agronomy

MDPI

*Article*

# Classification of Monofloral Honeys by Measuring a Low-Cost Electronic Nose Prototype Based on Resistive Metal Oxide Sensors

**Eduardo González María [1], Antonio Madueño Luna [2], Agustín Conesa Celdrán [1], Gemma Martínez Muñoz [1], Martin John Oates [1] and Antonio Ruiz-Canales [1,*]**

[1] Engineering Department, Escuela Politécnica Superior de Orihuela, Miguel Hernández University of Elche, 03312 Orihuela, Spain; eduardo.gonzalez@goumh.umh.es (E.G.M.); agusconesaceldran@gmail.com (A.C.C.); gemma.veinte@gmail.com (G.M.M.); moates990@gmail.com (M.J.O.)

[2] Departamento de Ingeniería Aeroespacial y Mecánica de Fluidos, Universidad de Sevilla, Camino de los Descubrimientos, s/n, 41092 Sevilla, Spain; antonio@madueno.es

* Correspondence: acanales@umh.es

**Abstract:** In this article a case study of characterisation of type of honey based on floral origin is presented. It is intended to discriminate Iberian honeys from local beekeepers located in the Community of Madrid (Spain), by means of a low-cost electronic nose prototype, composed of a matrix of nonspecific resistive sensors of MQ-type metal oxides. The measurements of the honeys made with an electronic nose prototype were contrasted with physicochemical analyzes and pollen content. The experiment was carried out in two trials. A first preliminary study in which six samples of honey from different sources were used (three Blueweed, one rapeseed, one lavender and one commercial honey) and in which eight repetitions were made for each of the six samples analyzed. Due to the small sample size, conclusive results were not obtained, although the sensors did show a clear response in those that presented a higher pollen content, above 57%, however, the honey samples that reflected pollen values lower than 50% they showed no perceptible reaction on the sensors. In the second study, in which the sample size was increased to a total of 16 samples (four lavender honeys, four oak honeys, four rosemary honeys, and four chestnut honeys), a total of 10 repetitions per sample were carried out with a total of repetitions out of 160. These last data were analyzed with the principal component technique (PCA), the results of which were inconclusive. However, when applying the data analysis through the use of Support Vector Machines (SVM), it is possible to obtain a model with 87.5% accuracy in the classification. In this case, the Lavender and Chestnut honeys were the ones that achieved a precision of 90% and 100% respectively.

**Keywords:** honey; botanical origin; volatile organic compounds; electronic nose; metal oxide sensors



**Citation:** María, E.G.; Luna, A.M.; Celdrán, A.C.; Muñoz, G.M.; Oates, M.J.; Ruiz-Canales, A. Classification of Monofloral Honeys by Measuring a Low-Cost Electronic Nose Prototype Based on Resistive Metal Oxide Sensors. *Agronomy* **2023**, *13*, 2183. https://doi.org/10.3390/agronomy13082183

Academic Editor: Roberto Marani

Received: 8 July 2023
Revised: 5 August 2023
Accepted: 14 August 2023
Published: 21 August 2023



## 1. Introduction

The guidelines governing the European honey industry allow, among other things, for honey to be marked based on its botanical and regional source. At the moment, there's no mandate in the law that requires floral designation for honey. This absence of clarity in the rules has resulted in honey being deemed as monofloral if a specific pollen type exceeds 45% representation [1].

Currently, several methods are used to verify the authenticity of monofloral honey, one of them being melissopalynology, a fundamental tool in the verification process but which implies a high cost both in analysis time and cost. The method for analyzing honey's pollen involves two steps: recognizing the pollen grains present and tallying them. The identification can only be done by comparing the morphology and dimensions of the observed pollen grains with those known grains that constitute the reference [2].

Using data from melissopalynological studies, we can categorize or create subcategories within comparable production entities. Principal component analysis facilitates

the reduction of the dimension of the data set, by transforming it into a new uncorrelated set [3].

Alternatively, another common method for distinguishing honey, often cited in research, is by analyzing prevalent physicochemical attributes such as moisture, electrical conductivity, free acidity, carbohydrates, HMF, color, optical rotation, and pH. Tools like principal component analysis (PCA), linear discriminant analysis (LDA), and stepwise discriminant analysis (SDA) are employed for its categorization. These tools assess data trends and uncover the links between the physicochemical attributes and honey's botanical source [4] Recently, methods to analyze honey's volatile components have emerged, aiding in the detection of chemical markers within honeys, facilitating the distinct identification of monofloral varieties. Among these methodologies, one of the most promising is the analysis of the volatile fraction of honeys by means of SPME-GC-MS (Solid Phase Microextraction—Gas Chromatography-Mass Spectrometry), since the compounds in this fraction are closely related to the organoleptic perception that the consumer receives from them. [5]

Honey, as such, contains small amounts of organic acids, amino acids, minerals, vitamins, phenolic compounds, and volatile compounds, mainly carboxylic acids, aldehydes, alcohols, and terpenes [6,7].

The volatile organic elements in honey can be categorized into distinct chemical classes, acting as a unique signature for each honey type [8]. Several researchers have directly linked the impact of plant species and the flora accessed by foraging bees to both the physicochemical and organoleptic traits of honey [9]. Additionally, the climate and soil conditions of the collection area, as well as the beekeeper's extraction and storage techniques, also play a significant role [10].

More than 300 volatile compounds belonging to different chemical families have been isolated: esters of aliphatic and aromatic acids, aldehydes, ketones and alcohols [11]. All of them contribute to the aroma and flavor of honey, along with sugars and acids [4] There are volatile compounds characteristic of certain honeys such as methylanthranilate for citrus honeys [12] and lavender and thyme [13].

The limitation for a low-cost electronic nose project is determined by the commercial availability of gas sensors. The parameters of the electrical circuit are based in the recommendations of the manufacturers in order to obtain a standard approach. For example, the voltage of the sensor heater is determining the temperature of the sensor. Moreover, another aspect of the enoses is the use of metal oxide gas sensors (MQ). The use of these sensors for quick classification has been extended in this type of devices [14].

The analysis for the differentiation of honeys based in VOCs based on the combination of statistical methods through principal component analysis and linear discriminant has been obtaining good results [6].

Common multivariate statistical methods used to identify the botanical or geographical roots of honey focus on its physicochemical, antioxidant properties, or chemical makeup. These methods include principal component analysis (PCA), linear discriminant analysis (LDA), cluster analysis (CA), and artificial neural networks (ANN). The origin determinations center around the honeys' physicochemical and antioxidant characteristics or their chemical structure [15].

Other honey characterization research often encompasses a variety of parameters, including physicochemical, sensory, or both, and occasionally alongside the pollen spectrum. For classifying these parameters, techniques such as cluster analysis, PCA, and LDA are typically employed. Additionally, the Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR) of monofloral honeys, when processed through PCA and later, using a machine learning algorithm like the support vector machine (SVM), has demonstrated effectiveness in pinpointing honey's origin [16].

An essential component in honey is pollen. Bee pollen is a mixture of pollen pellets of different colors collected by *Apis mellifera* on flowers of various species, removed by complex movements between the legs and the buccal appendages, compacted with nectar on the corbicles of its hind legs, and transported to the hive [17].

Numerous authors have described the chemical composition of pollen, in which the components vary depending on environmental, botanical and processing factors. Carbohydrates, proteins and essential amino acids, essential fatty acids, dietary fiber, vitamins and trace minerals [18] as well as carotenoids, flavonoids and polyphenols [19] have been described. More recent studies have highlighted the richness in phenolamides for bee pollen of the *Castanea* species [20] and flavonoid glycosides detected in *Echium plantagineum* bee pollen [21].

Aroma significantly influences the sensory characteristics of foods. Numerous studies highlight the potential of the e-nose in the food sector, characterized as a fast tool with a simple methodology [22] notably in distinguishing honey based on its geographical, botanical [19], and physicochemical attributes [23].

Taking into account the previously mentioned chemical characteristics of both honey and pollen, in this study an electronic nose prototype made up of eight MQ (resistive metal oxide) sensors has been used to analyze and botanically characterize different types of honey based on their chemical fingerprint. A PCA analysis was used to pool the honey samples and a comparative study was carried out using basic physical-chemical analyzes of honey as well as pollen analyzes in order to support the results of both electronic nose prototypes. With the results obtained, the potential capacity of these devices to be able to discriminate honey samples based on their botanical origin was supported in a similar way to a melissopalynological study, reducing both the time and the cost of the analysis. In addition to the PCA analysis, the Support Vector Machine (SVM) technique was used to try to improve the analysis of the data obtained and relate the measurements of the aforementioned device with the physicochemical and pollen analyses.

## 2. Materials and Methods

### *2.1. Used Honeys*

The honey samples were provided by the Association of Beekeepers of the Community of Madrid (Spain) and local beekeepers from the northern mountains of Madrid. For analysis with the e-nose, samples were prepared in glass containers with 50 g of each type of honey. The measurement with the device was carried out at room temperature, without carrying out any type of manipulation or previous preparation of the sample.

