVOO) Storage to Extend Oil Shelf Life: Chemical and Sensorial Point of View. *Eur. J. Lipid Sci. Technol.* **2018**, *120*, 1800156. [[CrossRef](#)]
7. Kiritsakis, A.K. Flavor components of olive oil—A review. *J. Am. Oil Chem. Soc.* **1998**, *75*, 673–681. [[CrossRef](#)]
8. Campestre, C.; Angelini, G.; Gasbarri, C.; Angerosa, F. The Compounds Responsible for the Sensory Profile in Monovarietal Virgin Olive Oils. *Molecules* **2017**, *22*, 1833. [[CrossRef](#)]
9. Angerosa, F.; Mostallino, R.; Basti, C.; Vito, R.; Serraiocco, A. Virgin olive oil differentiation in relation to extraction methodologies. *J. Sci. Food Agric.* **2000**, *80*, 2190–2195. [[CrossRef](#)]
10. Bendini, A.; Valli, E.; Barbieri, S.; Gallina, T. Sensory Analysis of Virgin Olive Oil. In *Olive Oil—Constituents, Quality, Health Properties and Bioconversions*; InTech: London, UK, 2012. [[CrossRef](#)]
11. Aparicio, R.; Morales, M.T.; García-González, D.L. Towards new analyses of aroma and volatiles to understand sensory perception of olive oil. *Eur. J. Lipid Sci. Technol.* **2012**, *114*, 1114–1125. [[CrossRef](#)]
12. Auvray, M.; Spence, C. The multisensory perception of flavor. *Conscious. Cognit.* **2008**. [[CrossRef](#)] [[PubMed](#)]

13. Gerhardt, N.; Schwolow, S.; Rohn, S.; Pérez-Cacho, P.R.; Galán-Soldevilla, H.; Arce, L.; Weller, P. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem.* **2019**, *278*, 720–728. [[CrossRef](#)] [[PubMed](#)]
14. Kemp, S.E.; Hort, J.; Hollowood, T. *Descriptive Analysis in Sensory Evaluation*; John Wiley & Sons, Ltd.: Chichester, UK, 2018; pp. 1–724. [[CrossRef](#)]
15. Romero, I.; García-González, D.L.; Aparicio-Ruiz, R.; Morales, M.T. Validation of SPME-GCMS method for the analysis of virgin olive oil volatiles responsible for sensory defects. *Talanta* **2015**, *134*, 394–401. [[CrossRef](#)]
16. Di Natale, C.; Paolesse, R.; D'Amico, A. Food and Beverage Quality Assurance. In *Handbook of Machine Olfaction*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2004; pp. 505–524. [[CrossRef](#)]
17. Oliver-Pozo, C.; Aparicio-Ruiz, R.; Romero, I.; García-González, D.L. Analysis of Volatile Markers for Virgin Olive Oil Aroma Defects by SPME-GC/FID: Possible Sources of Incorrect Data. *J. Agric. Food Chem.* **2015**, *63*, 10477–10483. [[CrossRef](#)] [[PubMed](#)]
18. Rusinek, R.; Siger, A.; Gawrysiak-witulska, M.; Rokosik, E.; Malaga-tobola, U. Original article Application of an electronic nose for determination of pre- pressing treatment of rapeseed based on the analysis of volatile compounds contained in pressed oil. *Int. J. Food Sci. Technol.* **2020**, *55*, 2161–2170. [[CrossRef](#)]
19. Rusinek, R.; Kmiecik, D.; Gawrysiak-Witulska, M.; Malaga-Tobola, U.; Tabor, S.; Findura, P.; Siger, A.; Gancarz, M. Identification of the olfactory profile of rapeseed oil as a function of heating time and ratio of volume and surface area of contact with oxygen using an electronic nose. *Sensors* **2021**, *21*, 303. [[CrossRef](#)]
20. Teixeira, G.G.; Dias, L.G.; Rodrigues, N.; Marx, I.M.G.; Ana, C.; Veloso, A.; Pereira, A. Application of a lab-made electronic nose for extra virgin olive oils commercial classification according to the perceived fruitiness intensity. *Talanta* **2021**, *226*. [[CrossRef](#)]
21. Zhang, X.; Cheng, J.; Wu, L.; Mei, Y.; Jaffrezic-Renault, N.; Guo, Z. An overview of an artificial nose system. *Talanta* **2018**. [[CrossRef](#)]
22. Rayappan, J.B.B.; Kulandaisamy, A.J.; Ezhilan, M.; Srinivasan, P.; Mani, G.K. Developments in Electronic Noses for Quality and Safety Control. In *Advances in Food Diagnostics*; John Wiley & Sons, Ltd.: Chichester, UK, 2017; pp. 63–96. [[CrossRef](#)]
23. Majchrzak, T.; Wojnowski, W.; Dymerski, T.; Gębicki, J.; Namieśnik, J. Electronic noses in classification and quality control of edible oils: A review. *Food Chem.* **2018**. [[CrossRef](#)]
24. Gliszczyńska-Świgło, A.; Chmielewski, J. Electronic Nose as a Tool for Monitoring the Authenticity of Food. A Review. *Food Anal. Method.* **2017**. [[CrossRef](#)]