Two experiments were carried out. A first preliminary study in which the capacity and sensitivity of the electronic nose prototype (e-nose) to detect the different honeys was assessed. Although it is not possible to observe any grouping based on the botanical origin of the honeys, it is observed that the highest response values of the sensors coincided with the honeys that presented the highest percentage of pollen, corresponding to samples 2 (Blueweed 74%), sample 3 (Blueweed 62%) and sample 4 (Rapeseed 57%) respectively, and that this lack of grouping may be due to the low number of samples. It was considered that increasing the number of samples would help to avoid this phenomenon. Maintaining this assumption, a second, more exhaustive study was carried out, in which both the number of samples and the number of experiments and repetitions were increased.

In the preliminary study, eight different honeys were used, obtained from different batches and harvested in the year 2020. Pollen analyzes classified these samples as monofloral. On the other hand, the values obtained in the physical-chemical analysis can be seen in Table 1; the values of humidity, electrical conductivity, the HMF and the pH are within the values established in the current legislation [24].

Employing statistical approaches on physicochemical data has proven effective in identifying various honey types. Among the chemical indicators commonly used to determine the floral source of honey samples are electrical conductivity, pH, and Hydroxymethylfurfural (HMF) [4].

**Table 1.** Physical, chemical and pollen analysis of the honey samples used in the first experiment.

| Honey | Reference | BRIX Degrees (%) | Humidity (%) | pH | Mm Pfund | Conductivity (mS/cm) | Pollen from Nectariferious (%) | Color |
|---|---|---|---|---|---|---|---|---|
| Blueweed | m1 | 82.3 | 14.8 | 4.05 | 37 | 0.27 | 50 | Extra light amber |
| Blueweed | m2 | 80.5 | 17 | 3.8 | 37.3 | 0.26 | 62 | Extra light amber |
| Blueweed | m3 | 80.5 | 17.4 | 4.47 | 43 | 0.3 | 74 | Extra light amber |
| Rape | m4 | 81.1 | 18 | 4.2 | 28 | 0.32 | 57 | White |
| Lavender | m5 | 82 | 15.1 | 3.6 | 40.3 | 0.16 | 50 | Extra light amber |
| Forest | m6 | | 18 | | 90 | 0.9 | | |

On the other hand, in the second study 16 honey samples were used, classified as lavender, chestnut, oak and rosemary honey (see Table 2). In this case, different honeys had to be used since it was not possible to obtain honeys with the same botanical origin as the initial ones. These honeys were harvested in 2022.

**Table 2.** Physical, chemical and pollen analysis of the honey samples used in the second experiment.

| Type | Sample Number | % BRIX | % Humidity | pH | Conductivity | Pfound Value (mm) | Color | Yeasts Content (10 g) | Description | Grains | Grains Nectariferous | Grains Pollineferous | Polien Type | Nectariferious (%) | Pollen (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lavender | 2065 | 82 | 16 | 4.18 | 0.36 | 47 | 0 | 7783.88 | Very low | 687 | 400 | 287 | 119 | 29.70 | 17.30 |
| Lavender | 3002 | 83 | 15 | 4 | 0.41 | 59 | 1 | 12,838.12 | Low | 500 | 400 | 100 | 37 | 9.25 | 7.4 |
| Lavender | 3007 | 85 | 17 | 4.01 | 0.48 | 57 | 1 | 18,231.1 | Low | 540 | 446 | 94 | 16 | 3.59 | 2.96 |
| Lavender | 3011 | 82 | 16 | 3.98 | 0.45 | 63 | 1 | 12,838.12 | Low | 500 | 400 | 100 | 44 | 11 | 8.8 |
| Chestnut | 2026 | 79 | 16.5 | 4.82 | 1.16 | 145 | 2 | 51,625.46 | Low | 500 | 425 | 75 | 166 | 39.06 | 33.2 |
| Chestnut | 2075 | 81 | 16.5 | 4.66 | 0.97 | 115 | 2 | 11,707.04 | Low | 803 | 712 | 91 | 399 | 56.03 | 49.68 |
| Chestnut | 2094 | 82 | 16.5 | 4.74 | 1 | 116 | 2 | 85,465.32 | Low | 1313 | 1073 | 240 | 454 | 42.3 | 34.5 |
| Chestnut | 22,109 | 81.5 | 16 | 4.85 | 1.16 | 131 | 2 | 85,465.32 | Low | 1313 | 1073 | 240 | 620 | 57.78 | 47.22 |
| Rosemary | 2001 | 83 | 14 | 4.26 | 0.1 | 14 | 4 | 5558.96 | Very low | 570 | 400 | 170 | 78 | 19.5 | 13.68 |
| Rosemary | 2007 | 80 | 17 | 4.12 | 0.09 | 19 | 3 | 2963.47 | Very low | 1004 | 401 | 603 | 125 | 31.17 | 12.45 |
| Rosemary | 2030 | 81 | 16 | 3.98 | 0.2 | 33 | 3 | 34,118.37 | Low | 508 | 463 | 45 | 15 | 3.23 | 2.95 |
| Rosemary | 2152 | 81 | 16 | 4.03 | 0.23 | 52 | 1 | 4533.88 | Very low | 736 | 400 | 336 | 87 | 21.7 | 11.8 |
| Oak | 2027 | 83.5 | 14 | 5.07 | 1.23 | 119 | 2 | - | - | - | - | - | - | - | - |
| Oak | 2042 | 83 | 15 | 4.87 | 1.16 | 104 | 5 | - | - | - | - | - | - | - | - |
| Oak | 2071 | 79 | 18.5 | 5.02 | 1.17 | 102 | 5 | - | - | - | - | - | - | - | - |
| Oak | 2078 | 83 | 14 | 4.26 | 1.21 | 107 | 5 | - | - | - | - | - | - | - | - |

In the case of oak honey, which comes mostly from plant secretions or excretions from hemiptera (plant-sucking insects), the pollen load is not considered [19]. This type of honey has other characteristics in its composition that allow it to have other exceptions compared to other honeys, such as sugar content or electrical conductivity. This circumstance will also be reflected in the detection of pollen components by the sensors most related to these substances.

### *2.2. Devide Used: E-Nose Prototype*

For the analysis of the honey, a device composed of an array of eight MQ resistive sensors (Table 3) (MQ2, MQ3, MQ4, MQ5, MQ7, MQ8, MQ9 and MQ135), Hanwei Electronics Co., Ltd.® (Zhengzhou, China), with an Arduino NANO® as acquisition equipment and data processing have been used. MQ gas sensors are electrochemical sensors and their resistance varies when exposed to certain gases. Its interior consists of a heater that is responsible for increasing the internal temperature. Thanks to this, the sensor reacts with the gases causing a change in the resistance value. The electrical parameters of the sensors (sensitivity among others) can be seen in [25].

**Table 3.** Used sensors in the used e-nose prototype and their sensibilities.

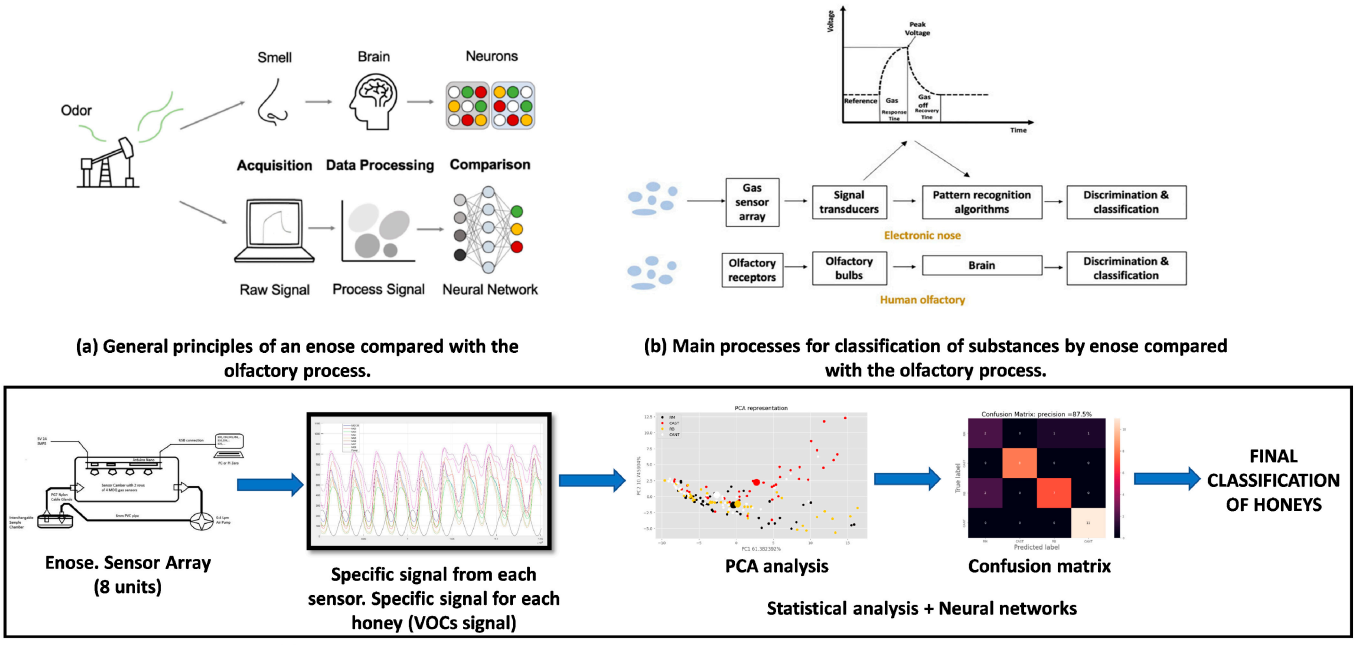
| Nº | Sensor | Sensible to |
|---|---|---|
| 1 | MQ2 | LPG (Liquefied Petroleum Gases), Hydrogen and Propane |
| 2 | MQ3 | Alcohol |
| 3 | MQ4 | Methane |
| 4 | MQ5 | Hydrogen and LPG |
| 5 | MQ7 | Hydrogen and carbon monoxide |
| 6 | MQ8 | Hydrogen |
| 7 | MQ9 | Carbon monoxide and liquefied petroleum gases (LPG) |
| 8 | MQ135 | $NH_3$ (ammonia), $NO_x$, alcohol, benzene, smoke, $CO_2$, etc. |

These affordable commercial resistive sensors feature a sensitive layer composed of a metal oxide, specifically tin dioxide ($SnO_2$). When exposed to a gas, the metal oxide facilitates the gas's ionization, leading to electron movement and changes in the layer's conductivity. Since this mechanism is temperature-dependent, the sensor includes a heater to sustain the desired operating temperature.

Each sensor is designed to detect a range of organic chemicals and gases (Table 3). All of these sensors are sensitive to a wide variety of gases. A specific substance, in this case honey, generates a specific signal. This signal depends on the specific number and quantity of volatile organic compounds (VOCs) of each honey. Subsequently, each honey presents a specific signal that can be distinguished among several honeys.

The chemical categories of the volatile substances found in honey encompass: hydrocarbons; aldehydes; alcohols; ketones; acids; esters; derivatives of benzene, furan, and pyran; norisoprenoids; terpenes and their derivatives; along with sulfur and cyclic compounds [26].

While the sensors don't directly detect the mentioned volatile organic compounds (VOCs) in honey, the goal of the experiments was to see if these VOCs might influence the sensors, potentially providing a chemical signature to differentiate each honey variety (Figure 1).




(a) General principles of an enose compared with the olfactory process.

(b) Main processes for classification of substances by enose compared with the olfactory process.

(c) Main processes for classification of honeys by enose in the experiment.

**Figure 1.** Schematics of the detection procedure in the experiment. (**a**) General principles of an enose compared with the olfactory process; (**b**) Main processes for classification of substances by enose compared with the olfactory process; (**c**) Main processes for classification of honeys by enose in the experiment.

Traditionally, honey quality attributes are assessed using standard analytical methods. Numerous studies detail the evaluation of honey's physicochemical properties, such as moisture level, pH, ash content, invertase enzyme activity, electrical conductivity, acidity, sugar breakdown (like sucrose, glucose, and fructose), color, and hydroxyl methyl furfural (HMF) [27,28]. These are employed to distinguish between various honey types or to detect adulteration. While the accuracy of these methods is established, they tend to be costly, labor-intensive, and destructive, making them unsuitable for real-time sorting. For industrial quality assurance, it's essential for managers to adopt testing procedures that can efficiently manage a high volume of samples at an affordable cost.

This equipment has been successfully tested in previous experiments [29–31] whose results have shown potential as a complementary detection tool in the agri-food sector. The main novelty of this device compared to previous versions is based on the improvement of the electronic design compared to the prototypes used in previous experiments [29–31] (Figure 2). In addition to this change in electronic design, new lower power consumption regulators were included to prevent overcharging and overheating.

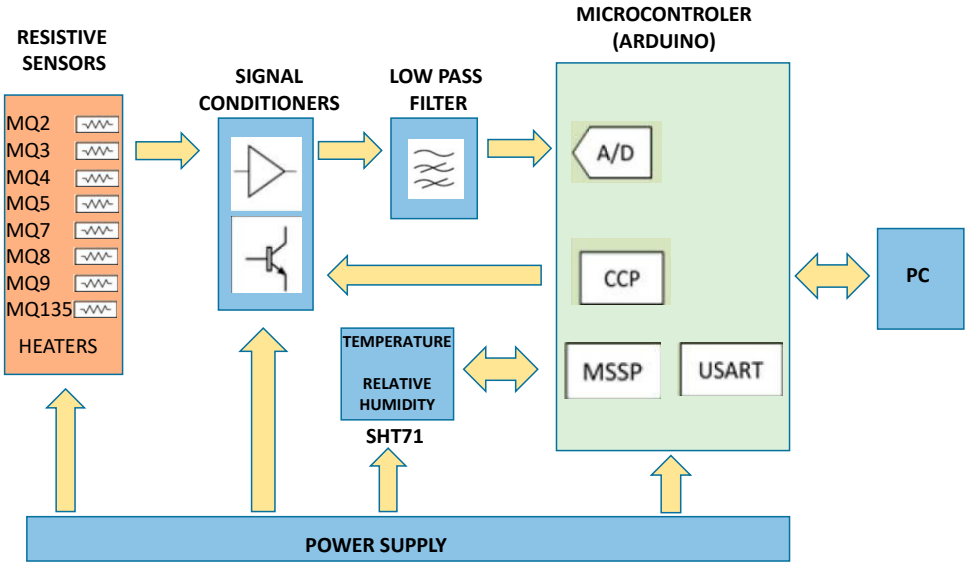



**Figure 2.** Flux diagram design of used electronic nose prototype.

In the original design, the main regulator, supported the entire load of the circuit making it get quite hot. The secondary regulators also had the same heating problem. The design was improved by separating on feed rails to lower the heating temperature of the bucks. These presented a good improvement and there was no heating problem, the 9 V to 5 V buck can support up to 3.5 A without a heatsink. Figure 3 includes the separation in power rails and the improved circuit of the motherboard of the device.

The air aspiration tube that allows data collection is connected to the sample (Figure 4). This connection tube is made of PVC, and 6 mm in external diameter (4 mm in internal diameter), and connects at the other end to the 700 mL sensor chamber made of PP5 (food grade PVC) that contains the matrix of MQ sensors. The device has a 0.4 L/min air pump that recirculates the air inside the sample chamber. In the latest version of the prototype, the air supply tubes have also been changed for more rigid and higher quality ones (Teflon). This prevents deterioration that occurs during sampling. Because of the size of the sample and the low air flow, the influence in the generation of smells is very low. Nevertheless, for a commercial purpose, the use of alternative materials to PVC will be taking into account.

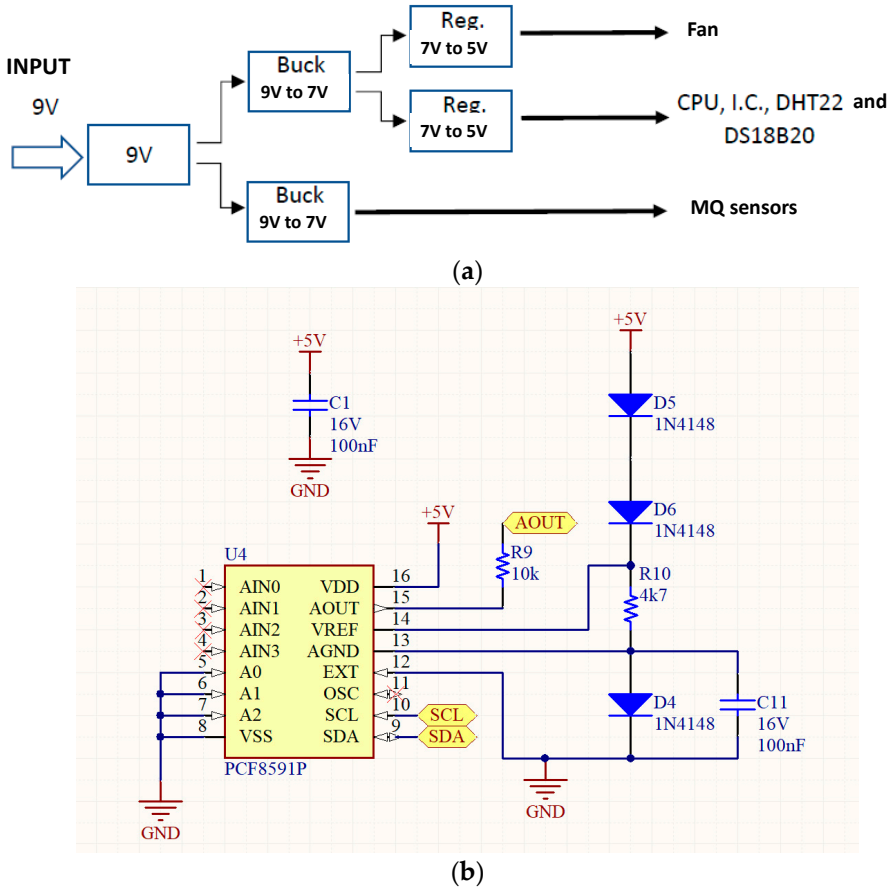



**Figure 3.** Improvements of the current enose compared with the previously used. (**a**) Separation in power rails; (**b**) Improvement introduced in the motherboard with the addition of a new A/D converter. The red cross are the zones that have been changed respect to the initial version.

All sensors must be burned for around 48 h to remove any remaining moisture. After burn-in, the trim pot resistors used for each sensor must be adjusted to obtain a 2 volt response on fresh air (the reference value without honey). If the device remains unused for over three months, the 48-h burn-in procedure should be redone.