25. Sanaeifar, A.; ZakiDizaji, H.; Jafari, A.; de la Guardia, M. Early detection of contamination and defect in foodstuffs by electronic nose: A review. *TrAC Trends Anal. Chem.* **2017**. [[CrossRef](#)]
26. Berna, A. Metal oxide sensors for electronic noses and their application to food analysis. *Sensors* **2010**, *10*, 3882–3910. [[CrossRef](#)]
27. Schweizer-Berberich, P.M.; Vaihinger, S.; Göpel, W. Characterisation of food freshness with sensor arrays. *Sens. Actuators B Chem.* **1994**, *18*, 282–290. [[CrossRef](#)]
28. Ying, X.; Zinnai, A.; Venturi, F.; Sanmartin, C.; Deng, S. Freshness evaluation of grass carp (*Ctenopharyngodon idellus*) by electronic nose. *J. Food Measur. Character.* **2017**, *11*, 1026–1034. [[CrossRef](#)]
29. Jia, W.; Liang, G.; Wang, Y.; Wang, J. Electronic Noses as a Powerful Tool for Assessing Meat Quality: A Mini Review. *Food Anal. Method.* **2017**. [[CrossRef](#)]
30. Gobbi, E.; Falasconi, M.; Zambotti, G.; Sberveglieri, V.; Pulvirenti, A.; Sberveglieri, G. Rapid diagnosis of Enterobacteriaceae in vegetable soups by a metal oxide sensor based electronic nose. *Sens. Actuators B Chem.* **2015**, *207*, 1104–1113. [[CrossRef](#)]
31. Concina, I.; Falasconi, M.; Gobbi, E.; Bianchi, F.; Musci, M.; Mattarozzi, M.; Pardo, M.; Mangia, A.; Careri, M.; Sberveglieri, G. Early detection of microbial contamination in processed tomatoes by electronic nose. *Food Control* **2009**, *20*, 873–880. [[CrossRef](#)]
32. Panagou, E.Z.; Sahgal, N.; Magan, N.; Nychas, G.J. Table olives volatile fingerprints: Potential of an electronic nose for quality discrimination. *Sens. Actuators B Chem.* **2008**, *134*, 902–907. [[CrossRef](#)]
33. Cosio, M.S.; Romano, A.; Scampicchio, M. Olive Oil and Electronic Nose. In *Electronic Noses and Tongues in Food Science*; Elsevier Inc.: Amsterdam, The Netherlands, 2016; pp. 83–92. [[CrossRef](#)]
34. Ordukaya, E.; Karlik, B. Quality Control of Olive Oils Using Machine Learning and Electronic Nose. *J. Food Qual.* **2017**, *2017*, 1–7. [[CrossRef](#)]
35. Zhou, Q.; Liu, S.; Liu, Y.; Song, H. Comparison of flavour fingerprint, electronic nose and multivariate analysis for discrimination of extra virgin olive oils. *R. Soc. Open Sci.* **2019**, *6*, 190002. [[CrossRef](#)]
36. Oates, M.J.; Fox, P.; Sanchez-Rodriguez, L.; Carbonell-Barrachina, Á.A.; Ruiz-Canales, A. DFT based classification of olive oil type using a sinusoidally heated, low cost electronic nose. *Comput. Electron. Agric.* **2018**, *155*, 348–358. [[CrossRef](#)]
37. Buratti, S.; Malegori, C.; Benedetti, S.; Oliveri, P.; Giovanelli, G. E-nose, e-tongue and e-eye for edible olive oil characterization and shelf life assessment: A powerful data fusion approach. *Talanta* **2018**, *182*, 131–141. [[CrossRef](#)] [[PubMed](#)]
38. Jolayemi, O.S.; Tokatli, F.; Buratti, S.; Alamprese, C. Discriminative capacities of infrared spectroscopy and e-nose on Turkish olive oils. *Eur. Food Res. Technol.* **2017**, *243*, 2035–2042. [[CrossRef](#)]
39. Navarro Soto, J.P.; Martínez Gila, D.M.; Artero Vázquez, E.; Gómez Ortega, J.; Gámez García, J. Sistema basado en nariz electrónica aplicada sobre aceituna para la determinación de la calidad del aceite de oliva producido. In Proceedings of the XIX Simposio Científico-Técnico de EXPOLIVA 2019, Jaen, Spain, 15–19 May 2019. [[CrossRef](#)]
40. International Olive Oil Council. *Sensory Analysis of Olive Oil. Method for the Organoleptic Assessment of Virgin Olive Oil*; Technical Report; 2011.

41. Martínez Gila, D.M.; Gámez García, J.; Bellincontro, A.; Mencarelli, F.; Gómez Ortega, J. Fast tool based on electronic nose to predict olive fruit quality after harvest. *Postharv. Biol. Technol.* **2020**, *160*, 111058. [[CrossRef](#)]
42. Carmel, L.; Levy, S.; Lancet, D.; Harel, D. A feature extraction method for chemical sensors in electronic noses. *Sens. Actuators B Chemical.* **2003**, *93*, 67–76. [[CrossRef](#)]
43. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn. *GetMobile Mob. Comput. Commun.* **2015**, *19*, 29–33. [[CrossRef](#)]
44. Hastie, T.; Tibshirani, R.; Friedman, J.H.J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009; p. 482. [[CrossRef](#)]