The resistance of the sensor, which is the one that is sensitive to the gases present in the sample, receives a sinusoidal power supply with voltage values that vary between 1.5 V and 4.5 V with a period of 2 min. This introduces a variation in the temperature of the sensor, which according to the literature on the use of the sensors and previous tests of ours has shown to increase the sensitivity and, above all, to avoid the drift process in the sensors. The drift process refers to the phenomenon whereby the value we read from sensors in stable situations drops over time due to sensitivity drift processes. The sensor sampling is done every 0.5 s with the ADC of the Arduino Nano at 10 bits of resolution, that is, it is sampled at 2 Hz.

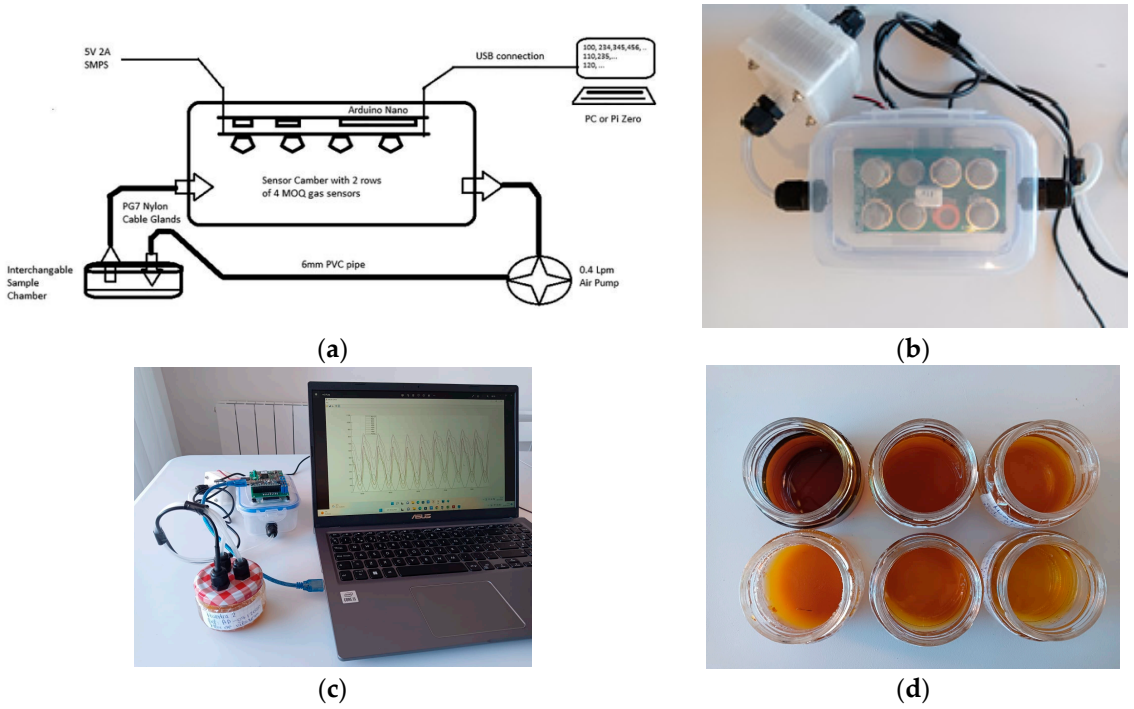



**Figure 4.** Device used: (**a**) Diagram of the electronic nose prototype used. (**b**) Matrix of MQ sensors inside the sampling chamber together with the SHt71 humidity and temperature sensor (in the center of the chamber). (**c**) Experimental procedure for one of the case study honeys. (**d**) Honeys used in the previous experiment.

### *2.3. Data Analysis*

For analyzing the raw data (unprocessed) captured by the electronic nose, the initial action taken was executing a Discrete Fourier Transform (DFT) for each cycle of the input signal [27]. The previous analysis of these data was performed with proprietary software based in Matlab R2022b. Currently this subsequent DFT analysis is done manually and an application is being developed to do it automatically.

Subsequently, the Python environment was used to perform a Principal Component Analysis (PCA) as an unsupervised linear method. The grouping that the e-nose could find among the samples is observed. This analysis allows to simplify the variables studied, their physicochemical composition, pollen analysis and the obtained values from the electronic nose prototype. The technique produces a fresh collection of variables known as principal components. Every principal component is derived from a linear blend of the initial variables. These principal components are mutually orthogonal, ensuring no overlapping information. This collection of principal components creates an orthogonal foundation for the data domain.

In the second experiment, the machine learning (ML) algorithm was used through Support Vector Machines (SVM). This method has given good results in other recent studies related to honey adulteration or classification that have combined machine learning tools [32,33] with other analysis techniques such as isotope profiles, hyperspectral microscopy technology, or infrared spectroscopy [34].

The outcomes from the e-nose readings, taken earlier, were compared against the physicochemical and pollen evaluations of each honey sample to gauge and assess the classification effectiveness of the utilized prototype.

## 3. Results

### *3.1. Results of the First Experiment*

On the one hand, the RAW values (raw and unprocessed values) returned by the AD converter were studied, analogue values that oscillate integers between 0 and 1023 (0–5 V)

corresponding to the data directly from the sensors, with a resolution of 10 bits. In the first study, the experiment was repeated eight times for each of the six samples analysed, showing a clear response from the sensors in those with a higher pollen content.

The most observed trend is the decrease of the sensor signal over time (Figure 5), events that have also been supported by other studies [35].

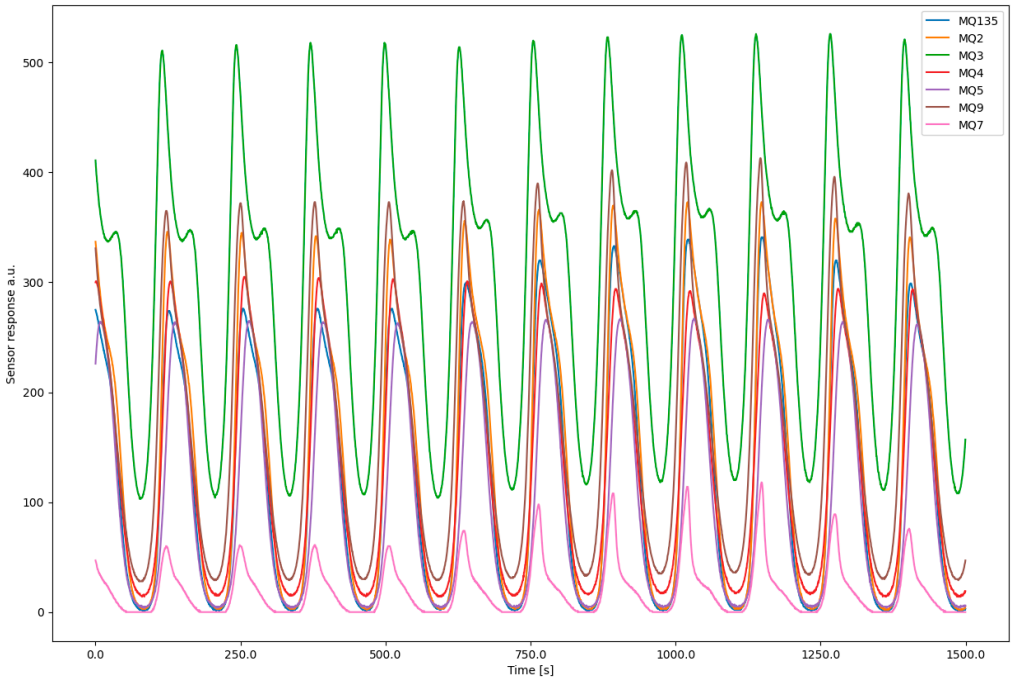


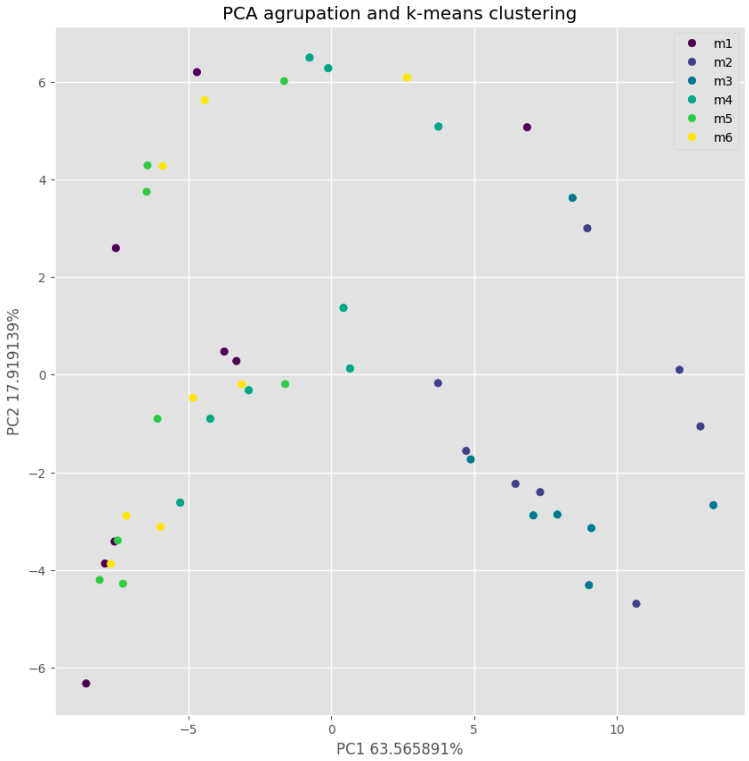

**Figure 5.** Response of the eight sensors to honey sample 3 (Blueweed 74% pollen content) together with the voltage signal introduced into the sensors.

To process the raw data, the initial step involved executing a Fourier transform (DFT) for every experiment via specialized software. Each sample underwent analysis for a duration of ten minutes, with the device's sample rate configured at 2 Hz. Consequently, each sampling yielded 1200 values, which were processed through the Matlab FFT function, producing cosine and sine coefficients along with five harmonics encompassing the DC component for every coefficient.

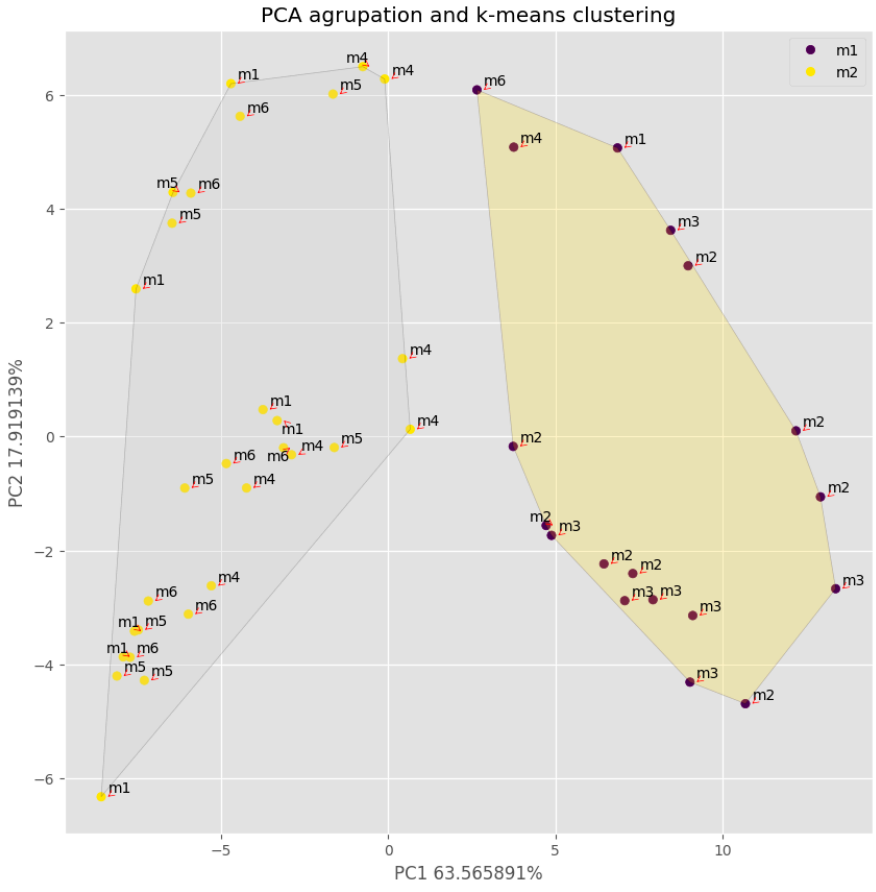
A Principal Component Analysis (PCA) was then conducted to visualize and cluster the samples. Both the PCA and the visualization were executed using the Matlab® R2020a version. The accumulated data was arranged into a matrix, subsequently normalized, and then broken down into its principal components (PCs).

In this first preliminary study, a grouping by individual samples was not observed, that is, the samples are not clearly distinguished. However, it can be seen in Figure 3a that there is a certain division by color, that is, on the left it can be seen that the samples m2 and m3 (Blueweed 62% and Blueweed 74%) that are grouped in contrast to the others that are grouped to the right of the cloud of points.

This conclusion, it is seen more clearly in Figure 6b. After using the k means algorithm to classify the data set into two groups, it is observed that the Blueweed samples m2, m3 can be grouped as one group and the rest of samples can be grouped in another group.




(**a**)




(**b**)

**Figure 6.** Results of the PCA analyses in the preliminary study: (**a**) Graph showing the PCA analysis of all the samples and each color indicates to which of the 6 types of samples each point belongs. (m1,

m2 and m3, different concentrations of Blueweed honey; m4, rape honey, m5, lavender honey, m6, forest honey); (**b**) Application of the k means algorithm to classify the data set into 2 groups in an unreinforced way. (Blueweed samples m2, m3 in (**a**) can be grouped as one group (m1, blue) and the others m1, m4, m5 and m6 in (**a**) can be grouped as another group (m2, yellow).

Although a clear classification for each of the honeys is not observed, it is shown that the prototype manages to define a classification based on its pollen content, which reinforces our hypothesis that despite being non-specific sensors for VOCs, it can have a potential classifier for honeys.

### *3.2. Results of the Sesond Experiment*

In this second study, sixteen honey samples were used, also obtained from local beekeepers in the Sierra Norte area of the Community of Madrid (Spain) and harvested in the autumn of 2022. It should be noted that these honeys, obtained mainly from local apiaries with hives on shelves, that is, that beekeeping whose hives remain all year in the same settlement. they are closely linked to the predominant flowering of the year in which it is harvested, and due to recent years marked by climatic extremes, it has made it difficult to obtain the same monofloral honeys as in the first study (Blueweed, Rape, Lavender and Forest).

In this case, the studied honeys were classified as Rosemary (RM), Oak (RB), Lavender (CANT), and Chestnut (CAST). For each of the varieties, they provided us with four samples from different batches, which allowed us to analyze a total of sixteen honeys with different pollen loads. As in the preliminary case, a physicochemical analysis was performed for each of the samples, the results of which are shown in Table 2 and which were used to compare with the results obtained from the prototype e-nose.

The Oak (honeydew) and Chestnut samples present a higher conductivity and color than the rest of the samples. These honeys have similar physicochemical properties and, frequently, similar pollen spectrum profiles that in some studies have complicated the differentiation between them [36] but on the contrary, this fact does allow them to be clearly differentiated from the other samples as can be seen. See in Figure 4.

A prior PCA was performed to compare the grouping performed. For this analysis, the values obtained from Table 2 were used, with the exception of the Oak honeydew values. In the case of honeydew such as Oak honey, melissopalynology is not useful to identify the botanical origin of this type of honey [37].

Figure 7 shows the PCA representation made with the data from the physicochemical analysis. Each of the points corresponds to each of the 12 samples. The color indicates the label of each sample relating it to the group to which it belongs according to the indicated legend. The largest points correspond to the centroid of each of the clusters corresponding to each group. The centroid indicates the average around which the points of the same group are grouped. It can be seen how three clusters have been formed that can be related to the groupings that we intend to make according to the type of honey. Bottom right, with greater separation from the other two, the orange cluster corresponding to Chestnut can be seen. This one presents a great internal variability between samples but it differs well from the other clusters. On the left, it can be seen that the clusters corresponding to Lavender and Rosemary are much closer, indicating greater similarity between them. It is observed that one of the Rosemary samples, located in the lower left part of its cluster, is closer to the Lavender cluster. That is, these two groups (Lavender and Rosemary) present a very high similarity according to their physicochemical analysis data.

According to the physicochemical values, the graph shows a greater similarity between the rosemary and lavender samples, with chestnut being more differentiated. This may be because it has a higher conductivity and color than the rest of the samples.

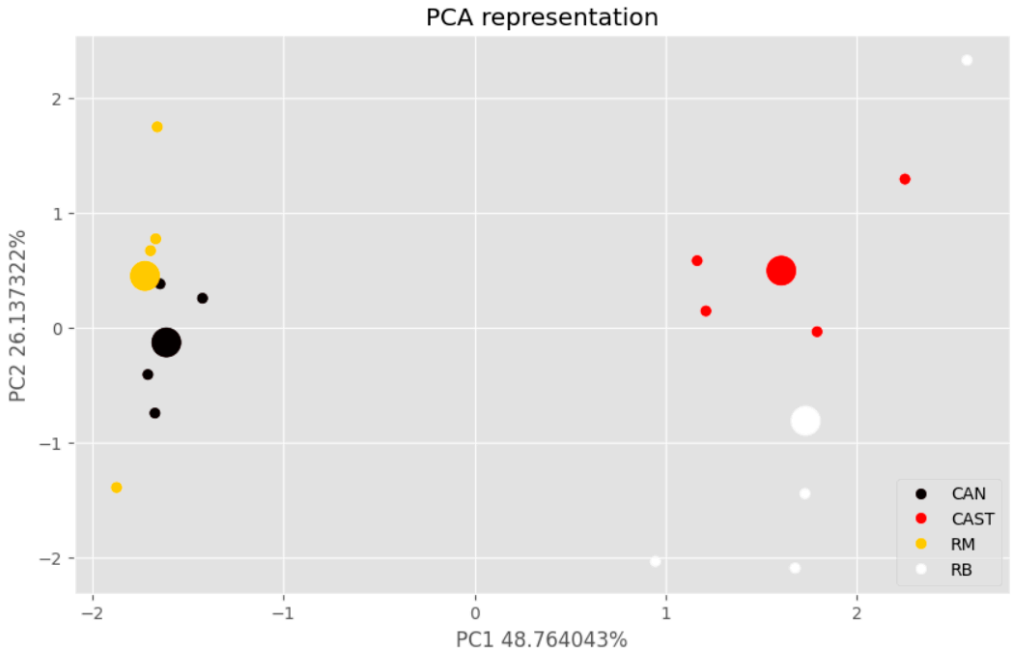



**Figure 7.** Representation of the first two main coefficients of the PCA carried out with the data obtained from the physicochemical analysis, carrying 74.89% of the variance of the data.

Two outliers are also observed, one from Oak and the other from Rosemary, which are far from the rest of the data set, and which clearly affects the correct grouping of the cluster. This may be due to one of the limitations that PCA has as a dimensionality reduction method, since it only contemplates linear combinations of variables, which means that it is not capable of capturing other types of relationships.

Regarding the electronic nose, a total of ten measurements were made for each of the samples, obtaining on the dataset containing 160 data corresponding to the measurements made by electronic nose on honey samples. These honeys have been grouped by pollen content into 4 groups, Rosemary (RM), Oak (RB), Lavender (CANT), and Chestnut (CAST).

In addition to the measurements by electronic nose, the results of a physicochemical analysis have been analyzed, shown in Table 4, containing measurements of the molasses used. A statistical study has been made according to these data of 3 of the RM, CAST, and CANT honeys. Each of these groups is made up of 4 samples, so there is a total of 12 samples studied. The samples corresponding to the RB group have been omitted since the pollen analysis has not been carried out on them.

**Table 4.** Metrics to evaluate the performance of the SVM model when classifying among the four classes.

| | Precision [1] | Recall [2] | F1-Score [3] | Support [4] |
|---|---|---|---|---|
| Rosemary | 0.80 | 0.67 | 0.73 | 6 |
| Chestnut | 1.00 | 0.90 | 0.95 | 10 |
| Oak | 0.75 | 0.86 | 0.80 | 7 |
| Lavender | 0.90 | 1.00 | 0.95 | 9 |
| Accuracy | 0.87 | - | 0.88 | 32 |
| Macro avg [5] | 0.86 | 0.86 | 0.86 | 32 |
| Micro avg [5] | 0.88 | 0.88 | 0.87 | 32 |

[1] Precision: The proportion of instances classified correctly in all classes. [2] Recall: The ability of the model to correctly identify all cases of a specific class. [3] F1 Score: The harmonic mean of accuracy and recall, providing a balanced measure for each class. [4] Support: Refers to the number of cases of each class. It gives an indication of the distribution of cases among the classes. [5] Macro/Micromean of Precision, Recall and F1 Score: Macromean calculates metrics independently for each class and then averages, while micromean aggregates contributions from all classes to calculate metrics.

The first objective was established to compare the initial grouping of the four groups mentioned above, with the groupings of the physicochemical data and those measured by electronic nose. In this way it was determined if the nose detects and differentiates the samples according to the initial hypothesis of the four groups.

Figure 8 shows the PCA representation made with the data from the analysis carried out by electronic nose. Each of the points corresponds to each of the 160 samples analyzed with the nose. In this case, the data of the 4 groups studied are presented. Each color corresponds to one of them established in the legend of the figure. The first observation of interest is; how a large part of the samples are indistinguishable by group, located in the central left part of the figure, while other clusters are formed by groups. In the upper left, the cluster corresponding to Chestnut (CAST, red) is the one that is best distinguished from the rest of the samples. It is seen that its centroid is the farthest from the rest and that a large number of samples are located in the upper right part of the figure. Secondly, it should be noted that the samples corresponding to Oak (RB, yellow) and Rosemary (RM, black) are the ones that are most similar to each other. The centroids are at the smallest distance observed and the samples are placed on exactly the same area of the figure. Finally, it can be seen how Lavender (CANT, white), despite the fact that the centroid distances itself from the rest, the samples are distributed together with the left central cluster.

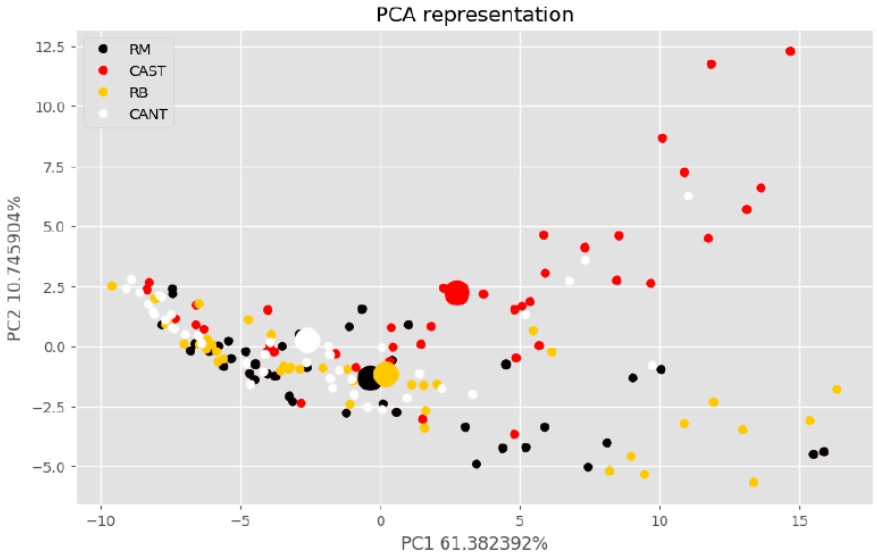



**Figure 8.** Representation of the first two main coefficients of the PCA carried out with the data obtained from the analysis by electronic nose, taking 80% of the variance of the data. Chestnut (CAST), red, Rosemary (RM), black, Lavender (CANT), white and Oak (RB), yellow.

In Figure 5 it can be seen that Chestnut honey, as in the case of the PCA analysis carried out with its physicochemical parameters, differs from the rest of the honey mainly due to its phenolic components. This may be due to the fact that chestnut honey is a monofloral honey in which the largest number of phenolic compounds have been identified to date [38], which allows increasing the sensitivity of the sensors and classifying it more clearly with respect to the rest of honey

The same happens with oak molasses honey, whose investigations have determined that it has a higher content of polyphenols and flavonoids [37] compared to other types of honey. Finally, supervised learning models have been used to classify each of the samples within each of the groups that have been established and to which they are considered to belong. The best results were obtained by using Support Vector Machines (SVM). The model obtained was trained with 80% of the data and with a quality of 20%. It is possible to obtain a model with 87.5% accuracy in the classification. Next, Table 4 details a series of measurements obtained.

Finally, the machine learning (ML) algorithm is shown using Support Vector Machines (SVM).

For the statistical categorization of the samples, we constructed a confusion matrix (Figure 9). Every column in the matrix denotes the count of predictions for each class, whereas each row indicates the instances belonging to the true class. With this matrix is possible to evaluate the prediction capacity of the obtained algorithm based in SVM. This algorithm is relating the measured values samples with the used enose against the type of honey. As it is presented in Figure 9, the precision of the predictive algorithm is 87.5%. This is a good fit and the enose can be employed for a quick identification of the studied honeys.

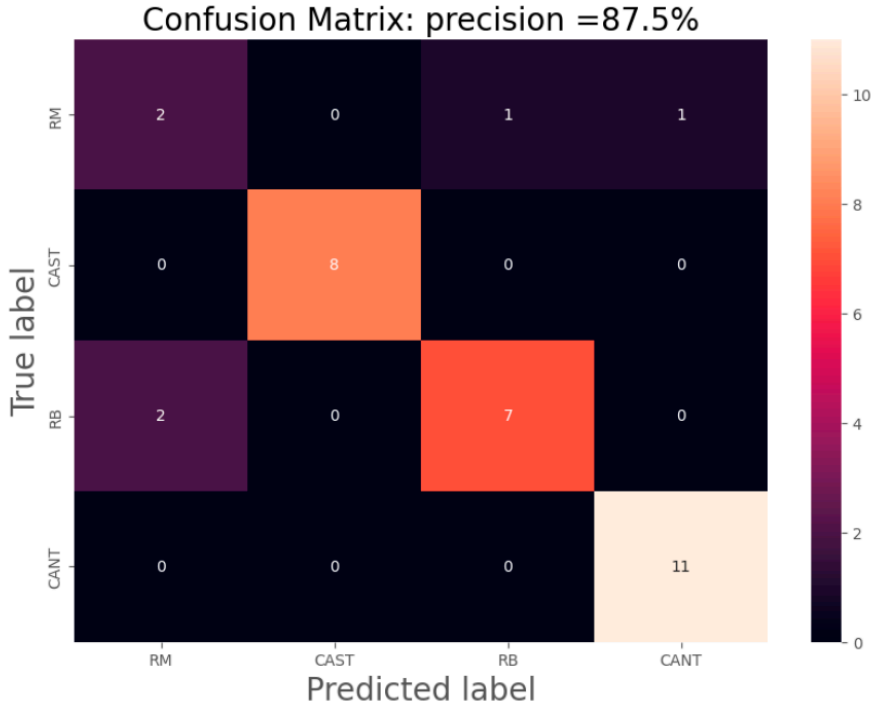



**Figure 9.** Confusion matrix as a measure of the model's performance. The code corresponding to each group is detailed on the axes.

Finally, a study is carried out whose objective is to identify which are the sensors of the electronic nose that contribute the most to the predictive capacity of the SVM model. The Permutation Importance technique will be used to evaluate the importance of the 72 variables in a data set consisting of 160 samples.

Permutation Importance is a technique that allows you to measure the relative importance of each variable in a machine learning model. In this case, it will be applied to evaluate how each of the 72 previously mentioned variables contributes to the predictive capacity of the model.

With this technique, a model is first trained using the entire dataset containing all 160 records and 72 variables. The initial performance of the model is then calculated. Each of the variables is selected, the values of one of them are mixed, while the other variables remain unchanged in each permutation. The performance of the model with the permuted variable is evaluated, and it is compared with the initial performance to establish the importance of this variable. A greater decrease in performance will indicate a greater importance of the variable (Figure 10).

Sensor MQ3 (alcohol) is the most important, while MQ9 (carbon monoxide and liquefied petroleum gases) is the least relevant. As expected, the MQ3 sensor is an important variable due to its sensitivity to alcohols and to the fact that the chemical composition of both honey and pollen is made up of these. The MQ135 is also important due to its wide sensitivity to different sensors.

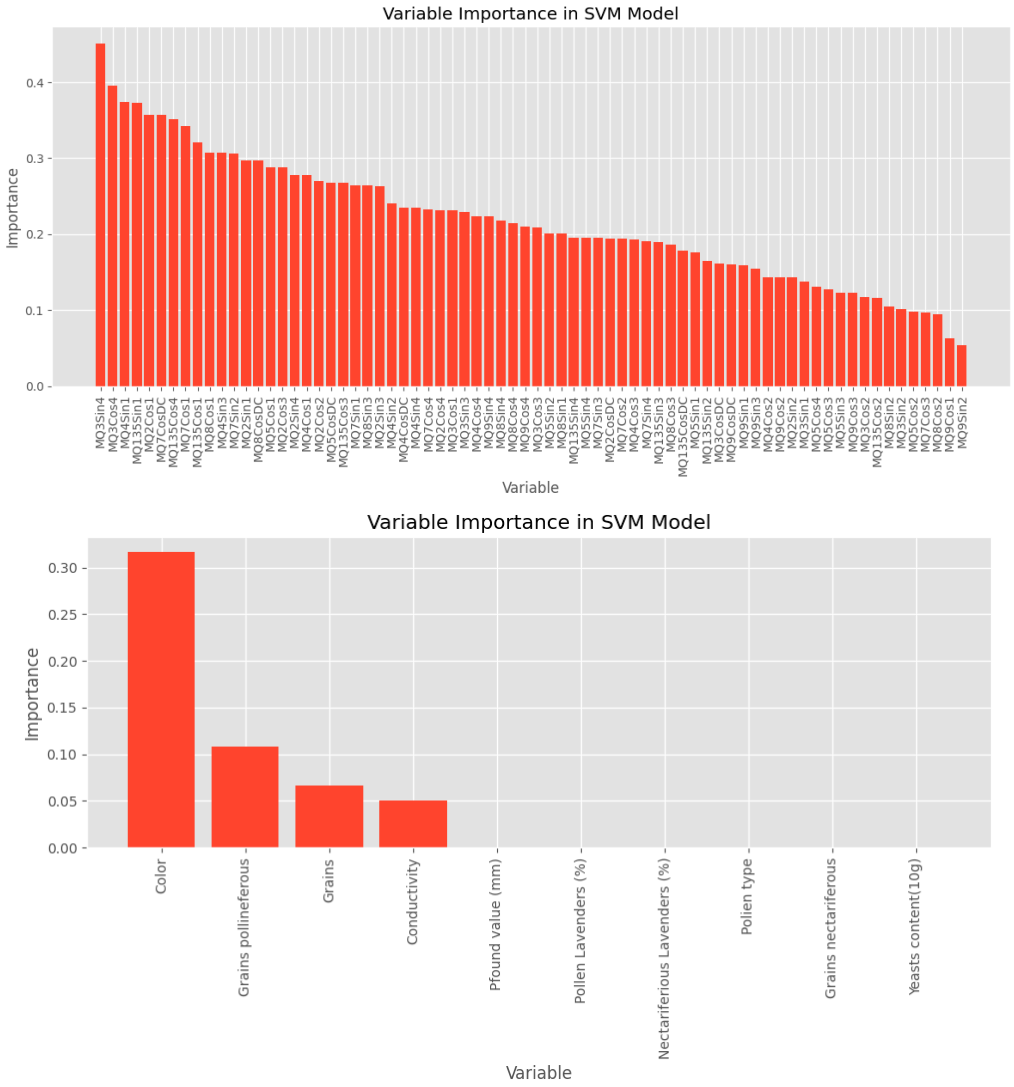



**Figure 10.** Predictive study of the SVM model.

Regarding the importance of the variables of the SVM model, color is the most important variable. Next in importance are pollen grains, followed by grains, conductivity. The other variables are not important in the model.

## 4. Conclusions

Volatile organic compounds can help distinguish monofloral honeys based on their sources and offer insights into the honey's botanical and geographical provenance.

The use of this e-nose prototype has allowed simplifying the analysis without the need to require a qualified specialist and reducing costs.

The knowledge of the volatile profile of a honey and its chemical fingerprint has allowed to discriminate the type of honey based on its floral origin.

The results of the preliminary study allowed to classify two groups based on the percentage of pollen contained in the honey, obtaining a higher response from the sensors for those samples that had a pollen content higher than 54%.

The SVL classification method has allowed to distinguish the varieties of honeys and it is shown as an alternative method to the PCA in the classification of monofloral honeys of the experiment developed.

For this experiment, the alcohol-related sensors (MQ3) are the ones that give the highest response due to the chemical composition of both honey and pollen. Similarly, the MQ135 gives a significant response, due to the wide range of VOCs to which it is sensitive.

This used prototype has been used for quick classification of agrifood products (olive oil, wine, coffee, lemons, among others) and medical purposes (urine analysis). The inconvenience of this prototype is the previous calibration. A period of several months (depending on the number of samples) has to be used for calibration. But finally, this device can be used for the studied substances.

Another problem is the repeatability of the device. There is a varied response in each device. This can be solved with a better quality sensor.

The next step for this device is the development of a commercial purpose.

**Author Contributions:** Conceptualization, E.G.M., A.M.L. and A.R.-C.; Formal analysis, E.G.M., A.M.L., A.R.-C. and A.C.C.; Investigation, E.G.M., A.M.L., A.R.-C., A.C.C., G.M.M. and M.J.O.; Writing—original draft, E.G.M., A.M.L. and A.R.-C.; Supervision, E.G.M., A.M.L. and A.R.-C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study did not require ethical approval.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Dataset analyzed in: https://www.dropbox.com/scl/fo/1o9d3fcq8lppa9jw370i9/h?dl=0&rlkey=f6fzhj7tu2gguln7p59ewk5nj. (accessed on 15 August 2023).

**Acknowledgments:** Authors are very grateful to TELENATURA EBT for its technical support, Asociación de Apicultores de la Comunidad de Madrid (APISCAM) and MIEL ANTONIO SIMON SL for providing us with the honey samples used in this study.

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

## References

1. De la Cruz, D.R.; Reyes, E.S.; Durán, S.S.; Sánchez, J.S. Análisis palinológico de mieles comerciales monoflorales. *Bot. Complut.* **2013**, *37*, 171–180. [CrossRef]
2. Importancia del Análisis de Polen en Miel. Available online: https://infomiel.com/importancia-del-analisis-polen-miel/ (accessed on 5 August 2023).
3. Jacinto-Pimienta, S.Y.; Mendoza-Hernández, J.H.R.; Zaldivar-Cruz, J.M.; Sol-Sánchez, Á.; Vargas-Villamil, L.M.; Reyes-Sánchez, C.A.; Jacinto-Pimienta, S.Y.; Mendoza-Hernández, J.H.R.; Zaldivar-Cruz, J.M.; Sol-Sánchez, Á.; et al. El uso de componentes principales en la clasificación melisopalinológica de la miel de *Apis mellifera* L. *Rev. Mex. Cienc. Agríc.* **2016**, *7*, 2831–2840.
4. Ballesteros, E.P.; Riveros, A.C.; Acuña, F.R.T. Determinantes Físicoquímicos de La Calidad de La Miel: Una Revisión Bibliográfica. *Cuad. Desarro. Rural* **2019**, *16*, 1–15. [CrossRef]
5. Pinto, P.; María, A. *Identificación de Marcadores Químicos en Mieles Monoflorales Españolas*; Project/Final Work/Degree; Universitat Politècnica de València: Valencia, Spain, 2021.
6. Schanzmann, H.; Augustini, A.L.R.M.; Sanders, D.; Dahlheimer, M.; Wigger, M.; Zech, P.-M.; Sielemann, S. Differentiation of Monofloral Honey Using Volatile Organic Compounds by HS-GCxIMS. *Molecules* **2022**, *27*, 7554. [CrossRef]
7. Tedesco, R.; Scalabrin, E.; Malagnini, V.; Strojnik, L.; Ogrinc, N.; Capodaglio, G. Characterization of Botanical Origin of Italian Honey by Carbohydrate Composition and Volatile Organic Compounds (VOCs). *Foods* **2022**, *11*, 2441. [CrossRef]
8. Bonini, A.; Guidici, V.; Leffer, V.; Muchiutti, N.; Fagúndez, G.; Tamaño, G.; Daners, G.; Guarinoni, G.; Almirón, F.; Fielder, S.; et al. *Combinación Quimiométrica de Espectroscopia NIR y de Composición Volátil Como Herramienta Para Caracterizar Mieles Cítricas Monoflorales*; Laboratorio Tecnológico del Uruguay (LATU): Montevideo, Uruguay, 2015.
9. Elbanna, K.; Attalla, K.; Elbadry, M.; Abdeltawab, A.; Gamal-Eldin, H.; Fawzy Ramadan, M. Impact of Floral Sources and Processing on the Antimicrobial Activities of Different Unifloral Honeys. *Asian Pac. J. Trop. Dis.* **2014**, *4*, 194–200. [CrossRef]
10. Montenegro, G.; Gómez, M.; Díaz-Forestier, J.; Pizarro, R. Aplicación de la Norma Chilena Oficial de Denominación de Origen Botánico de la Miel Para la Caracterización de la Producción Apícola. *Cienc. Investig. Agrar.* **2008**, *35*, 181–190. [CrossRef]
11. Escriche, I.; Sobrino-Gregorio, L.; Conchado, A.; Juan-Borrás, M. Volatile Profile in the Accurate Labelling of Monofloral Honey. The Case of Lavender and Thyme Honey. *Food Chem.* **2017**, *226*, 61–68. [CrossRef]
12. Sesta, G.; Piana, M.L.; Persano Oddo, L.; Lusco, L.; Belligoli, P. Methyl Anthranilate in Citrus Honey. Analytical Method and Suitability as a Chemical Marker. *Apidologie* **2008**, *39*, 334–342. [CrossRef]
13. Zapater, S.; María, L. *Caracterización de la Fracción Volátil de Mieles de Espliego y de Tomillo Comercializadas en la Comunidad Valenciana*; Universitat Politècnica de València: Valencia, Spain, 2015.
14. Borowik, P.; Adamowicz, L.; Tarakowski, R.; Wacławik, P.; Oszako, T.; Ślusarski, S.; Tkaczyk, M. Development of a Low-Cost Electronic Nose for Detection of Pathogenic Fungi and Applying It to Fusarium oxysporum and Rhizoctonia solani. *Sensors* **2021**, *21*, 5868. [CrossRef]

15. Tarapatskyy, M.; Sowa, P.; Zaguła, G.; Dżugan, M.; Puchalski, C. Assessment of the Botanical Origin of Polish Honeys Based on Physicochemical Properties and Bioactive Components with Chemometric Analysis. *Molecules* **2021**, *26*, 4801. [CrossRef] [PubMed]
16. Mateo, F.; Tarazona, A.; Mateo, E.M. Comparative Study of Several Machine Learning Algorithms for Classification of Unifloral Honeys. *Foods* **2021**, *10*, 1543. [CrossRef] [PubMed]
17. Roubik, D.W. (Ed.) *Ecology and Natural History of Tropical Bees*; Cambridge Tropical Biology Series; Cambridge University Press: Cambridge, UK, 1989; ISBN 978-0-521-42909-2.
18. Puerto, N.; Prieto, G.; Castro, R. Chemical Composition and Antioxidant Activity of Pollen. Review. *Chil. J. Agric. Anim. Sci.* **2015**, *31*, 115–126.
19. Rajs, B.B.; Primorac, L.; Gal, K.; Bubalo, D.; Prđun, S.; Flanjak, I. Influence of botanical origin on phenolic content and antioxidant capacity of monofloral bee pollen. *Acta Sci. Pol. Technol. Aliment.* **2022**, *21*, 213–222. [CrossRef]
20. Rodríguez-Flores, M.S.; Escuredo, O.; Seijo, M.C.; Rojo, S.; Vilas-Boas, M.; Falcão, S.I. Phenolic Profile of Castanea Bee Pollen from the Northwest of the Iberian Peninsula. *Separations* **2023**, *10*, 270. [CrossRef]
21. Qiao, J.; Feng, Z.; Zhang, Y.; Xiao, X.; Dong, J.; Haubruge, E.; Zhang, H. Phenolamide and Flavonoid Glycoside Profiles of 20 Types of Monofloral Bee Pollen. *Food Chem.* **2023**, *405*, 134800. [CrossRef]
22. Benedetti, S.; Mannino, S.; Sabatini, A.; Marcazzan, G. Electronic nose and neural network use for the classification of honey. *Apidologie* **2004**, *35*, 397–402. [CrossRef]
23. Faal, S.; Loghavi, M.; Kamgar, S. Physicochemical properties of Iranian ziziphus honey and emerging approach for predicting them using electronic nose. *Measurement* **2019**, *148*, 106936. [CrossRef]
24. Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. *Real Decreto 523/2020, de 19 de Mayo, Por el Que Se Modifica el Real Decreto 1049/2003, de 1 de Agosto, Por el Que Se Aprueba la Norma de Calidad Relativa a la Miel*; Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática: Madrid, Spain, 2020; Volume BOE-A-2020-6513, pp. 43006–43008.
25. Resultados Mieles ABRIL 2023. Available online: https://www.dropbox.com/scl/fo/1o9d3fcq8lppa9jw370i9/h?dl=0&rlkey=f6fzhj7tu2gguln7p59ewk5nj (accessed on 5 August 2023).
26. Barra, M.P.G.; Ponce-Díaz, M.C.; Venegas-Gallegos, C. Volatile compounds in honey produced in the central valley of Ñuble province, Chile. *Chil. J. Agric. Res.* **2010**, *70*, 75–84.
27. Spano, N.; Casula, L.; Panzanelli, A.; Pilo, M.I.; Piu, P.C.; Scanu, R.; Tapparo, A.; Sanna, G. An RP-HPLC determination of 5-hydroxymethylfurfural in honey: The case of strawberry tree honey. *Talanta* **2006**, *68*, 1390–1395. [CrossRef]
28. Da Silva, P.M.; Gauche, C.; Gonzaga, L.V.; Costa, A.C.O.; Fett, R. Honey: Chemical composition, stability and authenticity. *Food Chem.* **2016**, *196*, 309–323. [CrossRef] [PubMed]
29. Celdrán, A.C.; Oates, M.J.; Molina Cabrera, C.; Pangua, C.; Tardaguila, J.; Ruiz-Canales, A. Low-Cost Electronic Nose for Wine Variety Identification through Machine Learning Algorithms. *Agronomy* **2022**, *12*, 2627. [CrossRef]
30. Oates, M.; Fox, P.; Sánchez-Rodríguez, L.; Carbonell-Barrachina, 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]
31. Oates, M.J.; Abu-Khalaf, N.; Molina-Cabrera, C.; Ruiz-Canales, A.; Ramos, J.; Bahder, B.W. Detection of Lethal Bronzing Disease in Cabbage Palms (Sabal Palmetto) Using a Low-Cost Electronic Nose. *Biosensors* **2020**, *10*, 188. [CrossRef] [PubMed]
32. Machuca, G.; Staforelli, J.; Rondanelli-Reyes, M.; Garces, R.; Contreras-Trigo, B.; Tapia, J.; Sanhueza, I.; Jara, A.; Lamas, I.; Troncoso, J.M.; et al. Hyperspectral Microscopy Technology to Detect Syrups Adulteration of Endemic Guindo Santo and Quillay Honey Using Machine-Learning Tools. *Foods* **2022**, *11*, 3868. [CrossRef] [PubMed]
33. Hategan, A.R.; Magdas, D.A.; Puscas, R.; Dehelean, A.; Cristea, G.; Simionescu, B. Machine Learning Algorithms in Corroboration with Isotope and Elemental Profile—An Efficient Tool for Honey Geographical Origin Assessment. *Appl. Sci.* **2022**, *12*, 10894. [CrossRef]
34. David, M.; Hategan, A.R.; Magdas, D.A.; Berghian-Grosan, C.; Simionescu, B. Botanical Origin Assessment of Honey Based on ATR-IR Spectroscopy: A Comparison between the Efficiency of Supervised Statistical Methods and Artificial Intelligence. *Appl. Sci.* **2022**, *12*, 9645. [CrossRef]
35. Viciano-Tudela, S.; Sendra, S.; Parra, L.; Jimenez, J.M.; Lloret, J. Proposal of a Gas Sensor-Based Device for Detecting Adulteration in Essential Oil of Cistus Ladanifer. *Sustainability* **2023**, *15*, 3357. [CrossRef]
36. Escuredo, O.; Rodríguez-Flores, M.S.; Míguez, M.; Seijo, M.C. Multivariate Statistical Approach for the Discrimination of Honey Samples from Galicia (NW Spain) Using Physicochemical and Pollen Parameters. *Foods* **2023**, *12*, 1493. [CrossRef]
37. Seijo, M.C.; Escuredo, O.; Rodríguez-Flores, M.S. Physicochemical Properties and Pollen Profile of Oak Honeydew and Evergreen Oak Honeydew Honeys from Spain: A Comparative Study. *Foods* **2019**, *8*, 126. [CrossRef]
38. Lawag, I.L.; Lim, L.-Y.; Joshi, R.; Hammer, K.A.; Locher, C. A Comprehensive Survey of Phenolic Constituents Reported in Monofloral Honeys around the Globe. *Foods* **2022**, *11*, 1152. [CrossRef] [PubMed]